Power System Security Assessment - Application of Learning Algorithms

Andersson, Christian

2005

LUND UNIVERSITY

PO Box 117
221 00 Lund
+46 46-222 00 00

Power System Security Assessment

Application of Learning Algorithms

Christian Andersson

LUND UNIVERSITY

Licentiate Thesis
Department of Industrial Electrical Engineering
and Automation

Department of
Industrial Electrical Engineering and Automation
Lund University
Box 118
SE-221 00 LUND
SWEDEN

http://www.iea.lth.se

ISBN 91-88934-39-X
CODEN:LUTEDX/(TEIE-1047)/1-90/(2005)

Printed in Sweden by Media-Tryck
Lund University
Lund, 2005

MALMÖ UNIVERSITY

School of Technology and Society
Malmö University

Abstract

The last years blackouts have indicated that the operation and control of power systems may need to be improved. Even if a lot of data was available, the operators at different control centers did not take the proper actions in time to prevent the blackouts. This depends partly on the reorganization of the control centers after the deregulation and partly on the lack of reliable decision support systems when the system is close to instability. Motivated by these facts, this thesis is focused on applying statistical learning algorithms for identifying critical states in power systems. Instead of using a model of the power system to estimate the state, measured variables are used as input data to the algorithm. The algorithm classifies secure from insecure states of the power system using the measured variables directly. The algorithm is trained beforehand with data from a model of the power system.

The thesis uses two techniques, principal component analysis (PCA) and support vector machines (SVM), in order to classify whether the power system can withstand an $(n-1)$-fault during a variety of operational conditions. The result of the classification with PCA is not satisfactory and the technique is not appropriate for the classification problem. The result with SVM is much more satisfying and the support vectors can be used on-line in order to determine if the system is moving into a dangerous state, thus the operators can be supported at an early stage and proper actions can be taken.

In this thesis it is shown that the scaling of the variables is important for successful results. The measured data includes angle difference, busbar voltage and line current. Due to different units, such as kV, kA and MW, the data must be preprocessed to obtain classification results that are satisfactory. A new technique for finding the most important variables to measure or supervise is also presented. Guided by the support vectors the variables which has large influence on the classification are indicated.

Contents

Preface

During the final year of my B.Sc. studies in 2001 I was introduced to the challenges related to planning and operation of power system by my current PhD-advisor Bo Eliasson. I was studying computer and electrical engineering at Campus Norrköping, Linköping Institute of Technology. When planning my undergraduate project I contacted Daniel Karlsson at ABB Technology Products and came to work with wide area protection system against voltage collapse. The project was supervised by Bo Eliasson at Malmö University. It was during my undergraduate project I became aware of the special problems with multi input multi output system (MIMO), many of the power systems today are probably the largest MIMO system ever built by mankind. In the summer of 2002 I was appointed a PhD position at Malmö University and was accepted to the post-graduate study program at Lund Institute of Technology.

Acknowledgement

First of all, I would like to thank my supervisors, Bo Eliasson for introducing me to the interesting problems in power system control and for his support and advice. Gustaf Olsson for accepting me as a Ph.D. student at his department and for his encouragement and optimism. Olof Samuelsson who leads the electric power system automation research group at IEA, the guidance, support and comments on my work have been very important.

I am also very grateful to Sture Lindahl for support and many interesting discussions. Jan Erik Solem for introducing me to pattern recognition. The constant guidance, support and unlimited help has helped me a lot, this thesis had not been made without you. Lars Lindgren for help with the ARISTO simulator and all the discussions about power systems we have had, I have learned a lot from you.

I would like to thank the graduate students and researchers of the Industrial Electrical Engineering and Automation at Lund Institute of Technology and Applied Mathematics Group at Malmö University, especially to: Erik Alpkvist, Yuanji Cheng, Pär Hammarstedt, Adam Karlsson, Niels Christian Overgard, Fredrik Nyberg and Markus Persson for providing a creative, happy and sometimes confused atmosphere.

Furthermore, I would like to thank Gun Cervin and Yvonne Olsson at Kranenbiblioteket, Malmö University for their excellent service. I am also grateful to Jan Erik Solem, Olof Samuelsson and Christian Rosén for proofreading this thesis.

I would like to thank my family and friends, especially my two younger sisters for their patience, even though I am always teasing them, *(Josefin, jag vet att jag är ditt elände).* Finally, Cissi, you´re the best.

Malmö, August 21, 2005
Christian Andersson

Chapter 1

Introduction

In August and September 2003 three major blackouts occurred in N.E. North America, S. Scandinavia and Italy. About 110 million people were affected by these events. The interruption time ranged from a couple of hours to days. The reason why these blackouts occurred were partly due to the lack of supportive applications when the system is close to instability. If the power system is close to the stability limit, actions must be taken by system operators to counteract a prospective blackout. Today a large problem for system operators is to identify critical states since there are thousands of parameters which describe and affect the state.

The deregulation of the power markets has contributed to the fact that the power systems today are frequently operated close to stability limit. This is due to a larger focus on economical profits instead of system security and preventive maintenance.

The last three years the consumption of electrical power in the world has increased with 3 %, each year. That corresponds to investments of 60 billion USD each year in production and transmission equipment to match the increased consumption. If the trend continues actions must be taken to maintain system stability.

The latest blackouts and deregulated power markets contribute to an increased requirement of Wide Area Protection Schemes (WAPS). The WAPS will make it possible to change the operation criteria from: *"The power system should withstand the most severe credible contingency"* to *"The power system should withstand the most severe credible contingency, followed by protective actions from the WAPS."*

In this thesis *classification of power system stability* defines a method to separate secure from insecure states in power systems. Due to the complexity of power system with the large amount of states and the large amount of variables that describe the states it is difficult to estimate the stability of power systems. The purpose with this thesis is to examine methods for classifying power system stability with techniques from pattern recognition. The methods use the measured variables, such as busbar voltage, line current and active current of generators, of the power system to decide the stability of the power system. The classification algorithm which is developed with techniques for pattern recognition is trained from a data base. The data base, denoted **training set**, consists of measurements from different states of the power system. With the algorithm it should be able to estimate the stability of all the states of the power system may operate in. Therefore it is important that the training set properly describes the variations of the power system. It must be considered that if the training set does not show enough variability, the algorithm is useless. A properly trained algorithm can separate the training set into two parts. One consists of stable states and the other of unstable states. The algorithm can then be used to decide the class of unknown states.

This should be used as a supportive tool for the system operators to faster form an opinion of the actual state of the power system. This could also make it easier to decide what needs to be done by the system operators to avoid a blackout. Compared with the WAPS the classification method should prompt system operators at control centers to take preventive actions before the WAPS will act.

The work with this thesis started with the development of the training set. Thereafter Principal Component Analysis (PCA) was applied to the classification problem. The result with PCA was not satisfactory. Therefore Support Vector Machines (SVM) was used to find a solution of the classification problem. SVM calculates a hyperplane that separates a data set from their predefined classes. The SVM method is more adaptable to the classification problem in this thesis than PCA and the result is much more satisfying.

1.1 Previous work

This thesis deals with the problem of classification of power system stability. An overview of power system stability and design can be found in the standard textbooks [13] and [22]. In [23] the CIGRE Study Committee 38 and the IEEE Power System Dynamic Performance Committee gives a background and overview of power system stability with the focus on definition and classification. A nice introduction to power system voltage stability can be found in [34], which has a detailed description of reactive power compensation and control. These books and reports gives a good overview and background to power system design and control.

Previous work on power systems stability has been focused on identification and finding methods to prevent instability. In [6] and [26] the problem with voltage instability is analyzed, these two theses are a part of the previous work that is the inspiration to this thesis. Another work that has been a great inspiration to this thesis is the research by Louis Wehenkel which is summarized in [39]. Wehenkel uses the same approach to the stability problem as in this thesis but he uses decision trees instead of principal component analysis and support vector machines.

The knowledge about separation with a hyperplane has been known for quite long time, but the calculation of the hyperplane has caused problems. In 1992 Support Vector Machines (SVM) was introduced by Vapnik, see [37]. It is described in detail in [36]. The SVM calculate a hyperplane with the maximal margin between the two classes. This is done using a global optimization problem on quadratic form, which can easily be solved.

The last years SVM has been used to find methods to solve many classification and regression problems. In [16] and [31], SVM has been applied to face detection with successful results. Text categorization has also been improved with SVM, see [18]. In [19], windows are detected from pictures of city scenes. The author has found one publication which deals with classification of power system stability using SVM, see [30]. In the article classification of power system stability is made with nonlinear SVM and neural networks. The data set which is used has an impressive size, and the result of the classification is very successful.

1.2 Objectives and contributions of the thesis

The main objective of this thesis is to investigate if it is possible to use statistical learning algorithms for classification of power system stability. The objectives of this thesis are listed below.

- Find a method which makes it possible to classify the stability of a power system from measurements without calculating the state of the power system.

- Adjust the classification method for the application to power system stability.

- Investigate if the classification method can be implemented in real-time so that operation of power systems can be simplified and also improved.

People with different background should be able to read and comprehend the thesis. This means that some readers will consider some parts of the thesis as obvious. However, if the reader is not oriented in power system fundamentals an introduction can be found in [13] and [22]. The main contributions of the thesis are listed below.

- The power system stability is classified with a hyperplane, which is calculated with Support Vector Machines (SVM). The SVM offers a classification method with the largest distance between the two classes.

- Methods to improve the classification by using an appropriate scaling algorithm. Investigations of which variables should be used to represent the power system in order to improve the classification result.

- Methods to reduce the number of measured variables are introduced. Thus the classification can be made by a subset of the measurable variables without losing the number of correct classifications.

- Method for handling the difficulties of classification with statistical learning algorithm with a mixture of continuous and binary variables.

1.3 Outline of the thesis

The thesis consists of two parts, the design of the data set and two different methods applied to the classification problem.

In the first part, Chapter 2, an overview of power system stability and operation is given. Furthermore, the contribution of this thesis to the power stability problem is described. In Chapter 3, a short summary of the NORDEL power system and the operation rules is given. Thereafter, an introduction to design of experiments with factorial design is given. Finally, the design of the data set is described.

The second part consists of Chapter 4 and Chapter 5. In Chapter 4 the unsupervised learning method principal component analysis is applied to the classification problem and in Chapter 5 the supervised learning method of calculating a hyperplane with support vector machines is applied to the classification problem. Chapter 6 summarizes the results and suggests future work.

1.4 Publications

The thesis is based on the following publications:

[1] C. Andersson, J E. Solem and B. Eliasson. Classification of power system stability using support vector machines. In *IEEE Power Engineering Society General Meeting.* San Francisco, California, USA, June 12-16. 2005.

[2] L. Lindgren, B. Eliasson and C. Andersson. Damping of power system oscillations using phasor measurement units and braking resistors. In *IFAC proceedings of reglermöte.* Gothenburg, Sweden, May 26-27. 2004.

[3] B. Eliasson and C. Andersson. New selective control strategy of power system properties. In *IEEE Power Engineering Society General Meeting.* Toronto, Ontario, Canada, July 13-17. 2003.

[4] B. Eliasson and C. Andersson. New selective control strategy of the voltage collapse problem. In *CIRED 17th International Conference on Electricity Distribution.* Barcelona, Spain, May 12-15. 2003.

Chapter 2

Power system control

This chapter consists of a brief overview of power system operation and the security limits that are taken to maintain system stability in stressed situations.

2.1 Power system stability

The stability of power systems is defined as: *"The ability of an electric power system, for a given initial operating condition, to regain a state of operating equilibrium after being subjected to a physical disturbance, with most system variables bounded so that practically the entire system remains intact"*, [23]. This means that some faults should not cause any instability problems. Figure 2.1 shows different stability states. Due to the complexity of power systems, the different instabilities affect each other. A blackout scenario often starts with a single stability problem but after the initial phase other instability problems may occur. Therefore the actions in order to prevent a blackout in stressed situations may be different. For a detailed description of the different stability scenarios see [23]. Due to mainly economical factors there is a limit of which faults the system should withstand and remain in stability.

The synchronous NORDEL system (Sweden, Norway, Finland and Sealand) is dimensioned to fulfill $(n-1)$ criterion, which means that the system should withstand a single fault, e.g. short circuit of busbar or disconnection of nuclear unit, see [1]. In Figure 2.2 the different operative states are illustrated and how the power systems operation state can change due to a disturbance.

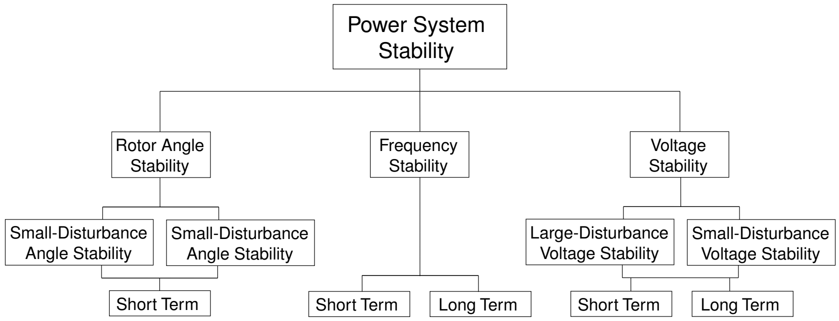


Figure 2.1: Classification of power system stability. Adapted from [23] .

In the secure state where the operating point of a power system should normally be it is possible to maintain system stability after a $(n-1)$ fault. It is very important to have tools to verify the operating point of the system. In this thesis a new method to improve the classification of the power system stability is introduced. The method also makes it possible to deicide how close the system is to the dividing line between the secure and insecure state.

However, faults which are included in the $(n-1)$ criterion sometimes cause blackouts. In Sweden this happened Dec. 27, 1983 when a broken disconnector caused a short circuit of a busbar which lead to a blackout where 60 $\%$ of the load was disconnected from the system. The time to restore normal operation was about 7 h. In the N.E. of North America on Aug. 14, 2003, 50 million people were affected by a blackout caused by a cascade disconnection of transmission lines. The time between the disconnection of the first transmission line and the blackout was 2,5 hours and the restoration took almost 48 hours. In Italy in Sep. 28, 2003, 53 million people were affected by a blackout which also was caused by a cascade disconnection of transmission lines. The fault scenario was a bit more difficult to overview than in USA and Canada. First a transmission line was disconnected due to a short-circuit caused by a line sagging into a tree and then after 25 minutes twelve other lines were disconnected within 2,5 minutes which caused the blackout. The restoration time was 18 hours and the number of switching operations that had to be made was about 3600.

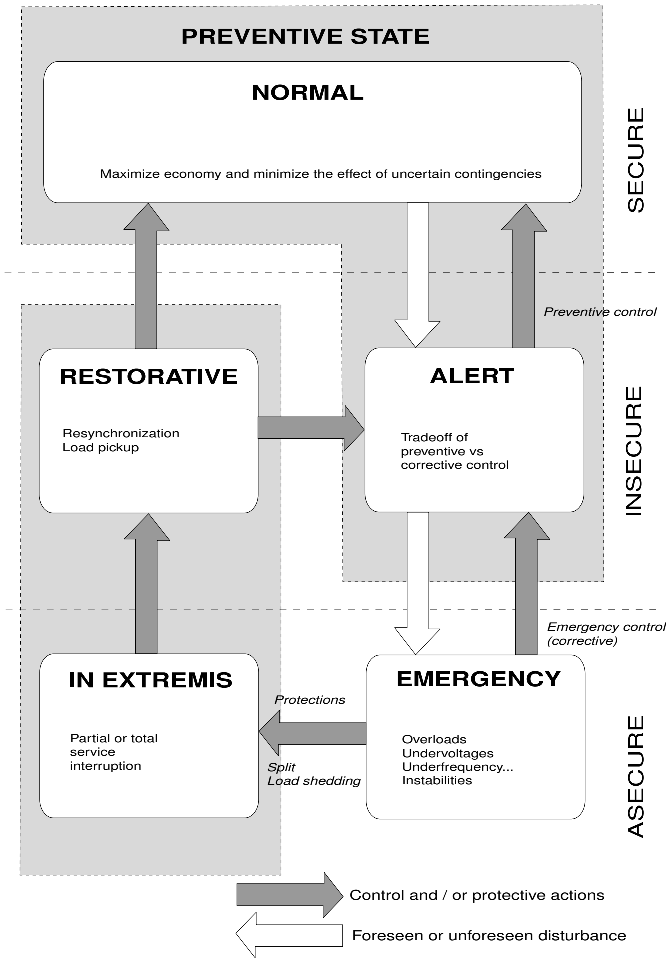


Figure 2.2: The operating states and transitions for power systems. Adapted from [12] .

In Figure 2.2 the state during abnormal situations are described with the insecure and asecure states. If a power systems operating point moves from the secure to the insecure state, actions must be taken by system operators to return to the secure state. By the three blackouts in Sweden, N.E. North America and Italy, it is shown that the preventive time interval can be between 50 seconds to 2,5 hours. From a system operation point of view it is possible to take actions before entering the insecure state. This thesis has focus on supportive information, before entering the insecure area. In the asecure state, only automatic actions and control are realistic.

Blackouts have occurred and will happen in the future, but the objective from the power industry is to increase the time between the blackouts and also minimize the number of people which are affected. By e.g. supportive applications for the operators based on new technology, the prevention against blackouts can be improved. By increased automation the imbalances between production and consumption can be quickly alleviated. Also the black-start capability after a blackout can be improved.

2.2 Power system operation

In power system control today the operating point is validated by a state estimator and tools to calculate system collapse limit. There are different software tools developed for this. In [35] the experience with implementing a voltage security assessment tool at B.C. Hydro, Burnaby, Canada is described. In Sweden today **SPICA** is used to calculate the voltage collapse limits. SPICA is developed by Swedish National Grid (Svenska Kraftnät) and uses an estimate to calculate the voltage collapse limits by load-flow calculations. Mainly by increasing the load in sensitive parts of the system and exposing them for a selection of $(n-1)$ faults. The predictor proposes changes in such as,

- Transfer limits
- Increased production
- Decreased export by HVDC,

if the estimated operating point does not fulfill the $(n-1)$ condition.

