ISSN 0280–5316
ISRN LUTFD2/TFRT--5459--SE

Användning av nevrala nätverk som prediktionsmetod vid investeringsanalys

Jens Henriksson
Björn Törnqvist

Institutionen för Reglerteknik
Lunds Tekniska Högskola
Juni 1992

Department of Automatic Control
Lund Institute of Technology
P.O. Box 118
S-221 00 Lund Sweden

Document name
MASTER THESIS

Date of issue
June 1992

Document Number
ISRN LUTFD2/TFRT--5459--SE

Author(s)
Jens Henriksson and Björn Törnqvist

Supervisor
Rolf Johansson, Curt Wells

Sponsoring organisation
Tidningen Expressen

Title and subtitle
Användning av neurala nätverk som prediktionsmetod vid investeringsanalys (Use of neural networks as a prediction method in investment analysis).

Abstract

The thesis treats the problem with predicting econometric systems using neural networks. An example is given in predicting the edition of one of the largest newspapers in Sweden, EXPRESSEN.

Key words

Classification system and/or index terms (if any)

Supplementary bibliographical information

ISSN and key title
0280–5316

ISBN

Language
Swedish

Number of pages
45

Recipient's notes

Security classification

Innehållsförteckning

1 Inledning

Vid investeringsanalys står man ofta inför problemet att beslutsunderlaget är otillräckligt. När skall jag köpa lägenheten? Kommer priset att gå upp eller ner? Ett av sätten att minska osäkerheten är att använda sig av prediktion. Det används mycket inom det tekniska området, från autopiloter i flygplanet JAS-Gripen till prediktion av fosforhalten i Ljungbyån. När det gäller ekonometriska data har man sett med viss skepsis på detta. Standardargumentet brukar vara att ekonometriska modeller bygger på hur människor reagerar, och inte på några variabler i en differentialekvation.

I den här rapporten försöker vi visa att det går bra att använda matematiska modeller för mycket komplicerade förlopp. Rapporten är framför allt inriktad på prediktion av tidningsförsäljning. Detta är inget nytt problem, utan det har varit föremål för många studier av bl.a. Hederstierna [1981]. Vid försäljningsprediktion försöker man helt enkelt "gissa" hur många som kommer att gå till en given kiosk för att köpa sin kvällstidning. Att information om detta beteende kan vara till stor nytta för kvällstidningarna är ganska självklart. De kan trycka mindre tidningar samtidigt som de kan sälja flera. Det speciella med dagstidningar är att de är en färskvara, "Who wants yesterdays paper? Nobody in the world" [Rolling Stones]. Detta medför ännu högre krav på prediktionen då tidningarna ej kan sparas till morgondagen, utan måste kasseras varje kväll.

I uppsatsen använder vi oss bland annat av något som kallas för artificiella neurala nätverk (ANN) [Ny Teknik, 1991]. I den svenska litteraturen finns även en annan beteckning, neuronnät. Vi måste redan nu påpeka att vi blivit mycket imponerade av denna tekniks möjligheter och vi tror att ANN har ett stort användningsområde inom ekonometrin för olika typer av prediktionsanalys och i flera olika tekniska sammanhang.

2 Expressen - ett exempel

2.1 Tidningen Expressen

Expressen är Nordens största kvällstidning. Den hade 1991 en medelupplaga på 576300 exemplar per dag. Expressens intäkter kommer huvudsakligen från lösnummersförsäljning (68.7%) och annonsintäkter (29.7%). Att öka lösnummerförsäljningen är det avgörande ekonomiska målet för tidningen. För att kunna sälja mycket tidningar måste kunden ha tillfälle att enkelt kunna köpa tidningen. En grupp på 2 personer bestämmer varje morgon hur många tidningar som skall tryckas, och hur de skall distribueras. Upplagan varierar mellan 450 000 och 650 000 tidningar och det finns c:a 16 500 försäljningsställen i landet. Ingen återförsäljare är den andra lik, vissa säljer i genomsnitt 3 tidningar per dag, medan andra säljer över 500.