The limitations with the state estimators are that they use a model of the power system. There is always a difference between the model and the real system. It is also common to use load-flow calculations. In a load-flow calculation the dynamic part of the elements in the model are neglected e.g. load dynamic and also often voltage regulators and current limiters. The load-flow solver calculates an operating point where there is balance in active and reactive power. There is a difference between load-flow calculation and dynamic simulation. When the operating condition changes, e.g. by a short-circuit of a busbar, the load-flow solver calculates the operating point after clearance of the fault. The dynamic simulation also analyzes the effect by e.g. automatic equipment. Hence the operating point after the transients are settled will be different compared with a load-flow calculation. This can result in that the operating point calculated by load-flow is stable but if the scenario is analyzed by dynamic simulation a blackout can occur in the transient phase.

In this thesis, dynamic simulations are used to develop classification methods which are used to estimate the bound between the secure and the insecure state, see Figure 2.2. Instead of using a traditional state estimator, statistical learning algorithms are used to calculate an algorithm which uses measured variables, e.g. busbar voltage, angle difference and line current to classify the stability.

Chapter 3

Data set

In this chapter the data set for verifying the statistical learning algorithms applied on classification of power system stability is described. With the algorithm is should be possible to estimate the class for all states in the power system. It is therefore important that data set properly describe the variations of the system. It must be considered that if the data set does not show enough variability, the SVM cannot be successfully applied.

To create a function for classification with a learning machine with the best conditions, the space of operation should be covered by the data set. In e.g. chemical processes, design of experiments are used to cover the operation space with a limited number of experiments, for details see [10].

The data set used to verify the classification algorithms has not been designed with experimental design since it would become very large. The time it takes to prepare the simulator and collect the data are not covered in this project. Instead, the knowledge about the daily operation of the NORDEL system and the problems that has caused stressed operation situations has been taken into account to create the data set.

3.1 Data set structure

In this thesis the data set is organized as a matrix, see Table 3.1. A row in the matrix, denoted with an **example**, is the state of the power system containing the measured variables. The label vector consists of the **binary code** for each example. If the example sustains the dimension criterion $(n-1)$, see Section 3.5, has the binary code +1, otherwise -1.

Table 3.1: Description of how the data set is organized, the label vector consists of the binary code to the corresponding examples.

	Variables					Label
Example 1						+1
⋮						⋮
Example l						-1

3.2 The NORDEL power system

In this presentation of NORDEL, Iceland and Jutland (Denmark) are excluded, for details see Figure 3.1. The synchronous system consisting of Finland, Norway, Sealand (Denmark) and Sweden is presented. The installed capacity was 93 GW on Dec. 31, 2003, the maximum load was 58 GW, measured on Jan. 15, 2003. In Table 3.2 the distribution of the installed capacity per country is listed. The information about the NORDEL power system published here is obtained from [4].

In Sweden, a large amount of hydro power is located in the northern part and nuclear power is located in the southern part. Because of a larger load demand in the south of Sweden the power flow normally is from north to south. The latest blackouts in Sweden (1983 and 2003) occurred because of reduced capacity in the transmission system, see [3] and [7]. The NORDEL system data is not used here, but instead a simplified model called NORDIC32 is used. This simplified model should be compared with the NORDEL power system in order to give guidelines for the design of the data set. To classify the stability of a power system, traditionally bus voltage and system frequency have been used. In Sweden the transfer of active power on roughly 10 selected transmission lines are used as guidance for the system stability. The transmission system is divided into 4 cut-sets, called $\mathbf{Snitt\ 1-4}$. Snitt 4 measures the sum of active power of the 5 transmission lines that connect the south of Sweden with the rest of the system, see Figure 3.1. Operation aims at keeping the transport of active power through Snitt 4 below a certain threshold to retain system stability from the most severe credible contingency.

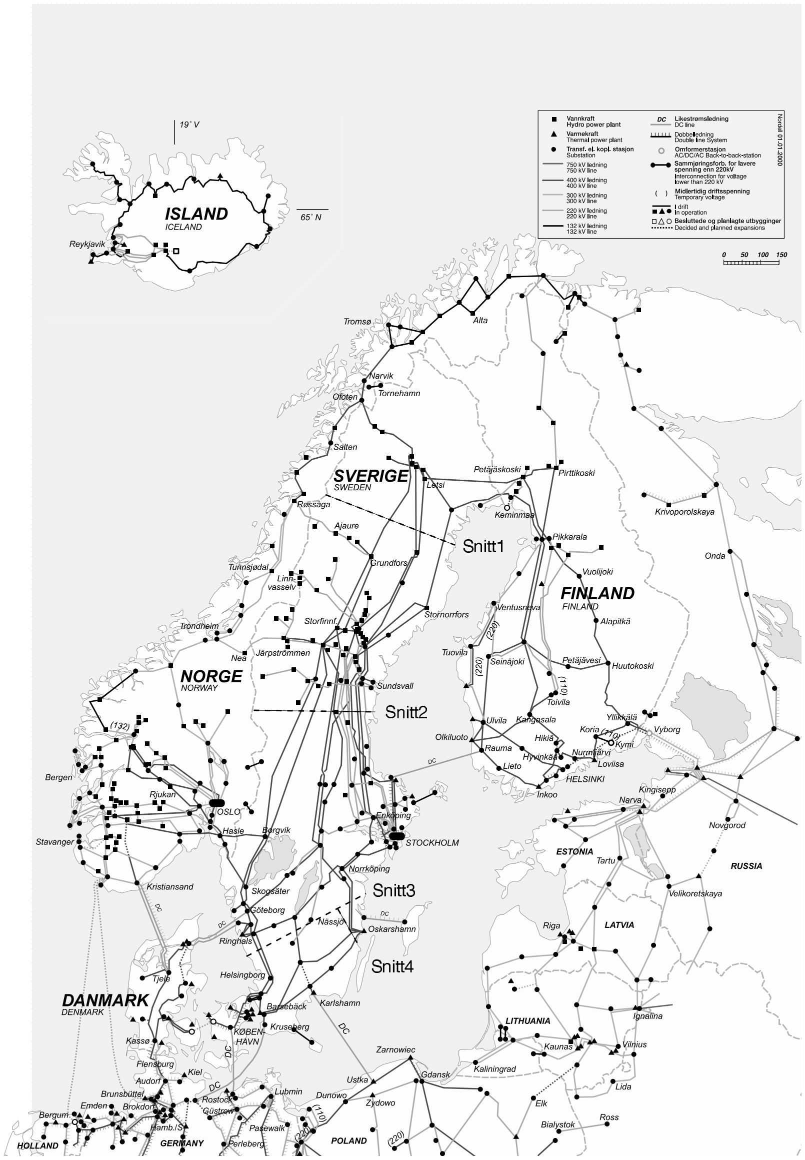


Figure 3.1: The NORDEL main grid, 2000, The cut-sets, Snitt 1 to 4, are marked.

Table 3.2: Installed capacity in NORDEL on 31 Dec. 2003, MW.

	Denmark	Finland	Norway	Sweden
Hydro power	11	2 978	27 676	16 143
Nuclear power	-	2 640	-	9 441
Other thermal power	9 704	11 225	305	7 378
Renewable power	3 115	50	100	399

3.3 NORDIC32

The NORDIC32 (N32) system is used for the study [5]. It is an equivalent or simplified model of a transmission system with similarities to the Swedish power system, see Figure 3.2. The generation consists of approximately 45 % nuclear, 45 % hydro and 10 % other thermal power. The installed capacity is 17 GW, see Table 3.3. The system has 32 substations, 35 units, 50 branches and the number of measurable variables are 653. To adjust the spinning reserve and speed droop the relation between N32 and NORDEL are settled to approximately 1:4.

In order to produce the data set, the platform described below was used. Since 1996 the simulator ARISTO (Advanced Real-time Interactive Simulator for Training and Operation) has been used for operator training at Swedish National Grid (Svenska Kraftnät). The simulator has unique features for stability analysis in real time [9]. Not only that, the simulation proves to be extremely robust during abnormal operating conditions. Using ARISTO the operators obtain a better understanding of the phenomena that may cause limitations and problems to power transmission and distribution systems.

The load models in ARISTO cover the load dynamics (short and long term) and load characteristics (voltage and frequency dependence) very well. Transmission lines are represented with pi-equivalents. The generation units are represented by synchronous generators and turbine governors, either thermal or hydro. The protection equipment implemented in ARISTO are

- Line distance protection
- Line over-current protection

Table 3.3: Installed capacity in NORDIC32, MW.

	Finland	Norway	Sweden
Hydro power	4 605	1 520	4 740
Nuclear power	-	-	4950
Other thermal power	-	-	1 080

- Unit over/under frequency protection
- Unit over/under voltage protection

and the automatic equipment

- Load disconnection due to under-frequency
- Shunt capacitor/reactor connect/disconnect due to under/over voltage.

The detailed representation of power system models in ARISTO results in a more realistic data set than if a simple load-flow solver had been used. The difference between load-flow calculations and dynamic simulations is that load-flow solvers calculate an operation point for the given conditions, whereas dynamic simulations calculate a time-simulation where the operating point is calculated for each time step. The advantage with dynamic simulations is that it is possible to analyze the influence of automatic protection, control equipment and dynamic models on the power system. In this study a three-phase short-circuit on a busbar is analyzed. With the dynamic simulation it is possible to take into account what has happened during the transients. With a load-flow solver it is only possible to analyze if there is balance in active and reactive power after the disconnection of the fault. Therefore, the data set is more realistic with the dynamic simulation.

The main additives to the simulator consist of a multi-user communication link between the Application Programmable Interface (API) of ARISTO and Matlab/Simulink. This function is used to collect the data set.

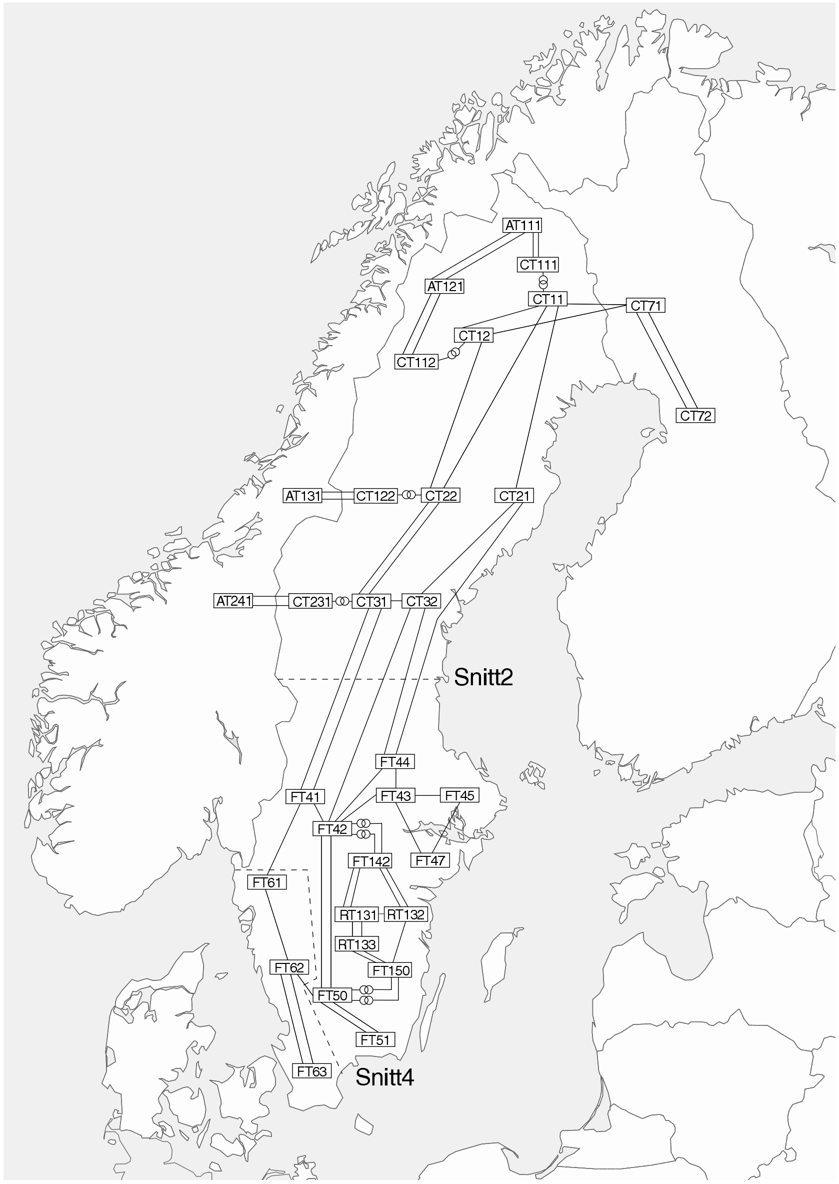


Figure 3.2: The NORDIC32 system.

3.4 The 2^k factorial design

Factorial designs are often used in designs of experiments. The purpose is to cover most of the variation in the process with a limited number of experiments, see [10] and [29]. The elements in the design consist of factors and response. Response is the output of the process. Factors are the elements in the process that influence the output. The factors can be both controllable and uncontrollable.

For classification of power system stability the label of the example is decided according to the Nordel grid planning rules for the Nordic transmission system, for details see [1]. Therefore the factorial design is applied with the following factors

- Production
- Transmission capability
- Consumption

and responses

- Bus voltages
- Angle differences
- Line current
- System frequency
- Load disconnection from system protection.

The purpose with factorial design is to define a high and a low level for every factor. In 2^3 design the factors form a cube where the examples shall be chosen from the corners of the cube, see Figure 3.3. If the factorial design is applied to the N32 system the number of simulations will be very large. Taken into account that the number of units are 35, branches 50 and loads 59 the simulation time will be very long. Approximations can of course be made. Using only nuclear power, a data set could be designed with factorial design. The number of substations with nuclear power are 5, thus the number of examples in the data set would be 32. The problem is to design a data set that also has a variation in production of hydro units, power flow, and load,

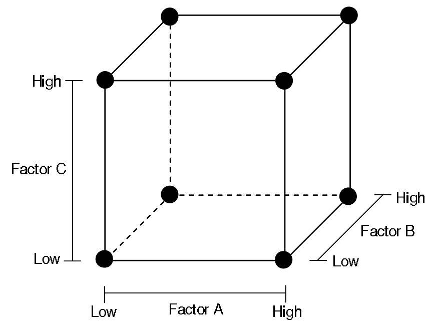


Figure 3.3: Illustration of a factorial 2^3 design.

with a limited number of examples. Therefore factorial design is not used to design the data set. Instead, knowledge from operation of the NORDEL system with already known limits has been used in the design of the data set.

3.5 Dimension criterion

According to Nordic grid instructions, the Nordel system must fulfill the $(n-1)$ criterion, see [1] and [2]. An example is considered acceptable if the power system can withstand any single failure. The system has withstood a contingency if:

- There is no load shedding, due to under frequency.
- The final post-fault voltage profile is acceptable.
- The final post-fault frequency is acceptable.

The contingencies are e.g. disconnection or short-circuit of a line, load, shunt or busbar. After fault clearance the voltage profile is acceptable if all except one busbar are within 10 % of the normal operating voltage. The system frequency is accepted if it does not drop below 49.0 and exceeds 49.5 Hz after 30 seconds. The upper level demand on the frequency is 52.0 Hz and it must be lower than 50.6 Hz after 30 seconds.

To verify the $(n-1)$ criterion for the examples, a three-phase short-circuit at all busbars, one at a time, is used to decide the binary code according to the rules above. This is done for each of the training examples. The fault is cleared 80 ms after the fault appeared, see [13]. In practice this is done in ARISTO by first applying the fault to the example, then after 80 ms the fault is cleared, finally after 240 s it is considered that the transients are approximately settled and the simulation is stopped. The binary code is then decided for the example.

Because N32 is an equivalent, with reduced topology, the $(n-1)$ criterion is not fulfilled by most of the examples in the data set. The data set is split into three classes:

1. Examples where a single fault leads to system black-out.
2. Examples where a single fault leads to black-out in two or three substations.
3. Examples which fulfill the $(n-1)$ criterion.

Therefore a method shall be developed to identify class 1 and separate class 1 from class 2 and 3.

3.6 Training set

As training data, four base generation patterns are used.

- NN, all nuclear power at maximum production.
- NO, all nuclear power at maximum production on the west coast, no nuclear power on the east coast.
- ON, all nuclear power at maximum production on the east coast, no nuclear power on the west coast.
- OO, no nuclear power on-line.

The first step is to find the minimum load level for the four generation patterns, by scaling the active power part of the load on a common system base. To create balance in the system at the minimum level the production of hydro power is corrected. As little hydro power as possible should be used to create balance in the minimum case. In order to achieve training data with the characteristics of the NORDEL power system the following dimension criteria are used for all examples.

- The system frequency is within 50 $\pm$ 0,01 Hz.
- The voltage profile is adjusted according to [24], where the normal operation voltage on the 400 kV level is 400-415 kV, with the minimum accepted voltage 395 kV and the maximum accepted voltage 420 kV.
- The spinning reserve in the system is 750 MW, divided on three hydro power units. It corresponds to the largest unit in N32.
- The ratio between N32 and NORDEL is approximately 1:4. Taken this into account the speed droop is tuned to about 1500 MW/Hz for all models. The Nordel system (Sweden, Finland, Norway and Sealand) it is 6000 MW/Hz, see [2].

When the minimum load level is achieved, a new model is created by increasing the load level by 5 % increments from the base load profile. Spinning reserve and the speed droop is tuned so that all models have similar characteristic. For the 4 generation patterns the load level is increased with 5 % until the maximum load level is identified. The total number of models in the training data is 30, with a consumption between 3 280 and 13 570 MW. This is done in order to extend the generation/transmission patterns as much as possible with a limited number of training models. The generation and consumption of reactive and active power for the 30 examples in the training set are listed in Table 3.4. The label for the examples are also listed here. Table 3.4 shows that the increased consumption for the three different base patterns, NN, NO, ON, does not imply a more stressed state.

3.7 Classification set

The classification algorithm is developed with a training set and validated with a classification set. To show the performance of the algorithm. The classification set consists of 16 examples. These are created from 16 of the examples in the training data that have been manipulated e.g. increased load factor and disconnection of single line or unit. Thus, an example which

Table 3.4: The system generation and consumption for the examples in the training set, MW and MVar.

Name	$P_{Generation}$	P_{load}	$Q_{Generation}$	Q_{load}	Label
NN1	7 489	7 329	1 550	3 359	-1
NN2	8 869	8 752	975	3 359	1
NN3	9 968	9 846	1 244	3 359	1
NN4	10 545	10 392	907	3 359	-1
NN5	11 136	10 940	1 019	3 359	-1
NN6	11 707	11 487	970	3 359	1
NN7	12 463	12 034	2 386	3 359	-1
NN8	12 948	12 581	1 640	3 359	1
NN9	13 545	13 128	2 019	3 359	1
NN10	14 075	13 566	2 832	3 359	-1
NO1	4 541	4 376	1 126	3 359	1
NO2	5 667	5 470	857	3 359	1
NO3	6 826	6 564	998	3 359	1
NO4	7 418	7 111	1 199	3 359	-1
NO5	8 021	7 658	1 440	3 359	-1
NO6	8 630	8 205	1 581	3 359	-1
NO7	9 266	8 752	2 408	3 359	-1
ON1	4 468	4 375	915	3 359	1
ON2	5 556	5 470	873	3 359	1
ON3	6 674	6 564	798	3 359	1
ON4	7 243	7 111	780	3 359	-1
ON5	7 819	7 657	777	3 359	-1
ON6	8 405	8 205	908	3 359	1
ON7	8 998	8 752	1 041	3 359	1
ON8	9 608	9 299	1 382	3 359	1
ON9	10 203	9 846	1 580	3 359	1
ON10	10 835	10 393	2 262	3 359	1
OO1	3 428	3 282	760	3 359	-1
OO2	4 639	4 376	1 075	3 359	-1
OO3	5 905	5 470	2 254	3 359	-1

fulfills the criterion in the training set is manipulated so that it does not fulfill the criterion in the classification set. The binary code for the examples is decided with the same criterion as for the training set.

3.8 Conclusions

In this chapter the data set that will be used to verify the classification methods in Chapter 4 and 5 described. The methods to design the different examples and the dimension criterion for deciding the binary code are presented. The detailed protection equipment in ARISTO, adjustment of spinning reserve, speed-droop, transient analysis and comparison to Nordel grid instructions improves the quality of the data set. The ARISTO simulator is used instead of a normal load flow solver, which improves the quality.

Chapter 4

Principal component analysis

In this chapter an introduction to Principal Component Analysis (PCA) will be given. PCA is one of the most popular unsupervised learning methods. The main reason is that the method reduces the dimension of the data set. With PCA the variables in the data set are projected to a lower dimensional space described by the principal components, see [14], [15] and [20]. Thus the characteristics of a multivariate process can hopefully be shown in lower dimensions. The projection of the data set on the principal components can give an unexpected relationship between the process and the variables.

Today in supervision and control of power systems the main problems are not to measure and collect data. The Supervisory Control And Data Acquisition (SCADA) system gives this possibility. The main problem is to identify examples which do not fulfill the $(n-1)$ criterion. Furthermore, it might not be necessary to use all measured variables from the power system for the identification. Which variables that can be reduced is an important question to investigate.

4.1 Principal component analysis

PCA was introduced by Pearson in 1901 [32]. With PCA the dimension of the data set is reduced from $\mathbf{R}^d$ to $\mathbf{R}^n$, where $n \leq d$. The principal components are orthogonal in $\mathbf{R}^n$. The principal components are the directions in the data set with the highest variance, see Figure 4.1. An introduction to how the principal components are calculated will be given below.

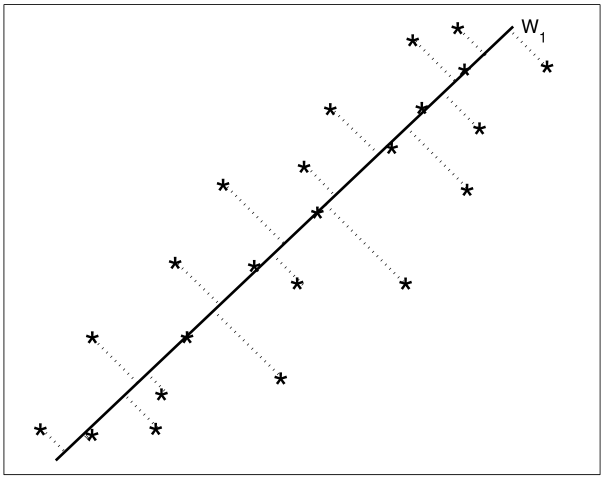


Figure 4.1: The first linear principal component of a data set. The line minimizes the total squared distance from each point to its orthogonal projection onto the line.

Denote the data set

$$\mathbf{X}_{l \times d} \ ,$$

where $\mathbf{X}$ is the data set stored as l examples with d variables. Let $\overline{\mathbf{x}} = (\overline{x}_1, \ldots, \overline{x}_d)$ be the mean value for all variables and,

$$\hat{\mathbf{X}} = [\mathbf{x}_1 - \overline{x}_1, \ldots, \mathbf{x}_d - \overline{x}_d] \ ,$$

where $\mathbf{x}_i, \ i = 1, \ldots, d$ are the columns in $\mathbf{X}$.

The covariance matrix of $\hat{\mathbf{X}}$, see [25], is

$$\Sigma = \frac{1}{l-1} \hat{\mathbf{X}}^\top \hat{\mathbf{X}} \ .$$

The principal components of $\hat{\mathbf{X}}$ can be calculated by solving the eigenvalue problem,

$$\Sigma \Phi = \Lambda \Phi \ ,$$

where Φ is an orthogonal matrix and Λ is a diagonal matrix containing the eigenvalues. The eigenvectors to $\hat{\mathbf{X}}^\top\hat{\mathbf{X}}$, Φ can also be calculated with Singular Value Decomposition (SVD) of $\hat{\mathbf{X}}$, as

$$\hat{\mathbf{X}} = \mathbf{U}\mathbf{S}\mathbf{V}^\top,$$

where $\mathbf{U}_{l\times l}$ and $\mathbf{V}_{d\times d}$ are orthogonal matrices and $\mathbf{S}_{l\times d}$ is a diagonal matrix containing the singular values. The first principal component, the direction in the data set with largest variance captured with a linear function, is the eigenvector with the largest corresponding singular value. The SVD-factorization is usually defined with the singular values in descending order.

In [15] and [11] the calculation of principal components are denoted with

$$\hat{\mathbf{X}} = \mathbf{T}\mathbf{P} + \mathbf{E} \ , \tag{4.1}$$

where $\mathbf{T}_{l\times n}$ are called score vectors, $\mathbf{P}_{d\times n}$ are called loading vectors, n is the number of principal components and $\mathbf{E}$ is the residual or variance in the data set not described by the principal components. $\mathbf{T}$ and $\mathbf{P}$ are calculated as,

$$\begin{aligned} I &= \{i_1, i_2, \ldots, i_d\} \\ J &= \{i_1, i_2, \ldots, i_n\} \\ \mathbf{P} &= \mathbf{V}^\top\{I, J\} \\ \mathbf{T} &= \hat{\mathbf{X}}\mathbf{P}^\top \ , \end{aligned}$$

where $\mathbf{V}^\top\{I, J\}$ are all the rows and the n first columns of $\mathbf{V}^\top$.

4.2 Number of principal components

How to choose the correct number of principal components is something that is difficult to answer. There are some values that can be calculated to guide the decision. If too many principal components are used the dimension of the data set is not reduced enough. On the other hand, if too few principal components are used, the residuals will be too large. For PCA the number of components is usually chosen by guidance from the eigenvalues. In Figure 4.2 the singular values of $\widehat{\mathbf{X}}$ are shown, it is sometimes denoted with **scree plot** [20].

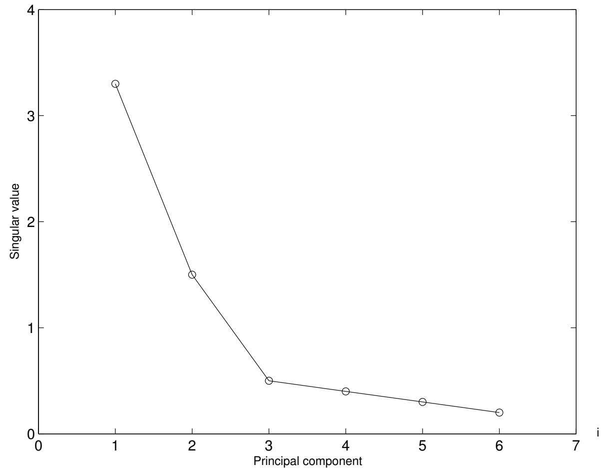


Figure 4.2: An example of a scree plot, showing the singular values.

There is a ”knee” in the plot when $i = 3$, therefore according to a popular rule, the number of components should be 3, see [20]. In [15] two factors are introduced for improving the choice of the number of principal components, H_{value} and H_{error}. They are based on the residuals variance

$$\begin{aligned} H_{value} &= \left(\sum_{i=n+1}^{d} \lambda_i \right) \left(\sum_{i=1}^{n} \frac{1}{\lambda_i} \right) \\ H_{error} &= \left(\sum_{i=n+1}^{d} \lambda_i \right) \left(1 + \sum_{i=1}^{n} \frac{1}{\lambda_i} \right) \quad , \end{aligned}$$

where λ_i are the eigenvalues to $\widehat{\mathbf{X}}^\top \widehat{\mathbf{X}}$. The number of principal components should be chosen so H_{value} and H_{error} are minimized. Both functions can be used but it is often more appropriate to use H_{error} because it is more distinct, compared to H_{value}.

4.3 Scaling

Before calculating the principal components it is useful to preprocess the data set. The different variables can have different range before preprocessing, e.g. the angle difference is between 0-90 degrees and the production of active power difference between 0-600 MW. In practice, the examples in the data set $\mathbf{x}_i$ are stored as rows in a matrix. The columns $\mathbf{z} = (z_1, \dots, z_l)$ of this matrix are scaled to a predefined interval. There are different methods to scale the variables in the examples e.g. mean-centering and univariate scaling, see [11].

$$\begin{aligned}
\mathbf{z}_{univariate} &= \frac{\mathbf{z} - m}{\sigma}, \text{ where} \\
m &= \frac{1}{l}\sum_{i=1}^{l} z_i \\
\sigma &= \sqrt{\sum_{i=1}^{l}\frac{(z_i - m)^2}{l-1}} \ .
\end{aligned} \tag{4.2}$$

4.3.1 Scaling to keep the physical properties

To keep the physical properties of the system a new scaling method is introduced, by scaling the variables group-wise. First, the voltage V and I are calculated in p.u. [13] as

$$V_{p.u.} = \frac{V}{V_{base}} \text{ and } I_{p.u.} = \frac{I}{I_{base}} \ ,$$

with a common system base S_{base}. Which results in,

$$I_{base} = \frac{S_{base}}{\sqrt{3}V_{base}} \ .$$

Then the variables are divided into groups, e.g. busbar V, load P, line I and angle differences. The groups are scaled with the univariate function (4.2). Instead of dividing with the standard deviation in every column, the columns are divided with the largest absolute value found in groups of variables.

$$\begin{aligned}
\mathbf{z}_{group} &= \frac{\mathbf{z} - m}{Max}, \text{ where} \\
Max &= \text{The largest absolute value in a group of variables} \ .
\end{aligned} \tag{4.3}$$

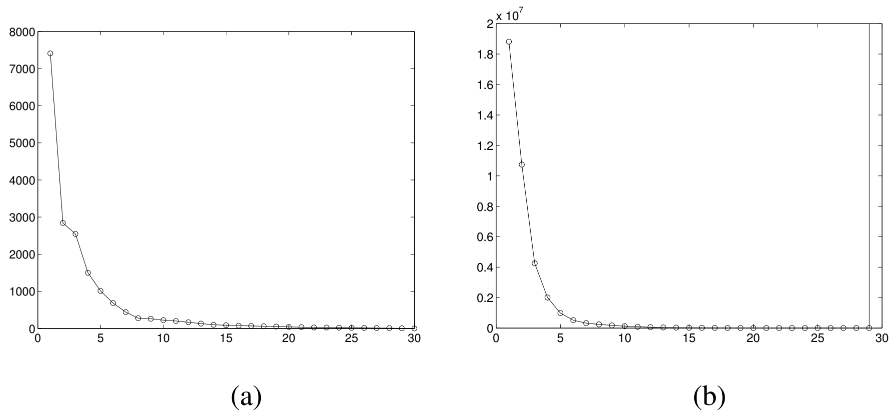


Figure 4.3: (a) The scree plot for the first 30 singular values to the training set.
(b) The 30 first numbers of H_{error} for the training set.

4.4 Classifying power system stability with PCA

The principal components are calculated for the training set, see Section 3.6. According to Figure 4.3 (a), the number of principal components should be between six or seven if the ”knee” condition is used. In Figure 4.3 (b) the number of principal components should be 29, if H_{error} should be minimized. The purpose with principal components, to reduce the dimension, is not fulfilled according to H_{error}. In Figure 4.4 to 4.7 the score plots for the five first principal components are presented. The two different scaling methods from Section 4.3 are used. It is clearly shown in Figure 4.4 to 4.7 that it is impossible to linearly separate the examples in the training set with different binary code using these five principal components. Experiments have shown that 19 principal components are needed to linearly separate the examples.

4.5 Conclusions

In this chapter the possibility to use principal component analysis to classify the stability of the data set presented in Chapter 3 is analyzed. Due to the complexity of power systems, with the large amount of variables that describe the state of the system, it is not possible to use PCA to separate the

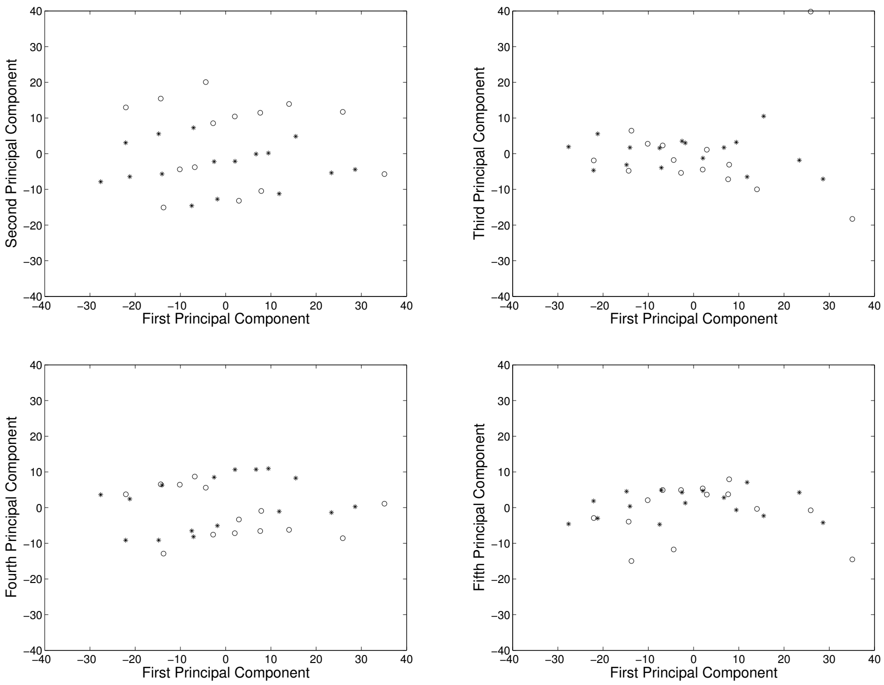


Figure 4.4: Score plot for the 5 first principal components. The examples are marked with "o" or "*", where "o" has the binary code -1 and "*" has the binary code 1. The data set are scaled univariate, see (4.2) .

examples in the training set with different binary code. A method for separation in higher dimensions must be chosen to succeed with the classifications.

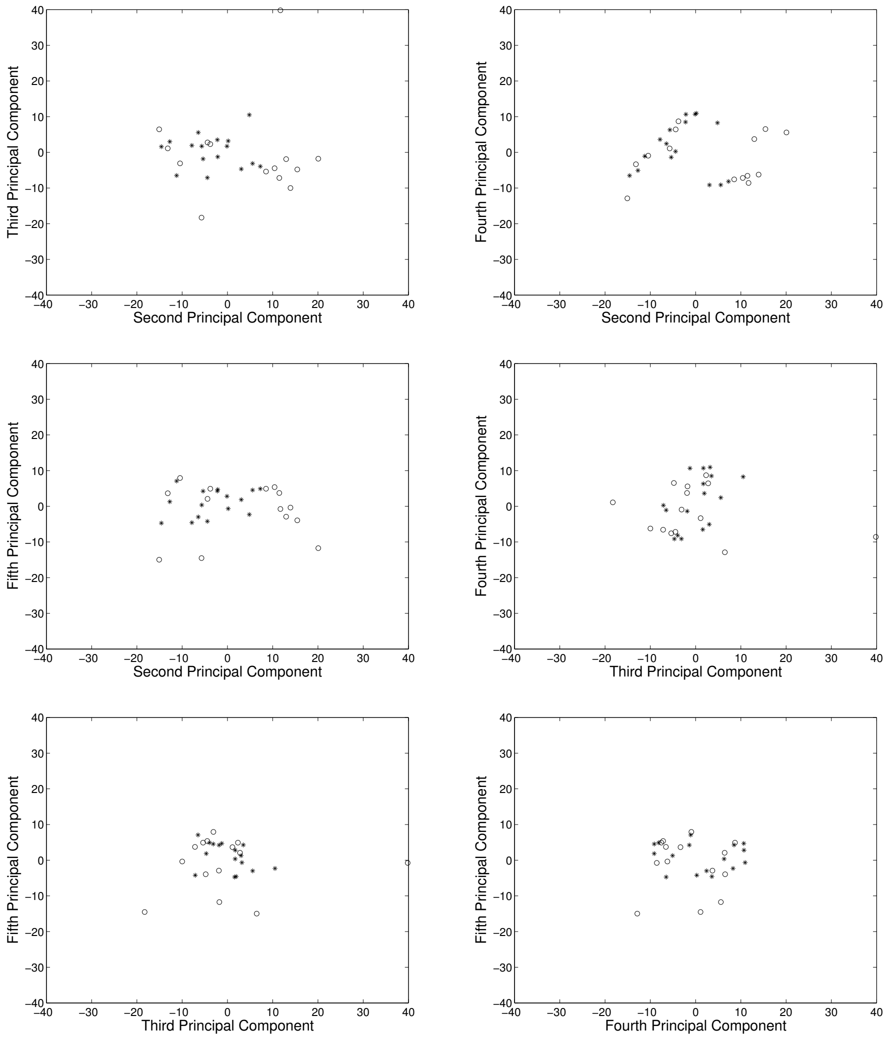


Figure 4.5: Score plot for the 5 first principal components. The examples are marked with ”o” or ”*”, where ”o” has the binary code -1 and ”*” has the binary code 1. The data set are scaled univariate, see (4.2) .