Några frågorna vi ställs inför är: Hur många tidningar skall Expressen investera i? Hur skall dessa tidningar fördelas mellan återförsäljarna?

2.2 Vad påverkar antal sålda

Vilka faktorer påverkar hur många som tänker köpa Expressen i dag? Man kan förmoda att löpsedeln spelar stor roll, men vilka andra faktorer spelar in? Det är viktigt att poängtera att följande antaganden endast är antaganden och inget annat. För att kunna bevisa vad som verkligen påverkar folk att köpa tidningen, krävs stora och dyrbara marknadsundersökningar.

Nyheter, löpsedel

Expressens försäljningsrekord är 957 000 sålda tidningar. Detta inträffade dagen efter Kungens och Silvias bröllop. Att rekordet slogs just den dagen är ingen slump. Vad som händer i världen och på hemmaplan påverkar naturligtvis hur många tidningar som säljs. Ett exempel på detta är följande: När vi studerade våra data var det en dag som hade en markant högre försäljningssiffra än övriga dagar. Vi trodde först att det hade blivit ett fel, men det visade sig att det var dagen efter Carolas schlagerfestivalvinst och Sveriges seger i ishockey-VM.

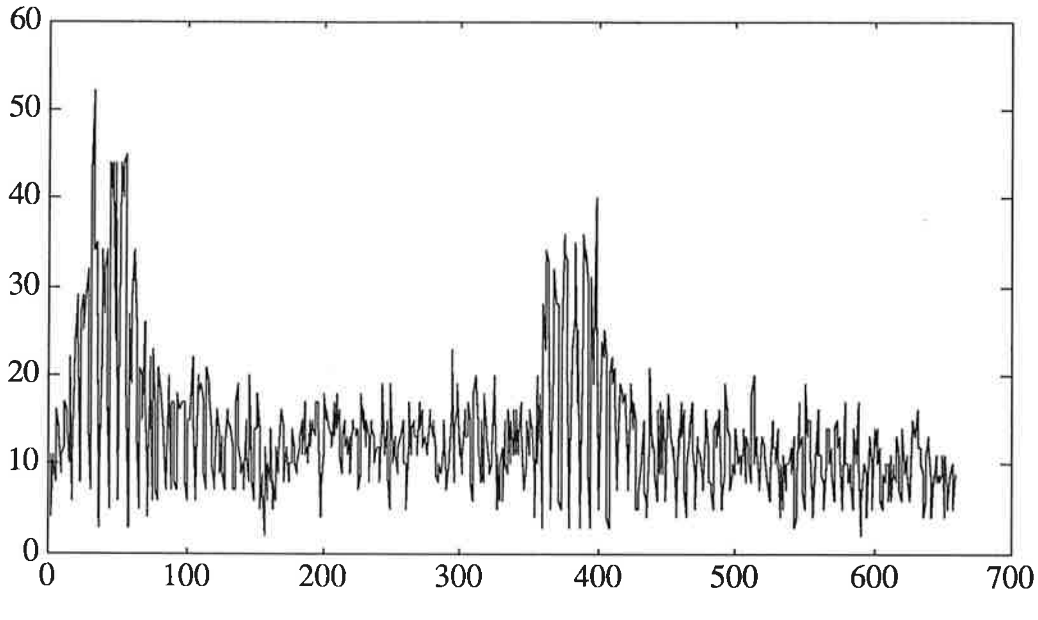
Figur 1. På Expressen finns ett arkiv där alla löpsedlar och annan information sparas.

Veckodag

Det finns en variation mellan veckodagarnas medelvärden. Detta kan dels bero på vanor hos svenska folket, och dels bero på de bilagor som finns i tidningen. TV-bilagan på torsdagar och Nöjes-bilagan på fredagar påverkar definitivt hur många som köper tidningen. Ett av de första tipsen vi fick av folk som jobbar med prediktion, var att Expressen skulle betraktas som sju veckotidningar snarare än en dagstidning. Ett exempel på att försöka ändra på detta är Expressens nysatsning på en tjej-bilaga. Den ges ut en dag där försäljningen hittills har varit lägre än övriga dagar.