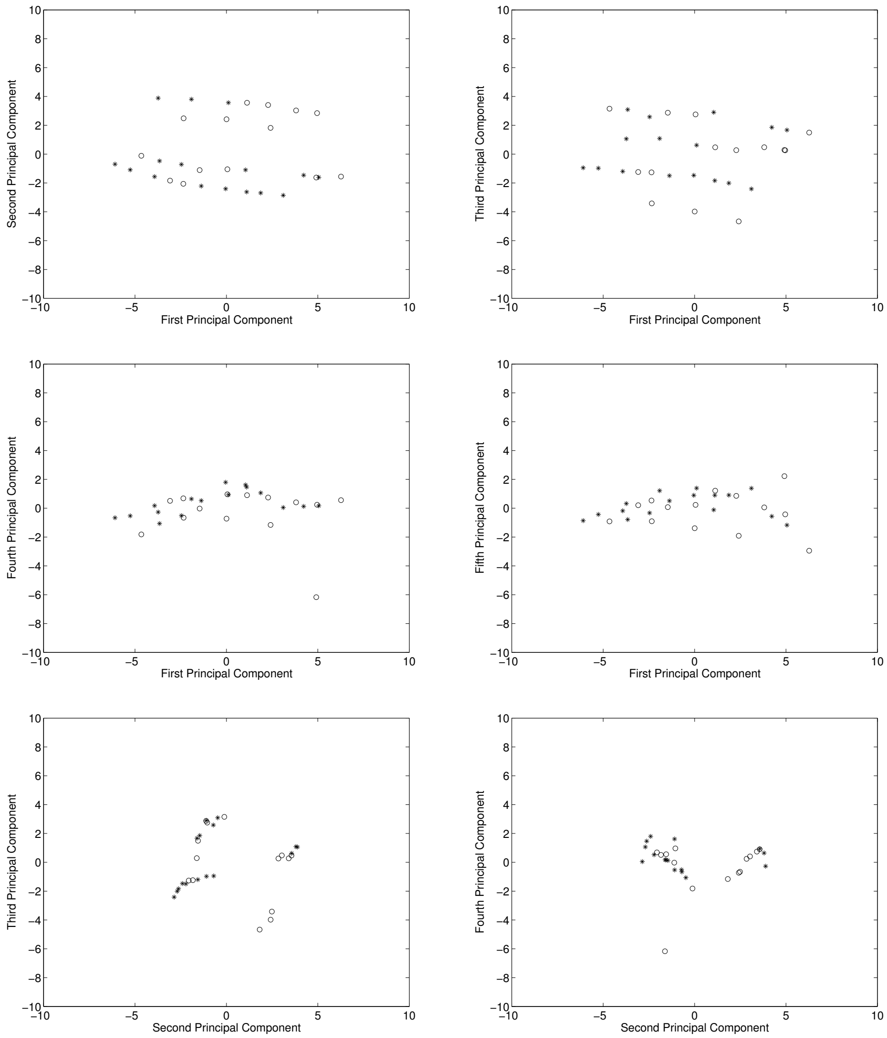


Figure 4.6: Score plot for the 5 first principal components. The examples are marked with "o" or "*", where "o" has the binary code -1 and "*" has the binary code 1. The data set are scaled group-wise, see (4.3) .

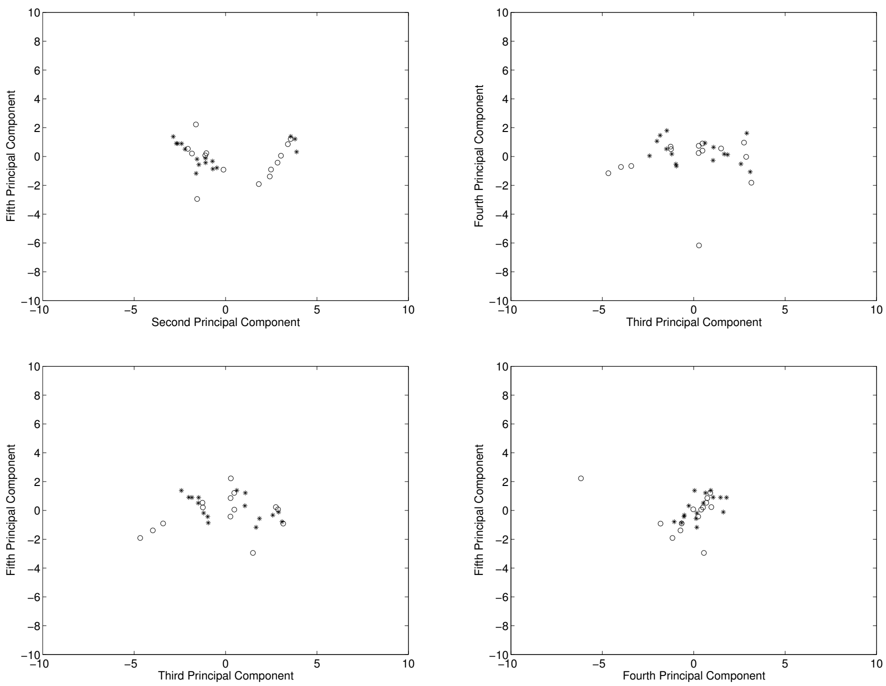


Figure 4.7: Score plot for the 5 first principal components. The examples are marked with "o" or "*", where "o" has the binary code -1 and "*" has the binary code 1. The data set are scaled group-wise, see (4.3) .

Chapter 5

Support vector machines

This chapter is an introduction to Support Vector Machines (SVM), a new generation learning system based on recent advances in statistical learning theory. Due to the large amount of variables that describe a state and the difficulties to estimate the stability of a state SVM is used as a classification tool. The aim is to use support vectors for calculating a hyperplane in high dimensions. This hyperplane separates the data with the maximum distance between the classes. The separation can be both linear and non-linear. The theory is applied to classification of power system stability. SVM have been applied to real-world applications such as text categorizing, hand-written character recognition, image classification and biosequence analysis.

5.1 Linear support vector machines

SVM were introduced by Vapnik in 1992 [36] and has received considerable attention recently. The objective is to create a learning machine that responds with a binary output to the test data. If the system sustains the dimension criterion it is marked by +1 otherwise it is marked -1. This designation is here after called the binary code. In order to train the machine, the training data must properly describe the variations of the system. It must be considered that if the training data do not show enough variability, the SVM cannot be successfully applied.

5.1.1 Maximal margin optimization

Denote the training data

$$\{\mathbf{x}_i \, , y_i\}, \quad i = 1, \ldots, l \ ,$$

where $\mathbf{x}_i \in \mathbf{R}^d$ are the input measurements and $y_i \in \{-1, 1\}$ are the corresponding binary responses. Suppose that there exists a hyperplane,

$$\mathbf{x} \cdot \mathbf{w} + b = 0 \ ,$$

with parameters $(\mathbf{w}, b)$, separating positive from negative examples. Let d+ and d- denote the shortest distance from the hyperplane to the closest positive/negative point respectively. The goal of the training is to find the hyperplane which best separates the training data. For a linear SVM this results in a convex optimization problem, which has a global solution. The hyperplane is therefore the plane in $\mathbf{R}^d$, with the largest d+ and d-. This is a useful property that other classification methods e.g. neural networks [33] do not have.

The constraints that the hyperplane must separate the points,

$$\mathbf{x}_i \cdot \mathbf{w} + b \geq 0, \quad \text{for } y_i = +1$$

$$\mathbf{x}_i \cdot \mathbf{w} + b \leq 0, \quad \text{for } y_i = -1$$

may be combined into

$$y_i(\mathbf{x}_i \cdot \mathbf{w} + b) \geq 0 \ .$$

The points which lie on the hyperplane $H_1 : \mathbf{x}_i \cdot \mathbf{w} + b = -1$ have the perpendicular distance from the origin $|b - 1|/\|\mathbf{w}\|$, and the points on the hyperplane $H_2 : \mathbf{x}_i \cdot \mathbf{w} + b = 1$ have the perpendicular distance from the origin $|b + 1|/\|\mathbf{w}\|$. Hence the margin is $1/||\mathbf{w}||$, see Figure 5.1.

The optimization problem

$$\begin{cases} \text{Minimize} & ||\mathbf{w}||^2 \\ \text{under the constraint} & d_i = y_i(\mathbf{x} \cdot \mathbf{w} + b) \geq 1 \ , \end{cases} \tag{5.1}$$

can be solved, by the following transformation,

$$L(\mathbf{w}, b, \mathbf{a}) = \frac{1}{2}||\mathbf{w}||^2 - \sum_{i=1}^{l} a_i y_i(\mathbf{x}_i \cdot \mathbf{w} + b) + \sum_{i=1}^{l} a_i \ , \tag{5.2}$$

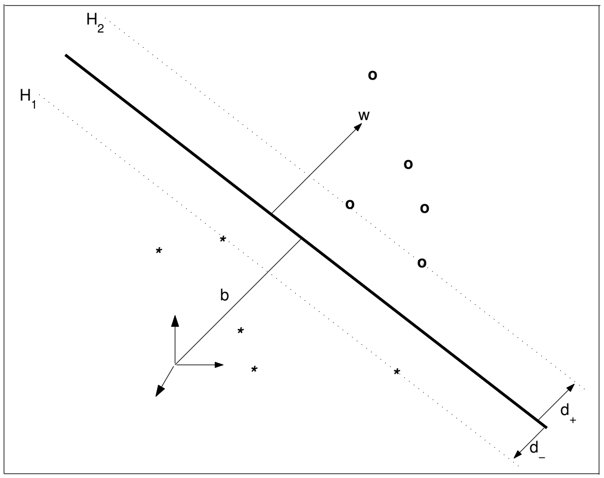


Figure 5.1: Linear separating hyperplane.

where $\mathbf{a} = (a_1, \ldots, a_l)$ are the Lagrange multipliers. The function is minimized with respect to $\mathbf{w}$ and b. The constraints from the gradient of the optimization problem with respect to $\mathbf{w}$ and b are

$$\begin{cases} \dfrac{\partial L(\mathbf{w}, b, \mathbf{a})}{\partial \mathbf{w}} & = & \mathbf{w} - \displaystyle\sum_{i=1}^{l} a_i y_i \mathbf{x}_i = \mathbf{0} \\ \dfrac{\partial L(\mathbf{w}, b, \mathbf{a})}{\partial b} & = & \displaystyle\sum_{i=1}^{l} a_i y_i = 0 \ . \end{cases}$$

Substituting these in (5.2) gives

$$L(\mathbf{w}, b, \mathbf{a}) = \sum_{i=1}^{l} a_i - \frac{1}{2} \sum_{i,j=1}^{l} a_i a_j y_i y_j \mathbf{x}_i \cdot \mathbf{x}_j \ . \tag{5.3}$$

The SVM maximizing (5.3) with respect to a_i for the constraints $\sum a_i y_i = 0$ and $a_i \geq 0$ gives the desired hyperplane. The optimization problem has a global solution, which gives the maximum margin between the two classes. The Karush-Kuhn-Tucker (KKT) [8] condition is used to determine b, the distance from the origin to the hyperplane, see Figure 5.1.

According to the KKT condition

$$a_i(y_i(\mathbf{x}_i \cdot \mathbf{w} + b) - 1) = 0 \ \ .$$

This can be used to determine b for any i for which $a_i > 0$. Every training point $\mathbf{x}_i$ has a corresponding a_i. The points which correspond to $a_i > 0$ are called **support vectors**. These points lie on the hyperplanes, H_1 and H_2 and affect the decision of the machine.

The classification data can be classified with the hyperplane using the sign of the function

$$f(x) = \mathbf{x} \cdot \mathbf{w} + b = \sum_{i=1}^{l} a_i y_i \mathbf{x}_i \cdot \mathbf{x} + b \ \ . \tag{5.4}$$

5.1.2 Soft margin optimization

The maximal margin optimization calculates a hyperplane with a consistent hypothesis. The resulting distance between the two classes is always the largest possible distance. In some classification problems it is not desirable to use the maximal margin optimization. If an example in the training set consists of misleading data the hyperplane will be affected by this. Therefore soft margin optimization can be used to calculate a hyperplane, which permit some degree of incorrect separation. This hyperplane represents a simpler model than the possibly over-fitted hyperplane of the maximal margin classifier. The so called slack variables ξ are used to represent the incorrect separation, see Figure 5.2. The hyperplane is calculated from the following optimization problem, for details see [8].

2-Norm Soft Margin

Using the 2-norm for the slack variable, the optimization problem is

$$\begin{cases} \text{Minimize} & ||\mathbf{w}||^2 + C\sum_{i=1}^{l} \xi_i^2 \\ \text{under the constraint} & d_i = y_i(\mathbf{x} \cdot \mathbf{w} + b) \geq 1 - \xi_i \ \ , \end{cases} \tag{5.5}$$

where $\xi = (\xi_1, \ldots, \xi_l)$ are the slack variables. The variable C must be chosen from a wide range of values. By using cross validation, leave-one-out

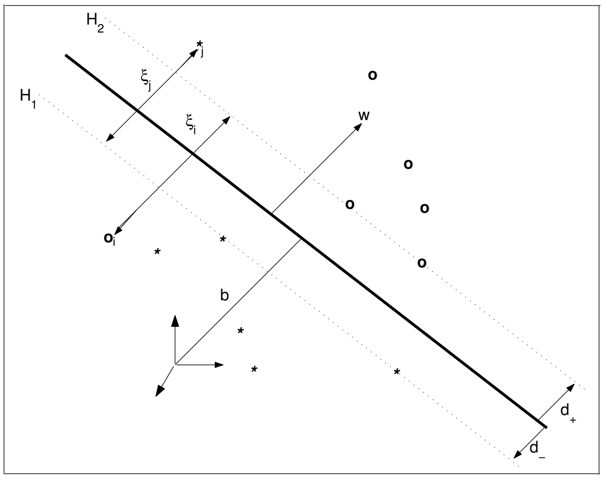


Figure 5.2: Linear separating hyperplane, with slack variables.

test, or a separate validation test set the value of the parameter is decided to obtain optimal performance. In this thesis C is set to 1000. The optimization test has not been performed to decide the value of C, due to the limited number of training examples. For implementation in power system operation the performance will be improved if C has the optimal value. The corresponding Lagrangian for the 2-norm soft margin optimization problem is

$$L(\mathbf{w}, b, \xi, \mathbf{a}) = \frac{1}{2}||\mathbf{w}||^2 + \frac{C}{2}\sum_{i=1}^{l}\xi_i^2 - \sum_{i=1}^{l} a_i y_i(\mathbf{x}_i \cdot \mathbf{w} + b) + \sum_{i=1}^{l} a_i(1-\xi_i) \quad , \tag{5.6}$$

where $\mathbf{a} = (a_1, \ldots, a_l)$ are the Lagrange multipliers. The constraints from the gradient of the optimization problem with the respect to $\mathbf{w}$, ξ and b are

$$\begin{cases} \dfrac{\partial L(\mathbf{w}, b, \xi, \mathbf{a})}{\partial \mathbf{w}} &= \mathbf{w} - \displaystyle\sum_{i=1}^{l} a_i y_i \mathbf{x}_i = \mathbf{0} \\ \dfrac{\partial L(\mathbf{w}, b, \xi, \mathbf{a})}{\partial \xi} &= C\xi - \mathbf{a} = \mathbf{0} \\ \dfrac{\partial L(\mathbf{w}, b, \xi, \mathbf{a})}{\partial b} &= \displaystyle\sum_{i=1}^{l} a_i y_i = 0 \quad . \end{cases}$$

Substituting these in (5.6) gives

$$L(\mathbf{w}, b, \mathbf{a}) = \sum_{i=1}^{l} a_i - \frac{1}{2} \sum_{i,j=1}^{l} a_i a_j y_i y_j \mathbf{x}_i \cdot \mathbf{x}_j - \frac{1}{2C} \sum_{i=1}^{l} a_i^2 \ . \tag{5.7}$$

The SVM maximizing (5.7) with respect to a_i, for the constraints $\sum a_i y_i = 0$ and $a_i \geq 0$, gives the desired hyperplane.

According to the KKT condition b is determined with

$$a_i(y_i(\mathbf{x} \cdot \mathbf{w} + b) - 1 + \xi_i) = 0 \ ,$$

for any $a_i > 0$.

1-Norm Soft Margin

Using the 1-norm for the slack variable, the optimization problem is

$$\begin{cases} \text{Minimize} & ||\mathbf{w}||^2 + C \sum_{i=1}^{l} \xi_i \\ \text{under the constraint} & d_i = y_i(\mathbf{x} \cdot \mathbf{w} + b) \geq 1 - \xi_i \\ & \xi_i \geq 0 \ . \end{cases} \tag{5.8}$$

The 1-norm soft margin is refered to as "the Box Constraint", where $\xi_i < C$. The difference between 1-norm and 2-norm is that ξ_i is limited by C for 1-norm, and not by 2-norm. The corresponding Lagrangian for the 1-norm soft margin optimization problem is

$$L(\mathbf{w}, b, \xi, \mathbf{a}, \mathbf{r}) = \frac{1}{2}||\mathbf{w}||^2 + C \sum_{i=1}^{l} \xi_i - \sum_{i=1}^{l} a_i y_i (\mathbf{x}_i \cdot \mathbf{w} + b - 1 + \xi_i) + \sum_{i=1}^{l} r_i \xi_i \ , \tag{5.9}$$

where $\mathbf{a}$ and $\mathbf{r}$ are the Lagrange multipliers. The function is minimized with respect to $\mathbf{w}$, ξ. The constrains from the gradient of the optimization problem with the respect to $\mathbf{w}$, ξ and b are

$$\begin{cases} \dfrac{\partial L(\mathbf{w}, b, \xi, \mathbf{a}, \mathbf{r})}{\partial \mathbf{w}} & = & \mathbf{w} - \sum_{i=1}^{l} a_i y_i \mathbf{x}_i = \mathbf{0} \\ \dfrac{\partial L(\mathbf{w}, b, \xi, \mathbf{a}, \mathbf{r})}{\partial \xi_i} & = & C - a_i - r_i = 0 \\ \dfrac{\partial L(\mathbf{w}, b, \xi, \mathbf{a}, \mathbf{r})}{\partial b} & = & \sum_{i=1}^{l} a_i y_i = 0 \ . \end{cases}$$

Substituting these in (5.9) gives

$$L(\mathbf{w}, b, \mathbf{a}) = \sum_{i=1}^{l} a_i - \frac{1}{2} \sum_{i,j=1}^{l} a_i a_j y_i y_j \mathbf{x}_i \cdot \mathbf{x}_j \ \ , \tag{5.10}$$

which is identical with the maximal margin. The only difference is the determination if b with the KKT condition

$$\begin{cases} a_i(y_i(\mathbf{x} \cdot \mathbf{w} + b) - 1 + \xi_i) = 0 \\ \xi_i(a_i - C) = 0 \ \ , \end{cases}$$

for any $a_i > 0$.

5.2 Scaling

As in Chapter 4 the data set is preprocessed before calculation of the hyperplane. In practice, the examples in the data set $\mathbf{x}_i$ are stored as rows in a matrix. The columns $\mathbf{z} = (z_1, \ldots, z_l)$ of this matrix are scaled to a predefined interval with a linear scaling [17],

$$\begin{aligned} \mathbf{z}_{linear} &= \frac{2(\mathbf{z} - Min)}{Max - Min} - 1 \\ Min &= \text{The smallest value in every column} \\ Max &= \text{The largest value in every column.} \end{aligned} \tag{5.11}$$

5.2.1 Scaling to keep the physical properties

The scaling method that keeps the physical properties of the variables, which was introduced in Chapter 4, is also used to classify with SVM. It is here applied to the linear scaling. First, the voltage V and I are calculated in p.u. [13] as

$$V_{p.u.} = \frac{V}{V_{base}} \ \text{ and } \ I_{p.u.} = \frac{I}{I_{base}} \ \ ,$$

with a common system base S_{base}. Which results in,

$$I_{base} = \frac{S_{base}}{\sqrt{3} V_{base}} \ \ .$$

Then the variables are divided into groups e.g. busbar V, load P, line I and angle differences. The groups are scaled with the linear function (5.11). Instead of finding *Min* and *Max* in every column, the *Min* and *Max* values are found in groups of variables which are then scaled group-wise.

5.2.2 Scaling of unit stator current after capacity

The capacity of the power system to fulfil the $(n-1)$ criterion depends on the spinning reserve. Therefore this is used to improve the classification. Instead of using the unit stator currents in the data set the stator current is used to calculate the unit capacity to increase the production. When the stator current is used in the data set no consideration of the unit capacity is taken. If a unit is idle, the stator current is almost 0. The unit has a large capacity to prevent a stressed situation. When a unit is not connected to the power system the stator current is 0, but the unit has no capacity to prevent a stressed situation. Both units have almost the same stator current, but the capacity to prevent a stressed situation is totaly different. Instead of using $I_{p.u.}$, I_* is calculated as

$$I_* = \frac{I_{rated} - I}{I_{base}} \ ,$$

were I_{rated} is on machine base. Using I_* the influence of the hyperplane from a unit in nominal production and a disconnected unit are the same.

5.2.3 Variables

In total 653 variables are used in an example to describe the operating point of the power system, post-fault. The variables are

- P, Q and I for lines, (156)
- V, f for busbars, (64)
- P, Q and I for loads, (180)
- P, Q and I for transformers, (24)
- P, Q and I for units, (108)
- Q and I for shunts, (84)

- Angle differences, (37),

where the number of variables in each group are shown in parenthesis.

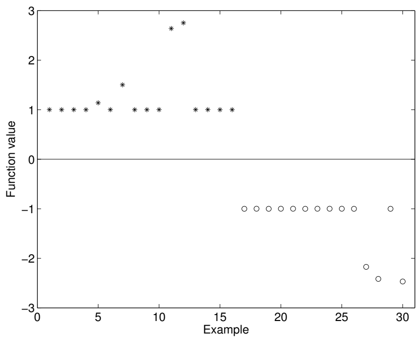
5.3 SVM training

Using the OSU SVM matlab toolbox [27] the support vectors are calculated. The number of support vectors is 20. The support vectors are used with (5.9) to calculate the hyperplane. The hyperplane is the optimal linear separation between the two classes. Using (5.4), the function value for each of the examples in the training set is decided.

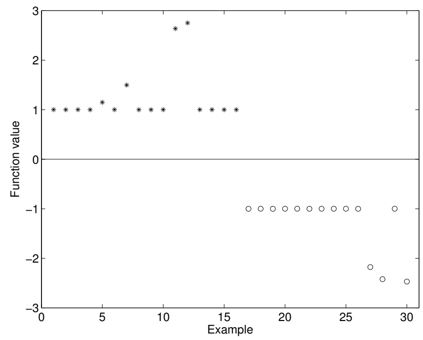
In Figure 5.3 (a)-(c) the function value for the 30 examples in the training set are marked by a ”o” or ”*” (”o” has the binary code -1 and ”*” +1). The examples which are used as support vectors have the function value ± 1, as in Figure 5.3 (a)-(c). It is possible to find a hyperplane which separates the two classes in the training set, scaled by 3 different methods, with SVM.

In Section 3.3 one of the traditional methods to indicate if the $(n-1)$ is fulfilled in the Swedish part of the Nordel system, the cut-sets methodology was introduced. Figure 5.4 shows that it is difficult to use the cut-sets to classify the stability of the examples in the training set. The examples with different binary code can not be separated with a certain threshold.

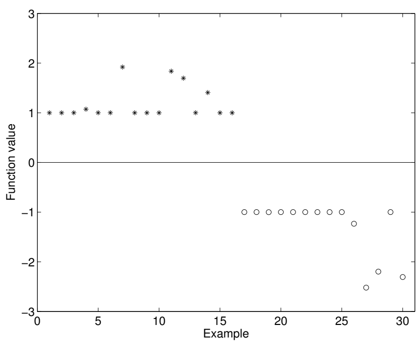
As the number of available examples is small, the performance of the training set is validated by leave-one-out cross validation, see [38]. Each time 29 of the 30 examples in the training set are used to train the SVM. The remaining example is then used for classification. This is done for all 30 examples, see Figure 5.3 (d). With the leave-one-out test 28 of 30 examples in the training set are correctly classified. The robustness in the training set is pretty good with a fault classification of 28/30.



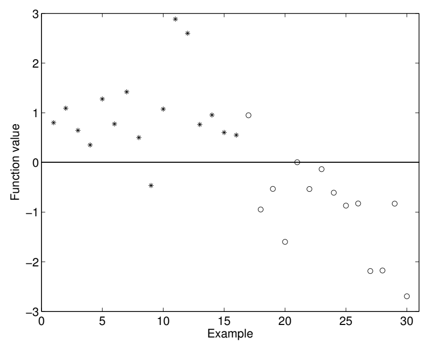
(a)



(b)

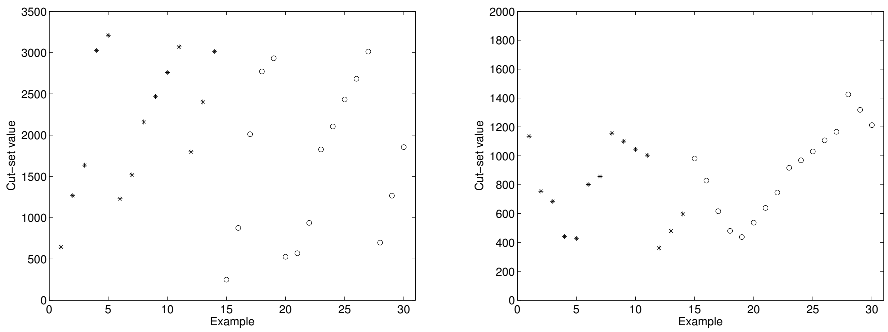


(c)



(d)

Figure 5.3: Linear SVM with soft-margin optimization, where $C = 1000$. Examples marked with ”o” has the binary code -1 and ”*” has the binary code +1. The function value, see (5.4), for the training set. All 633 feature variables are used, 20 variables are excluded due to variance equal to 0. (a) The variables are scaled linear group-wise, the unit stator current are also scaled after capacity. (b) The variables scaled linear group-wise. (c) The variables scaled linear separately. (d) Leave-one-out test. The function value, for the training set, 28 of 30 examples are correctly classified. The variables are scaled as in (b).



(a) (b)

Figure 5.4: (a) The sum of transferred active power through Snitt2, for the examples in the training set. (b) Snitt4, for the examples in the training set.

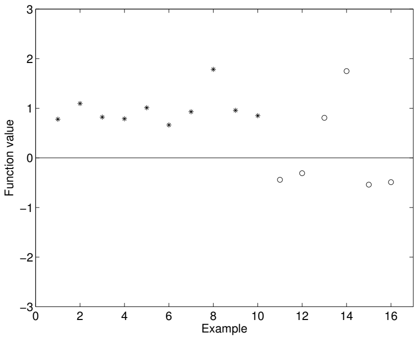
5.4 Classifying with a hyperplane

The hyperplane described in Section 5.3 is used to classify the examples in the classification set. With Equation (5.4) the function value is decided. The classification set is scaled with the three methods, which corresponds to the examples in the training set. In Figure 5.5 (a)-(c) the function value, for the examples in the classification set are marked as in Figure 5.3.

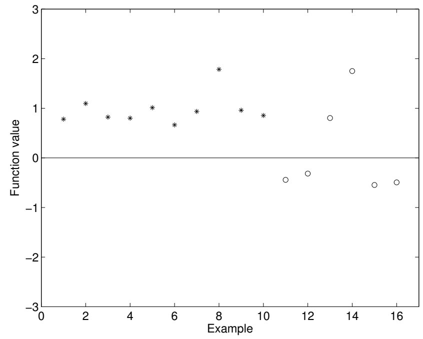
The four examples that are misclassified are classified to fulfill the $(n-1)$ criterion, but they do not. The examples are originally represented in the training set and there they fulfill $(n-1)$, but in the classification set the active power load is increased on a common system base so they do not. The faults which make the examples to not fulfill the $(n-1)$ are short circuit of busbar CT11_ B400, CT21_ A400 and FT50_ A400, see Figure 3.2. These faults also cause blackouts for the examples in the training set so the faults are represented both in the training and classification set.

The misclassification depends on that the training set has too few examples, or that the variability of the examples in the training set does not describe the examples in the classification set. If the training set does not describe the examples in the classification set it is impossible to make a correct classification. The number of variables used to describe the classification problem can also be too large, thus the hyperplane is over-fitted, see Figure 5.6. The hyperplane is over-fitted if the hyperplane is adjusted to the training set instead of the classification problem, which hopefully is described by the training set. The linear SVM decreases the risk for overfitting, compared to using a nonlinear function. There is a need to investigate if the classification can be improved with fewer variables.

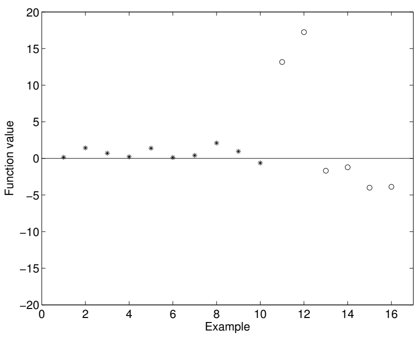
As with the examples in the training set the cut-set methodology is not successful to use to classify the examples in the classification set, see Figure 5.7. The examples with different binary code can not be separated with a threshold.



(a)



(b)



(c)

Figure 5.5: Linear SVM classification with soft-margin optimization, where $C = 1000$. Examples marked with ”o” has the binary code -1 and ”*” has the binary code +1. (a) The function value, for the classification set. All 633 feature variables are used and scaled linear group-wise, the unit stator current are also scaled after capacity . (b) The function value, for the classification set. All 633 feature variables, scaled linear group-wise. (c) The function value, for the training set. All 633 features variables scaled linear separately.

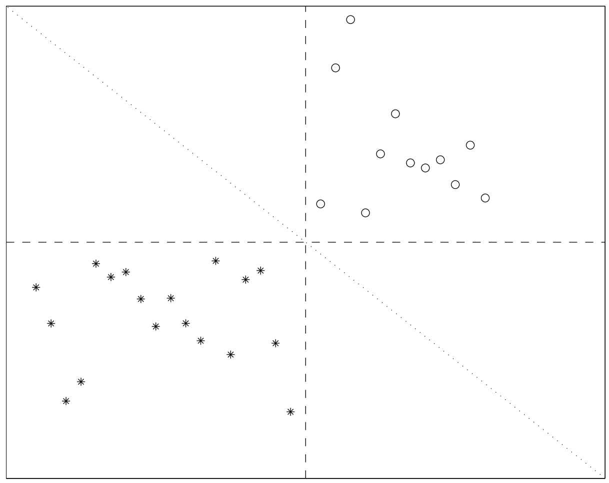
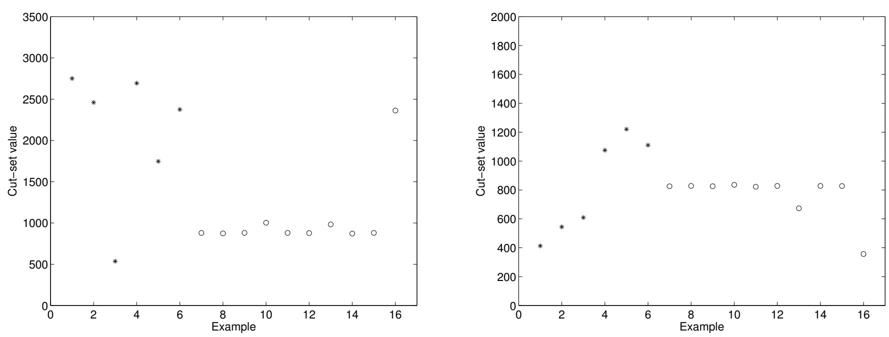


Figure 5.6: Illustration of overfitting classification.



(a) (b)

Figure 5.7: (a) Snitt2, for the classification set. (b) Snitt4, for the classification set.

5.5 Reducing variable sets

There is a need to analyze if the original data set consists of variables which cause over-fitting problems. The problem is, given the original data set in $\mathbf{R}^d$, what is the best subset of feature variables in $\mathbf{R}^c$, where c<d? It is not possible to find an optimal solution for this problem. First of all, we do not know the dimension of $\mathbf{R}^c$ and if we knew it, the number of possible subsets is

$$n_c = \frac{d!}{(d-c)!c!}$$

which is very large when d is 650.

Therefore, a new method to reduce the dimension of the data set is introduced, thus reducing the dimension of the data set without removing variables which are important for the separation. First, all variables in the training data with zero variance are removed. Then a method for further reduction is introduced. To reduce the variables the whole training data is scaled according to (5.11), and a hyperplane is calculated with SVM. By guidance from the normal $\mathbf{w}$ of the hyperplane, the dimension of the training set is reduced.

The unit normal of the hyperplane $\mathbf{w}_n$ is calculated, as

$$\mathbf{w}_n = \frac{\mathbf{w}}{||\mathbf{w}||} \ . \tag{5.12}$$

The absolute value of the j'th component in $\mathbf{w}_n$ indicates the contribution of the j'th variable to the hyperplane, see Figure 5.8. All components lower than a certain threshold, heuristically found, are removed. This procedure is important to analyze, so that over-fitting problems can be avoided.

The reduction method is applied on the data set, where the variables are reduced by guidance from the normal to the hyperplane. After the reduction a new hyperplane is calculated with the new variable setup. Then the function value for the examples in the classification set is calculated.

Only 2 of 16 are misclassified, both with the original variable setup and when the number of variables is reduced from 633 to 70, see Figure 5.9. The variables that are used to describe the classification problem after reduction are,

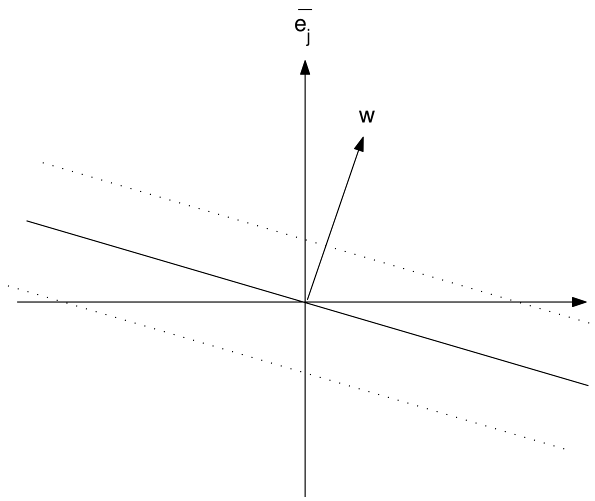


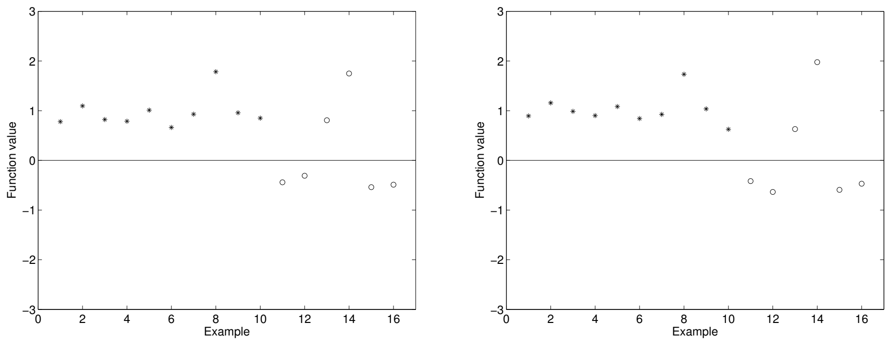
Figure 5.8: Reducing variables by guidance form the normal of the hyper-plane.