Säsongsvariation

Köper folk mer eller mindre tidningar under semestern? Hur stor är säsongsvariationen? Siffror som Expressen presenterar, pekar inte på någon speciellt stor säsongsvariation för genomsnittsförsäljaren. I normala fall brukar försäljningen minska med 7 - 10 % från mars till maj, men under 1991 låg försäljningen relativt stilla. Om man däremot går in på enskilda återförsäljare kan säsongsvariationen vara betydlig. Ett exempel på detta är Mora som markant ökar sin tidningsförsäljning under Vasaloppet, eller Visby som varje sommar stormas av nyhetshungriga turister.



Figur 2. Säsongsvariationen hos en återförsäljare i Visby. På x-axeln: tid (dagar under 2 år), y-axeln: antal sålda tidningar. Topparna som syns är den kraftigt ökade försäljningen under sommaren.

Trender

Genom tiderna har Expressen haft varierande framgång.Trenden de senaste åren har varit varierande, men nu börjar en uppgång skönjas. Hur framtiden kommer att arta sig är mycket svårt att gissa. Det är många saker som påverkar trenden. Hur stor roll spelar en duktig journalist? Vad påverkar mest, en ändrad layout av tidningen eller konjunkturen? Detta är ett problem som vi måste lämna åt marknadsanalytikerna.

Väder och vind

Att vädret påverkar tidningsförsälningen är självklart. Att ringa SMHI varje dag är ganska billigt och enkelt. Men hur bär man sig åt för att utnyttja vädret i prediktionssammanhang? Är regn och +4 grader sämre eller bättre än sol och -20 grader. Dessutom kan det vara stora vädervariationer även inom små områden. Vi har valt att inte försöka använda oss av vädret i våra prediktioner. Expressen gör dock flera försök på detta område. Bl.a. använder det sig av en ny strategi på Gotland: Om det är soligt väder minskas leveransen av tidningar till Visby-stad, medan den ökas till stränderna. Beslutet om detta fattas inte av försäljningsanalytikerna utan det delegeras till tidningsutkörarna.

Antalet levererade tidningar

Om Expressen levererar 30 tidningar kan de självklart inte sälja 50 tidningar. Men hur är det åt andra hållet? Ökar försäljningen om det levereras för många tidningar? Att det kan påverka försäljningen på köpmarknader, med många utgångskassor verkar troligt. Då kan varje kassa få så mycket tidningar att tidningen finns tillgänglig i hela varuhuset. Vi har dock haft som utgångspunkt att det inte påverkar försäljningen i en större grad. Om en kioskägare i Eslöv får 1000 tidningar säljer han inte mer Expressen än om han fått 100 st.

Andra faktorer.

Det finns säkert många andra faktorer som påverkar antalet sålda tidningar. Några exempel på detta är: Behandlades löpsedelsnyheten på TV dagen innan? Hur ser konkurrentens löpsedel ut? Har det varit nyhetstorka eller har det varit många stora nyheter och skandaler nyligen? O.s.v. Det går säkert att hitta många mer faktorer. Alla har dock det gemensamt att de är mycket svåra att mäta. Ekonometriska variabler är för det mesta mycket svårmätbara

2.3 Expressens tillvägagångssätt

Expressen använder en modell bestående av en grundupplaga och ett dagligt pålägg som tillsammans bildar den faktiska dagliga upplagan. För att bestämma grunden används ett datorprogram. Detta program jobbar efter en enkel modell: Om det har blivit många tidningar över skärs tilldelningen ner efter ett tag, och tvärtom. Vi har blivit imponerade av att denna enkla prediktion fungerar så bra som den gör.