- P, Q and I for lines, (0, 9, 16 of 156)
- V, f for busbars, (5, 0 of 64)
- P, Q and I for loads, (0, 0, 0 of 180)
- P, Q and I for transformers, (0, 0, 0 of 24)
- P, Q and I for units, (3, 6, 1 of 108)
- Q and I for shunts, (11, 16 of 84)
- Angle differences, (2 of 37),

where the numbers of variables in each group is shown, ordered from left to right, in parenthesis.

5.5.1 Different variable setup

In [11], [14] and [28] it is indicated that variable have different variations patterns. Consideration must be taken about the difference between variables.



(a) (b)

Figure 5.9: Linear SVM classification with soft-margin optimization, where $C = 1000$. Examples marked with ”o” has the binary code -1 and ”*” has the binary code +1. The variables are scaled linear group-wise, the unit stator current are also scaled after capacity. (a) All 633 feature variables. (b) The 70 variables which have the largest influence on the hyperplane.

The variables can be divided into two groups:

- Qualitative
- Quantitative

Qualitative variables are normally represented numerically by codes. The easiest representation is a single binary code as 0 or 1, e.g. circuit breakers status. In this study circuit breakers/disconectors are not used in the data set, but shunt capacitors and reactors are included in the data set and they are approximately qualitative.

Quantitative variables may assume any numerical value along a continuous scale. All variables in the data set are of this type except shunts. It is important to be aware of the difference between the two variable types.

There is a need to investigate if the number of variables that are used for classification can be reduced. Large data sets can cause over-fitting problems, it is also interesting to minimize the cost to measure and collect the data set. Therefore the different variable groups are used, one-at-a-time, to calculate the hyperplane with soft margin optimization. The variables are scaled group-wise, unit stator current is also scaled after capacity. In Figures 5.10-5.12 the results with the different variables are shown, some variables give a better hyperplane for separation than others. The load and frequency gives an unsatisfactory separation while line and voltage gives successful separation.

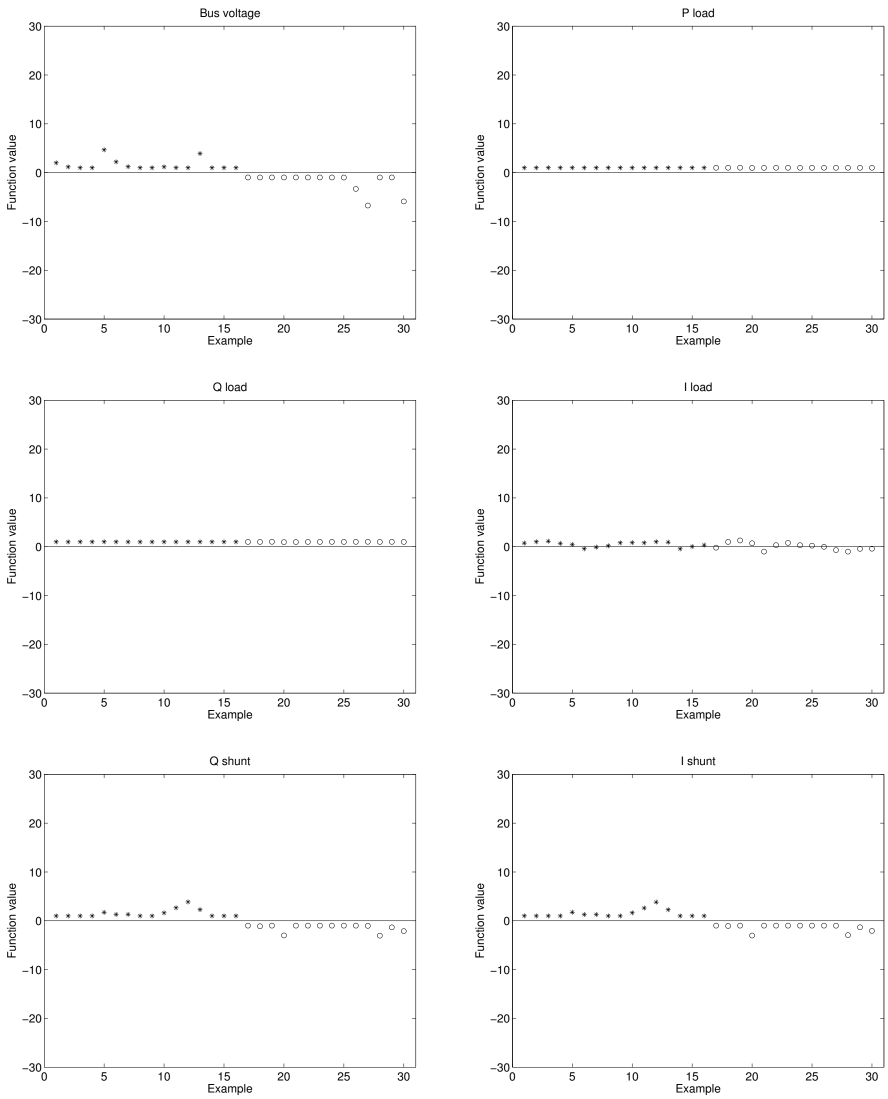


Figure 5.10: The function value for the SVM training based on the different variable groups in the training set. The soft-margin optimization is used to calculate a hyperplane with slack variables, where $C = 1000$, for further information see Section 5.1.2. The variables are scaled group-wise, where the unit stator current also are scaled after capacity.

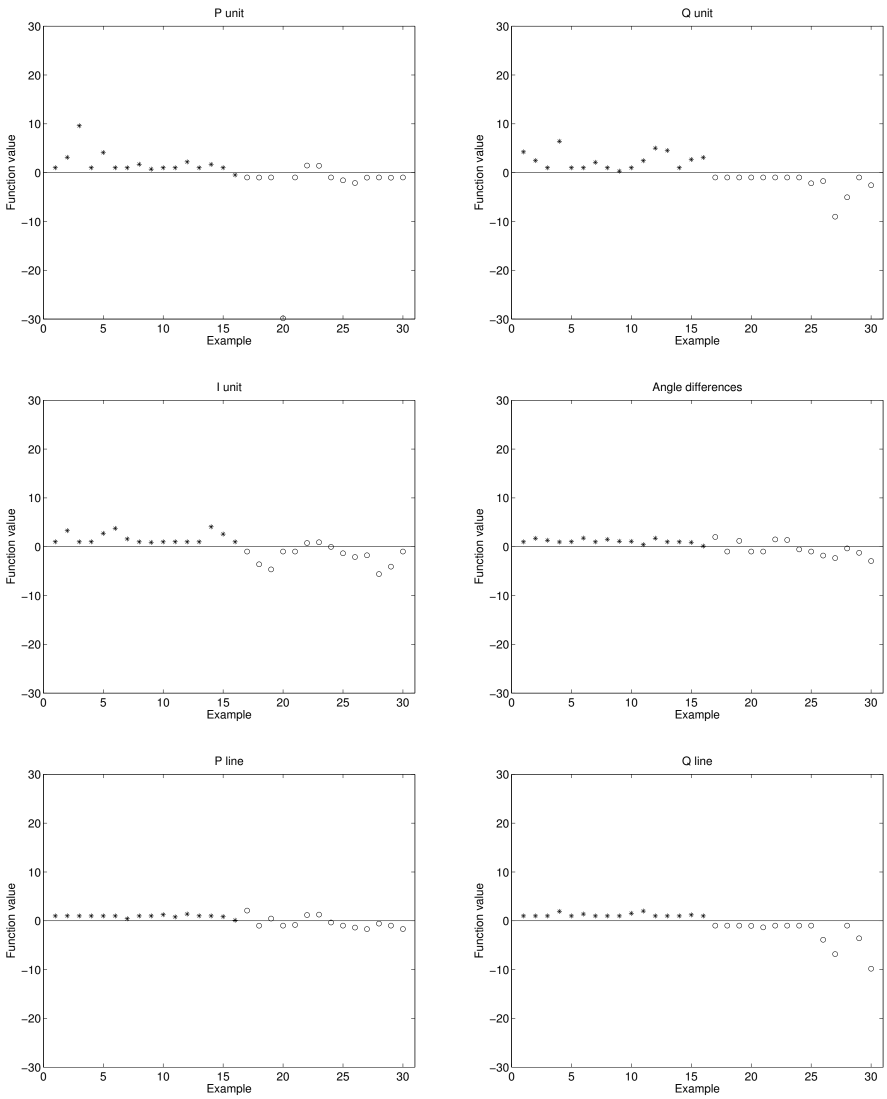


Figure 5.11: The function value for the SVM training based on the different variable groups in the training set. The soft-margin optimization is used to calculate a hyperplane with slack variables, where $C = 1000$, for further information see Section 5.1.2. The variables are scaled group-wise, where the unit stator current also are scaled after capacity.

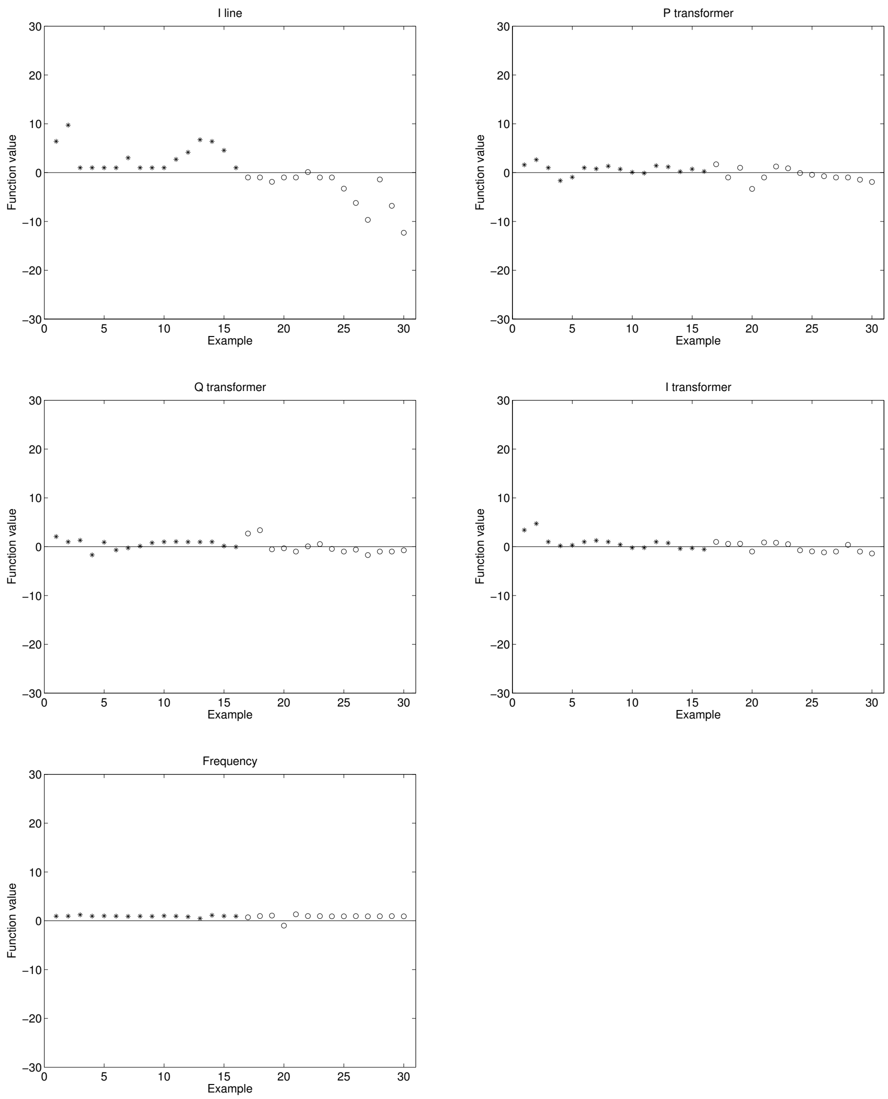


Figure 5.12: The function value for the SVM training based on the different variable groups in the training set. The soft-margin optimization is used to calculate a hyperplane with slack variables, where $C = 1000$, for further information see Section 5.1.2. The variables are scaled group-wise, where the unit stator current also are scaled after capacity.

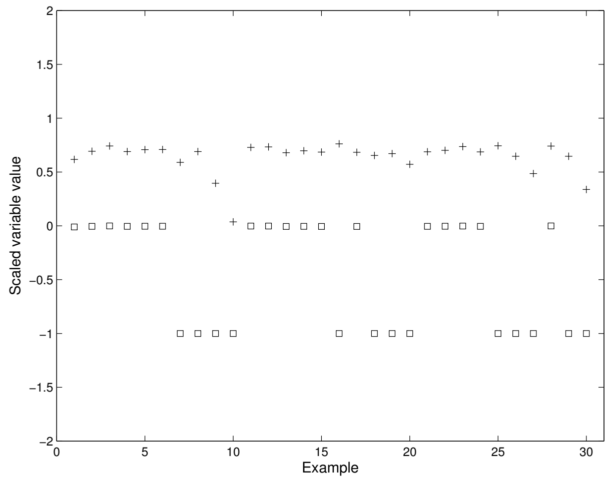


Figure 5.13: Two of the variables in the training set, scaled linear group-wise, CT22 X1 CurrentValue, qualitative, marked with "□" and CT22 A400 VoltageValue, quantitative, marked with "+".

It is possible to calculate a hyperplane that separates the examples in the training set with both P shunt and I shunt. Therefore, it should be natural to use the two variable classes for the classification problem. But P shunt and I shunt are qualitative and therefore they have too large impact on the hyperplane. The hyperplane will then be trained to classify between two different shunt patterns that can be identified in the training set. In Table 5.1 the 30 variables which have the largest impact on the hyperplane are listed, 19 of these are shunts. The difference between qualitative and quantitative variables are shown in Figure 5.13. It is well known that mixing qualitative and quantitative variables can give problems when building classifiers. There are methods for dealing with this, e.g. [21] and [38]. Usually these methods require large training sets.

Table 5.1: The 30 variables which have the largest influence on the hyperplane. The soft-margin optimization is used to calculate a hyperplane with slack variables, where $C = 1000$, for further information see Section 5.1.2. The variables are scaled group-wise, where the unit stator current also are scaled after capacity.

Object	$\frac{\mathbf{w}}{\|\|\mathbf{w}\|\|}$
CT22 X1 Shunt reactor (I)	0.4675
CT32 X3 Shunt reactor (I)	0.3942
CT22 X1 Shunt reactor (Q)	0.2635
FT41 EK1 Shunt capacitor (I)	0.2276
CT32 X3 Shunt reactor (Q)	0.2243
CT31 X3 Shunt reactor (I)	0.2124
FT42 X1 Shunt reactor (I)	0.1483
FT44 X1 Shunt reactor (I)	0.1419
FT50 X1 Shunt reactor (I)	0.1410
CT11 T2 Unit (Q)	0.1364
FT41 EK1 Shunt capacitor (Q)	0.1309
CT21 X1 Shunt reactor (I)	0.1225
CT31 X3 Shunt reactor (Q)	0.1207
CT11 X1 Shunt reactor (I)	0.1100
CT32 X1 Shunt reactor (I)	0.1075
CT22 A400 Busbar (V)	0.0871
CL1 Line (I) CT11	0.0867
FT42 X1 Shunt reactor (Q)	0.0837
FT44 X1 Shunt reactor (Q)	0.0810
FT44 T11 Unit (Q)	0.0799
FT50 X1 Shunt reactor (Q)	0.0773
FT51 G2 Unit (Q)	0.0752
FT50 X2 Shunt reactor (I)	0.0743
CT72 G2 Unit (P)	0.0733
CL8 Line (Q) CT22	0.0727
CT122 A130 Busbar (V)	0.0718
CT21 X1 Shunt reactor (Q)	0.0695
FL14 Line (I) FT51	0.0690
FL15 Line (I) FT51	0.0690
CL3 Line (Q) CT22	0.0671

Due to the different variable types, different variable setups are applied to classification with SVM. In Section 5.4 the data set was scaled with three different methods. All three are also used here because no method could be excluded from the results in Section 3. Five variable setups are compared,

1. Original variable setup, 653
2. Shunts removed, 569
3. All P and Q removed, 289
4. All P and Q, transformer and shunts removed, 257
5. All P and Q, frequency and shunts removed, 225

and the data set is scaled by three different methods,

1. Group-wise
2. Group-wise, unit stator current is also scaled by capacity
3. Separately.

The variable setup which gives the best separation are 2, 4 and 5.

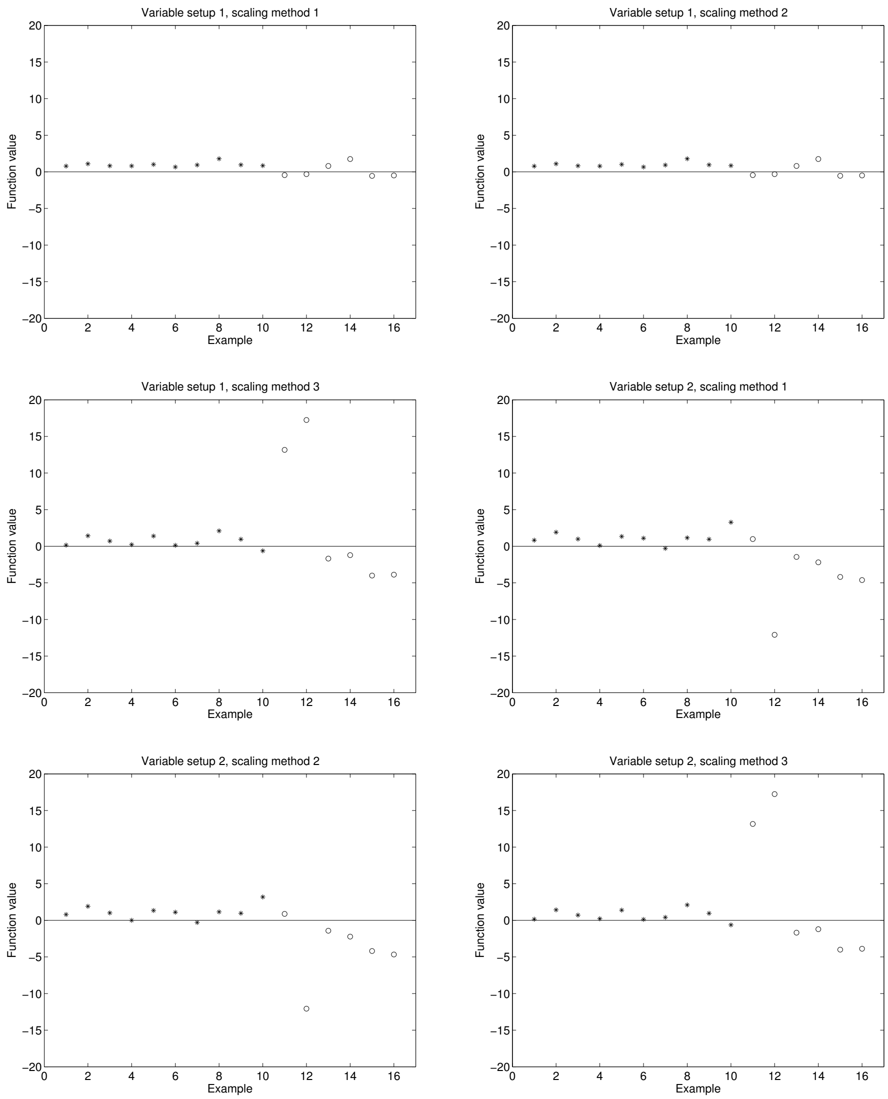


Figure 5.14: The function value for the SVM classification based on the different variable setups and different scaling methods. The soft-margin optimization is used to calculate a hyperplane with slack variables, where $C = 1000$.

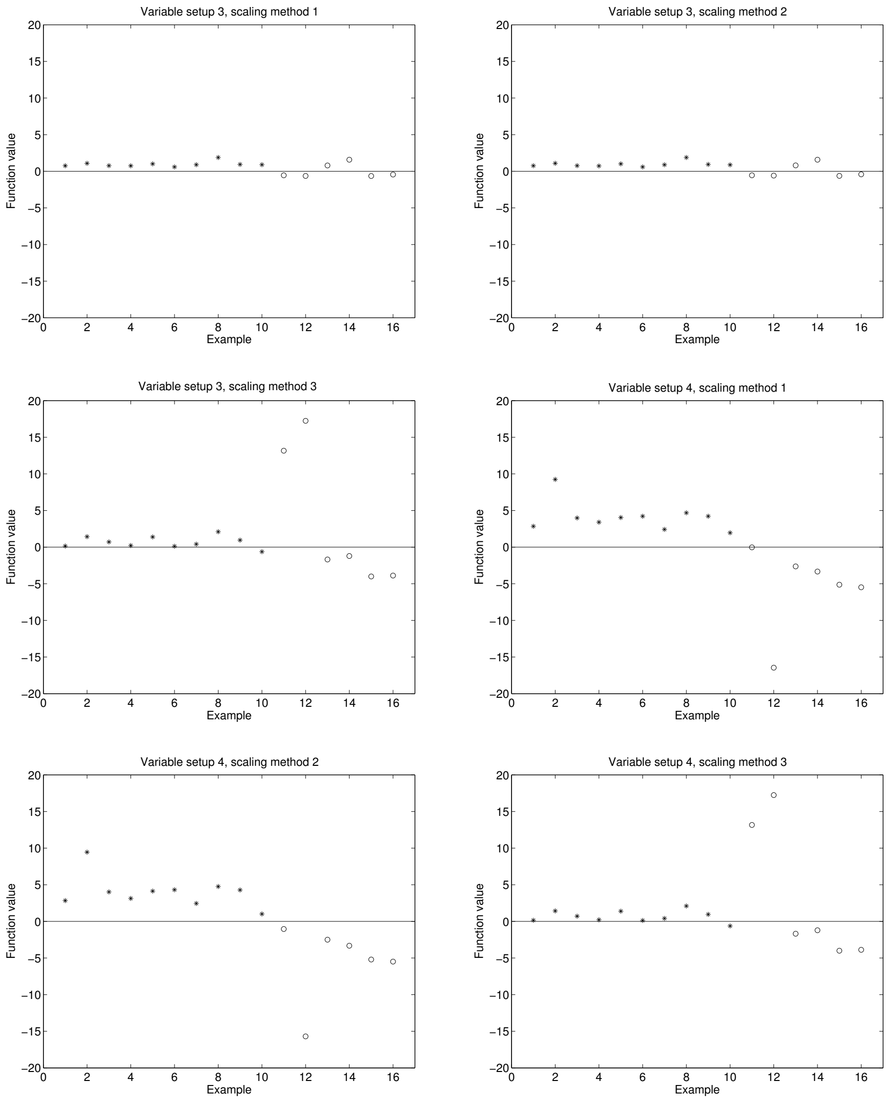


Figure 5.15: The function value for the SVM classification based on the different variable setups and different scaling methods. The soft-margin optimization is used to calculate a hyperplane with slack variables, where $C = 1000$.

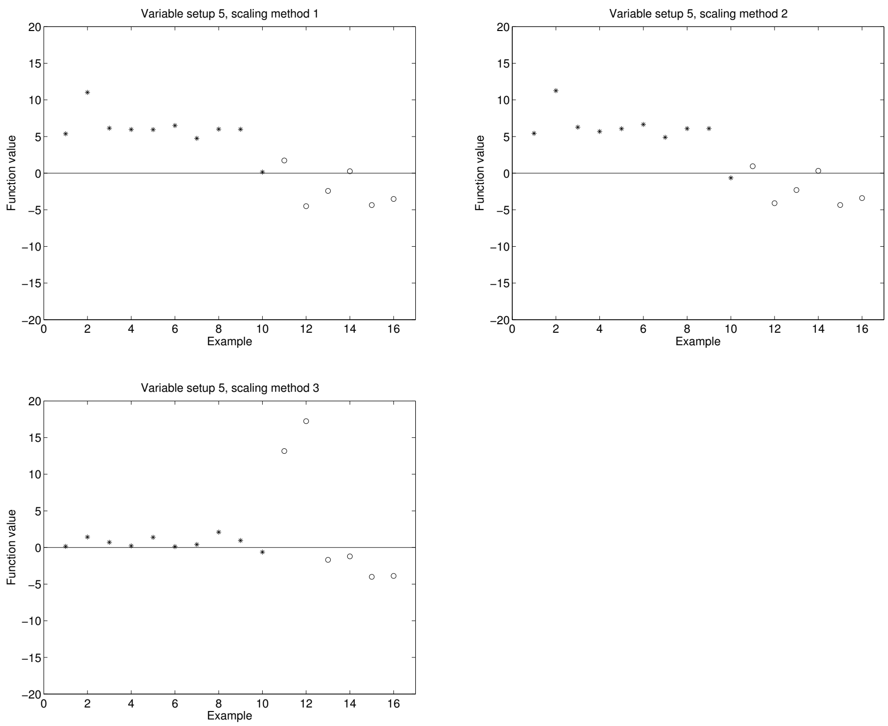


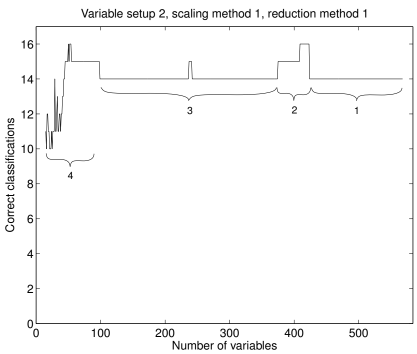
Figure 5.16: The function value for the SVM classification based on the different variable setups and different scaling methods. The soft-margin optimization is used to calculate a hyperplane with slack variables, where $C = 1000$.

The method to reduce variables by guidance from the normal to the hyperplane is applied to the data sets with the variable setups which give the best classification, according to Figure 5.14-5.16. The variable setups with shunts are excluded. The linear scaling method is not used. The reduction is illustrated with a new plot, **reduction plot**, which shows the number of correct classified examples in the classification set. The number of variables in the data set is reduced by two different reduction methods:

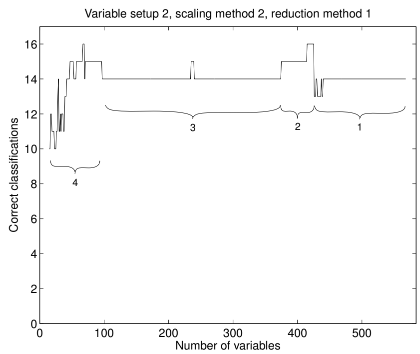
1. The first one by increasing the threshold so one variable is excluded at a time, see Figure 5.17-5.18.

2. The other is by excluding the variable which has the least influence on the hyperplane, the new data set is used to calculate the difference between the variables. Thereafter a new hyperplane is calculated and the process is repeated, see Figure 5.19-5.20.

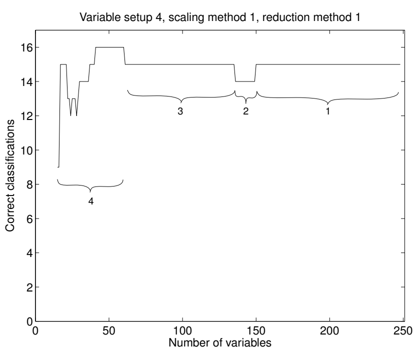
The variable groups that are dominating in the reduction process are presented in Table 5.2. The variable setups that have varying classification result are divided into subsets, see Figure 5.17-5.20. In these subsets the dominating variable groups are presented in Table 5.2. This is only a guideline of which variable groups that are excluded, there are other variable groups represented in the subsets. In Table 5.3 the 30 variables which have the largest impact on the hyperplane are listed. The variable setup where P, Q, transformer and shunt are reduced is used. These 30 variables are also pointed out in Figure 5.21.



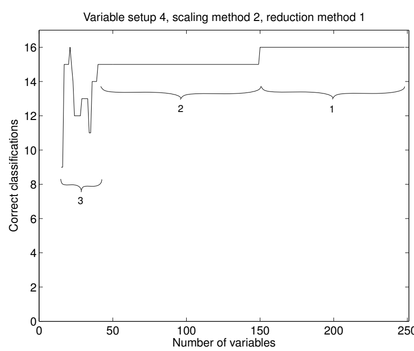
(a)



(b)

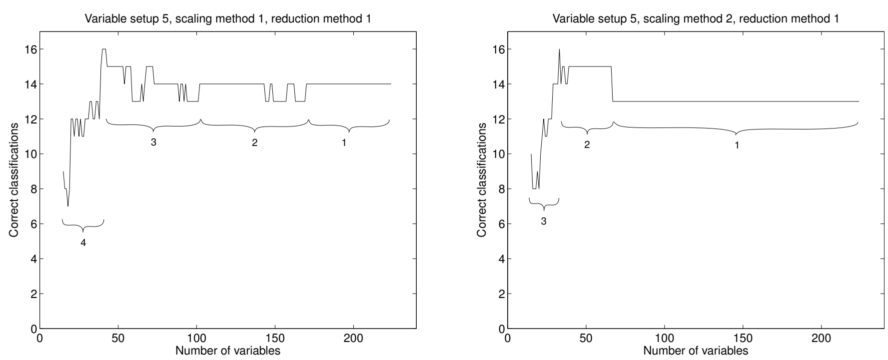


(c)



(d)

Figure 5.17: The reduction plot based on the different variable setups and different scaling methods. The x-axis show the numbers of variables that are used in the classification. The y-axis show the number of correct classified examples in the classification set. The soft-margin optimization is used to calculate a hyperplane with slack variables, where $C = 1000$.



(a) (b)

Figure 5.18: The reduction plot based on the different variable setups and different scaling methods. The x-axis show the numbers of variables that are used in the classification. The y-axis show the number of correct classified examples in the classification set. The soft-margin optimization is used to calculate a hyperplane with slack variables, where $C = 1000$.

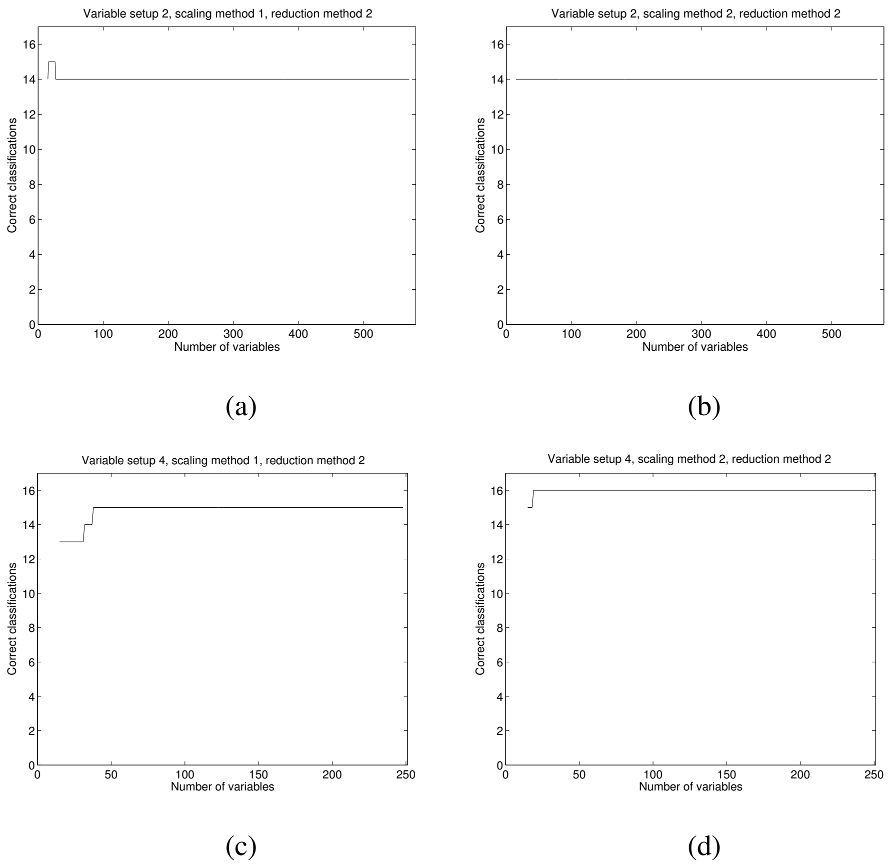
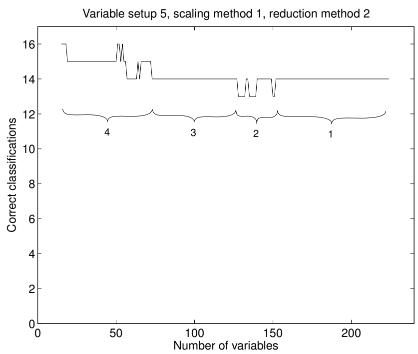
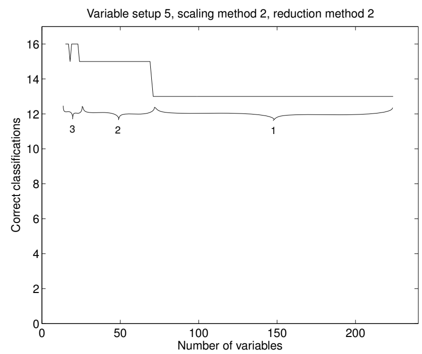


Figure 5.19: The reduction plot based on the different variable setups and different scaling methods. The x-axis show the numbers of variables that are used in the classification. The y-axis show the number of correct classified examples in the classification set. The soft-margin optimization is used to calculate a hyperplane with slack variables, where $C = 1000$.





(a) (b)

Figure 5.20: The reduction plot based on the different variable setups and different scaling methods. The x-axis show the numbers of variables that are used in the classification. The y-axis show the number of correct classified examples in the classification set. The soft-margin optimization is used to calculate a hyperplane with slack variables, where $C = 1000$.

Table 5.2: The dominating variable groups that are excluded in the reduction in Figure 5.17-5.20

	1	2	3	4
5.17 (a)	Load (P, Q)	Load (I) Angle difference	Mixture of all variables	Line (P, Q, I) Unit (P, Q I) Busbar (V)
5.17 (b)	Load (P, Q)	Load (I) Line (P)	Mixture of all variables	Line (P, Q, I) Unit (P, Q I) Busbar (V)
5.17 (c)	Load (I) Busbar (f)	Load (I) Unit (I)	Line (I) Angle difference	Line (I) Busbar (V) Angle difference
5.17 (d)	Load (I) Busbar (f)	Line (I) Unit (I) Angle difference	Busbar (V) Line (I)	
5.18 (a)	Load (I)	Line (I) Load (I) Angle difference	Line (I) Angle difference Busbar (V)	Line (I) Busbar (V) Unit (I)
5.18 (b)	Angle difference Line (I) Load (I)	Line (I) Busbar (V) Unit (I)	Line (I) Busbar (V)	
5.20 (a)	Load (I)	Load (I)	Angle difference Line (I) Unit (I)	Line (I) Busbar (V)
5.20 (b)	Mixture of all variables	Mixture of all variables	Busbar (V)	

Table 5.3: The 30 variables which have the largest influence on the hyperplane in Figure 5.17, variable setup 4, scaling method 2 and reduction method 1.

Object	$\frac{\mathbf{w}}{\|\mathbf{w}\|}$
CT11 A400 Busbar (V)	0.3245
FT44 A400 Busbar (V)	0.3212
CT22 A400 Busbar (V)	0.3100
CT111 A130 Busbar (V)	0.2843
CT122 A130 Busbar (V)	0.2272
FL2 Line (I) FT42	0.2249
FL6 Line (I) FT61	0.1983
FT41 A400 Busbar (V)	0.1708
CT21 A400 Busbar (V)	0.1678
CL1 Line (I) CT11	0.1656
CL5 Line (I) CT12	0.1560
CT12 A400 Busbar (V)	0.1474
CT72 G2 Unit (I)	0.1423
FT41 G1 Unit (I)	0.1260
CT72 G3 Unit (I)	0.1201
AT111 A130 Busbar (V)	0.1189
CL4 Line (I) CT12	0.1121
CT112 A130 Busbar (V)	0.1071
CL7 Line (I) CT71	0.1065
CL6 Line (I) CT71	0.1065
FL3 Line (I) FT43	0.1047
FT42 A400 Busbar (V)	0.0956
FT142 A130 Busbar (V)	0.0937
CT11 T1 Unit (I)	0.0913
FL10 Line (I) FT62	0.0889
CL17 Line (I) FT44	0.0878
FL15 Line (I) FT51	0.0855
FL14 Line (I) FT51	0.0855
FT50 A400 Angle FT62 A400	0.0852
RT132 G1 Unit (I)	0.0851

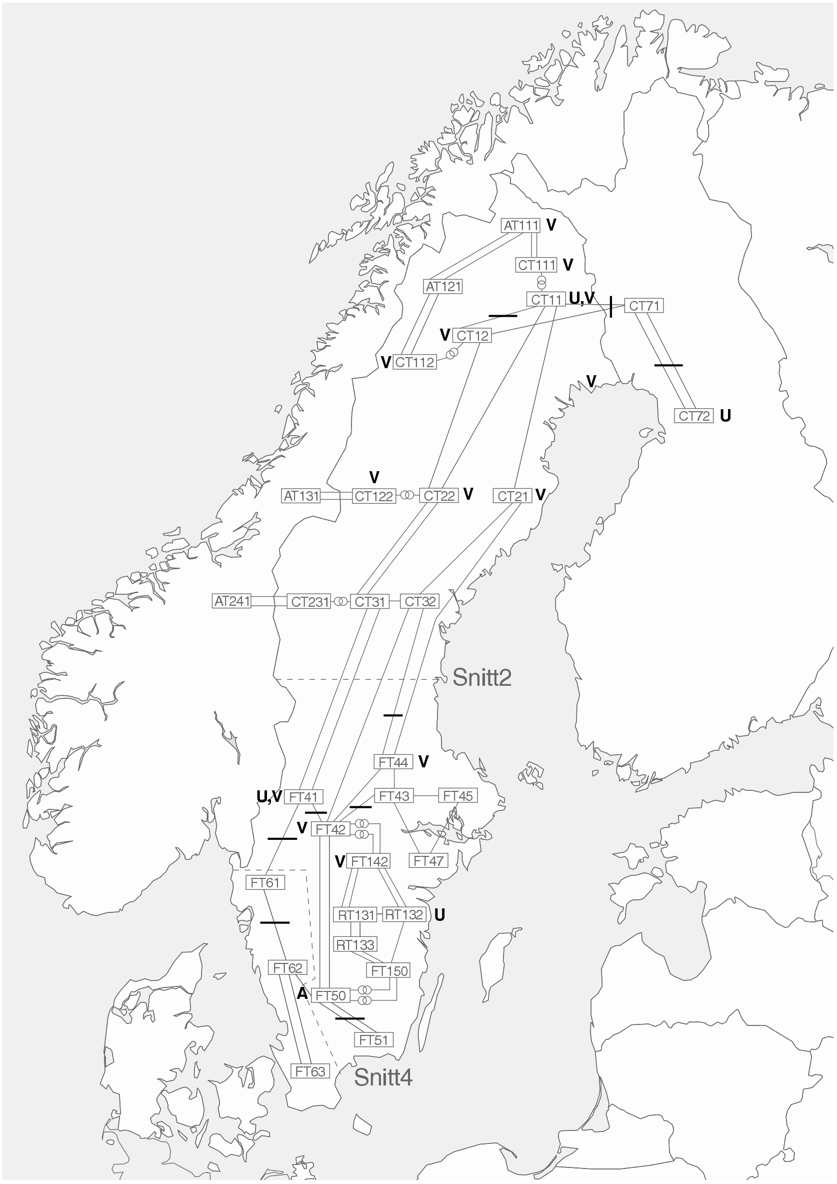


Figure 5.21: The NORDIC32 system with the variables from Table 5.3 marked. Bus voltage are marked with ”V”, unit stator current with ”U” and line current with ”—”.

5.6 Implementation

It is not a problem to calculate a hyperplane which can be used to classify the stability of the NORDEL power system. The calculation time for a training set which consists of 12000 examples and 1200 variables takes 3,3 s[1] and to decide the function value for an example with Equation 5.4 takes 0,04 s. The problem is to obtain a training set which is representative for the NORDEL system. The SCADA system gives the possibility to measure and collect the data. But for each of the examples, which can be measured and collected, it is impossible to decide the binary code. An example can obviously be modelled in a simulator, for example ARISTO, and the binary code can be decided for the model. But it is important to be aware of the difference between model and reality. The questions that have to be investigated before implementing this technique are:

- Is it possible to produce a training set with a model of the power system and use the training set to classify the stability of the real power system?
- Is it possible to classify an example, without having verified that the training set is representative for the classification example?
- What happens when a part of the data is misleading and contains measurement errors?

The author considers that it is possible to produce a training set in a simulator, it is mainly based on the methods which are used today in the state estimators. But the training set must be representative for the power system and the amount of time to produce such a training set will be very large. In [39] the system that have been used consists of 561 buses, 1000 lines and 61 generators. The number of examples in the data base are 11 000. The second question inspires for further research in this area, especially in finding methods to verify if it is possible to classify an example with the existing training set. The last question is not a large problem if the misleading data is identified. Robustness studies must be made of which variables that can be excluded. If the misleading data can be excluded the calculation time to remove these variables and calculate a new hyperplane is not a problem.

[1] On a 1,6 GHz Intel centrino PC with 512 MB RAM-memory and Windows XP using Matlab 6.5.

5.7 Results

Classification of the stability of the NORDIC32 system with a linear SVM shows that it is not enough to use the cut-sets methodology as indication of instability, which is mainly used in the Nordic power system today, see Figure 5.4 and 5.7. The examples with different binary code can not be separated using the power flow through the cut-sets.

In Section 5.3 it is shown that a hyperplane can be calculated that separates the examples with different binary code. But due to the misclassifications in the classification set, see Figure 5.5, different variable setups, scaling methods and a reduction method to improve the classification are used. In Figure 5.12 the different groups of variables are used one-at-a-time to calculate a hyperplane with soft-margin SVM. Some variable groups are more suitable to use for the classification than others. The difference between qualitative and quantitative variables must be noticed. In Table 5.1 it can be seen that 18 of the 30 variables that have the largest influence on the hyperplane are qualitative. It is reasonable to assume that the qualitative variables contribute to an over-fitted hyperplane which classifies the disconnection of shunt reactors and capacitors instead of the stability of the power system, see Figure 5.13. When the qualitative variables are excluded from the data set the classification is improved. The number of correct classification and the distance between the two classes is increased.

In this chapter a new method to scale variables which keeps the physical properties of the system by group-wise scaling is introduced, the unit stator currents are also scaled by capacity. Furthermore, a new method to reduce the number of feature variables is also introduced. In Figure 5.17-5.20 the different variable setups, scaling methods and reduction methods are validated by the reduction plot. It can be seen in Figure 5.17-5.20 that the group-wise scaling method with unit stator current scaled after capacity gives the best classification. The variables that should be reduced from the data set are, P, Q, transformer and shunts. In Table 5.2 -5.3 and Figure 5.21 it is shown that the variables that have most influence on the hyperplane are:

- Busbar voltage magnitude.
- Line current magnitude.
- Generator current magnitude.

5.8 Conclusions

In this chapter support vector machines are used to calculate a hyperplane to classify the stability of power systems. It is shown that it is possible to calculate a hyperplane which separates the two classes. The unique feature from SVM, which calculate a hyperplane with the maximum distance between the two classes, gives new possibilities to improve the classification of power system stability. The classification can be improved by using a variable setup where all P and Q, shunts and transformers are excluded. With the scaling method which keeps the physical properties of the variables and the capacity of the units, the classification is improved. The number of variables can furthermore be reduced by guidance from the normal of the hyperplane.

The data set must be selected with knowledge of the process, using already known information and also a great consideration of what we want to classify. In [11] it is mentioned that all variables that can be used should be used, because they have a contribution to the description of the process. In practice it is often a large covariance between variables and it can cause a fault classification due to over-fitting classifications models [14].

The dataset that is used to improve the classification of power system stability with SVM has not the proper variability for a final implementation. With a proper variability of the training set the corresponding support vectors can be used for development of an early warning system, in real-time. This application can be developed in Simulink interacting with the ARISTO simulator.

Chapter 6

Conclusions

All results indicate that SVM is an effective method for classification of power system stability. The thesis presents some of the issues that need to be adjusted to improve the classification result. The work involves the use of statistical learning algorithms to solve problems in power system operation. This thesis should also inspire for future research, because there is a large amount of problems that have not been solved yet.

6.1 Summary of the results

The SVM method is more adaptable to the classification problem in this thesis than PCA and the result is much more satisfying. With the linear SVM the hyperplane is calculated that separates the stable from the unstable example in the training set with the maximum distance between them. The hyperplane gives a better indication of the security of a state than the cut-set methodology. The calculation time is also short due to the global optimization problem of the SVM. In general, the SVM does not perform well outside the span of the training data. The training and classification set in this thesis are too small. In the work by Wehenkel, see [39], the number of examples in the training set is 11 000. For further work the number of examples must be increased.

The three main contribution of this thesis are:

- The method to scale the variables to keep the physical properties and the capability of the units, see Section 5.2.

- The method which excludes the binary variables, due to the difficulties to classification with SVM with a mixture of binary and continuous variables, see Section 5.5.1.

- The method to reduce the number of variables by guidance from the hyperplane, see Section 5.5.

These three methods are introduced in this thesis and they contribute to improve the classification result.

6.2 Future work

The classification result in this thesis is quite successful, but the method must be further verified before a possible implementation in power system operation. The size of the training and classification sets are too small and must be increased. With the ARISTO simulator it is possible to produce realistic synthetic data, thanks to the detailed models and the dynamic simulations. However, the interface of ARISTO is not developed for this type of study. The amount of time it takes to develop the data set is very long. Therefore consideration must be taken if ARISTO should be used in the future for this project. The dimension of the training set should be at least the same number of examples as measured variables, 653.

In this thesis linear SVM was used and the impact of adding non linear functions to the hyperplane has not been investigated. Therefore an analysis of the effect of adding non linearities to the SVM to solve the classification problem is motivated. The non linear SVM provide the user to the choose what non linear functions that shall be added and also the order of them. The frequently used functions today are polynomial, sigmoid and gaussian.

The scaling function was modified in this thesis to compensate for the capability of the generators. It should be interesting to develop methods which also compensate for the transmission capability of the power system. By adding this, the changes in the transmission, e.g. disconnection of lines and changes in transmission capability due to over heated lines, could easier be observed by the separating hyperplane. The improvements by adjusting the scaling method will hopefully make the training set more robust.

In this thesis no methods have been developed for dealing with the problem of how to check if the training set is representative for the classification example. In general, classification methods can not be expected to perform well outside the span of the training data. With a larger training set, it would be interesting to study the structure and span of the training data in the space of the measured variables.

Bibliography

[1] Nordel grid planning rules for the nordic transmission system. www.nordel.org, 2005-02-27, 1992.

[2] Rekommandasjon for frekvens, tidsavvik, reglerstyrke og reserve. www.nordel.org, 2005-02-27, 1996.

[3] Elavbrottet 23 september 2003 - händelser och åtgärder. www.nordel.org, 2005-04-25, 2003.

[4] http://www.nordel.org, 2005.

[5] CIGRE TF 38.02.08. Long term dynamics. phase ii. final report. Technical report, CE/SC 38 GT/WG 02, Ref. No. 102, 1995.

[6] S. Arnborg. *Emergency control of power systems in voltage unstable condition*. PhD thesis, Royal Institute of Technology, Sweden, 1997.

[7] H. Blomqvist, editor. *Elkraftsystem 1*. Liber AB, 1997. ISBN 91-47-00064-3.

[8] N. Cristianini and J. Shawe-Taylor. *An introduction to support vector machines and other kernel-based learning methods*. Cambridge University Press, 2000. ISBN 0-521-78019-5.

[9] A. Edström and K. Walve. The training simulator aristo - design and experiences. In *IEEE Power Engineering Society Winter Meeting*, 1999.

[10] L. Eriksson, E. Johansson, N. Kettaneh-Wold, C. Wikström, and S. Wold. *Design of experiments principles and applications*. Umetrics, 2000.

[11] L. Eriksson, E. Johansson, N. Kettaneh-Wold, and S. Wold. *Multi- and megavariate data analysis*. Umetrics, 2001. ISBN 91-973730-1-X.

[12] L. H. Flink and K. Carlsen. Operating under stress and strain. In *IEEE Spectrum 15*, 1978.

[13] J. Glover and M. Sarma. *Power system analysis and design*. Brooks/Cole, 2002. ISBN 0-534-95367-0.

[14] T. Hastie, R. Tibshirani, and J. Friedman. *The element of statistical learning; data mining, inference and prediction*. Springer Series in Statistics, 2001. ISBN 0-387-95284-5.

[15] A. Höskuldsson. *Prediction methods in science and technology*. Thor Publishing, 1996. ISBN 87-985941-0-9.

[16] K. Hotta. Support vector machines with local summation kernel for robust face recognition. In *Proceedings 17th International Conferance on Pattern Recognition*, 2004.

[17] C-W. Hsu, C-C. Chang, and C-J Lin. A practical guide to support vector classification. Technical report, Department of Computer Science and Information Engineering, National Taiwan University, 2002.

[18] T. Joachims. Text categorization with support vector machines: learning with many relevant features. In *Proceedings of ECML-98, 10th European Conference on Machine Learning*, 1998.

[19] B. Johansson. *Computer vision using rich features - geometry and systems*. PhD thesis, Lund Institute of Technology, 2002.

[20] R. Johnson and D. Wichern. *Applied multivariate statistical analysis*. Prentice Hall, 2002. ISBN 0-13-834194-X.

[21] W. J. Krzanowski. Discrimination and classification using both binary and continuous variables. *Journal of the American statistical association*, 1975.

[22] P. Kundur. *Power system stability and control*. McGraw-Hill, Inc, 1993. ISBN 0-07-035958-X.

[23] P. Kundur, J. Paserba, V. Ajjarapu, G. Andersson, A. Bose, C. Canizares, N. Hatziargyriou, D. Hill, A. Stankovic, C. Taylor, T. Van Cutsem, and V. Vittal. Definition and classification of power system stability. *IEEE Transaction on Power Systems*, 19(2), 2004.

[24] T. Larsson. Driftsinstruktion, spänningsreglering på stamnätet. Technical report, Svenska Kraftnät, 2003.

[25] G. Lindgren and H. Rootzén. *Stationära stokastiska processer*. KFS i Lund, 2003.

[26] P-A. Löf. *On static analysis of long-term voltage stability*. PhD thesis, Royal Institute of Technology, Sweden, 1995.

[27] J. Ma, Y. Zhao, and S. Ahalt. Osu svm classifier matlab toolbox (ver 3.00), 2002. web http://eewww.eng.ohio-state.edu/ ∼maj/osu_svm.

[28] J. N. Miller and J. C. Miller. *Statistics and chemometrics for analytical chemistry*. Prentice Hall, 2000. ISBN 0-130-022888-5.

[29] D. Montgomery. *Design and analysis of experiments*. John Wiley & Sons, 2001. 0-471-73304-0.

[30] L. S. Moulin, A. P. Alvares da Silva, M. A. El-Sharkawi, and R. J. Marks II. Support vector machines for transient stability analysis of large-scale power systems. *IEEE Transaction on Power Systems*, (2), 2004.

[31] E. Osuna, R. Freund, and F. Girosi. Training support vector machines: an application to face detection. In *Proceedings Conference Computer Vision and Pattern Recognition*, 1997.

[32] K. Pearson. On lines and planes of closest fit to systems of points in space. In *J. Edcu. Psych.*, volume 2, pages 559–572. 1901.

[33] B. D. Ripley. *Pattern recognition and neural networks*. Cambridge University Press, 1996. ISBN 0-521-46086-7.

[34] C. Taylor. *Power system voltage stability*. McGraw-Hill, Inc, 1994. ISBN 0-07-063184-0.

[35] E. Vaahedi, C. Fuchs, Y. Mansour W. Xu, H. Hamadanizadeh, and K. Morison. Voltage stability contingency screening and ranking. *IEEE Transaction on Power Systems*, 14(1), 1999.

[36] V. Vapnik. *The nature of statistical learning theory*. Springer Verlag, 1995. ISBN 0-387-98780-0.

[37] V. Vapnik, B. Boser, and I. Guyon. A training algorithm for optimal margin classifier. In *Proceedings of the 5th Annual ACM Workshop on Computer Learning Theory*, 1992.

[38] A. Webb. *Statistical Pattern Recognition*. John Wiley & Sons, 2002. ISBN 0-470-84513-9.

[39] L. Wehenkel. *Automatic learning techniques in power systems*. Kluwer Academic Publisher, 1998. ISBN 0-7923-8068-1.