2.3.1 Pålägg

Pålägget bestäms på morgonen efter en titt på löpsedeln, förstasidan, extrema väderleksförhållanden och eventuella lokala händelser. Återförsäljarna kan också själva ringa in och begära mer tidningar om de t.ex. har REA på fläsk och tror att de kommer sälja fler tidningar. Här kommer vi även in på frågan om hur bra Expressen är på att prediktera topparna och dalarna. Den frågan är mycket svår att besvara. Vid en titt på riket i sin helhet är de ganska duktiga, men när man går ner på den enskilde återförsäljaren är det svårare att bedöma resultaten. Vid några mindre analyser som vi gjort har vi fått resultat som pekar på att Expressen gissar bra i c:a 60 % av fallen. Helt klart är att det finns en viss information att hämta från pålägget.

2.4 Vad vi har för data

De data vi fick från Expressen var organiserade på följande vis:

1	2	3	4	5	6	7	8	9
2113	891231	H	01	0181	05	55	43	0
2178	890703	Må	01	0181	01	30	30	5

Figur 3. En typisk tabell från Expressen

Följande data finns i tabellen:

kolumn 1	försäljningsnr	ex. 2113 Konsum Snabbköp Bragev. 44
kolumn 2	datum	ex. 891231. Nyårsafton 1989
kolumn 3	dagtyp	ex. Må = måndag, H= helgdag
kolumn 4	region	ex. 01 Stockholm/Södertälje, 21 Visby
kolumn 5	kommun	ex. 0181 Södertälje
kolumn 6	bransch	ex. 05 livsmedel, 01 kiosk
kolumn 7	levererade	antalet levererade tidningar
kolumn 8	sålda	antalet sålda tidningar
kolumn 9	pålägg	pålägget, se ovan

Vid modellbygge använde vi data från 890601 till 910429. När vi körde "skarpt" utnyttjade vi data från 910801 till 920330. D.v.s vid validering har vi använt data som **ej** använts i modellbyggandet.

2.5 Vad skall vi bygga vår modell efter?

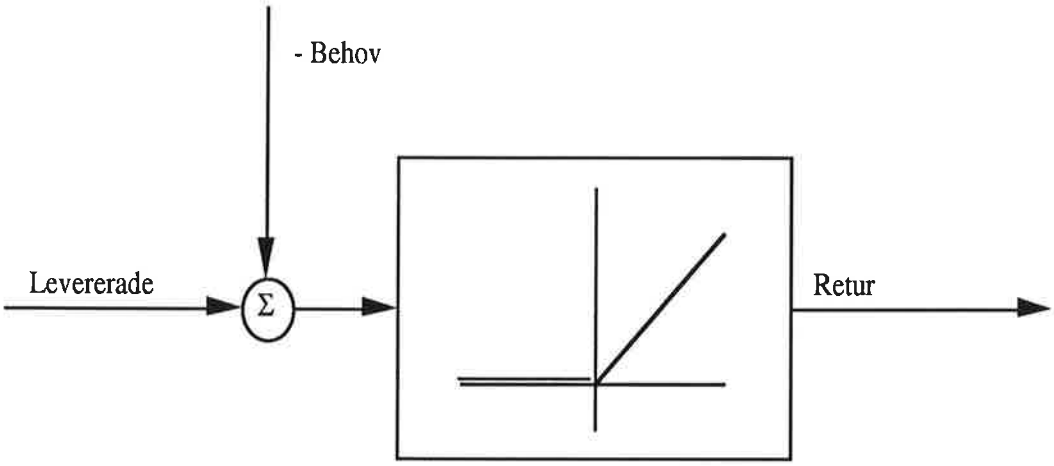
I vår modell för hur många tidningar Expressen skall investera i, måste vi välja indata. Modellbyggandet sker först och främst efter de försäljningsdata vi har. Vårt första beslut var att rensa bort alla helger ur dataserien p.g.a. att de skiljer sig från de övriga dagarna. Onsdagen den 31/12 har en försäljning som är skild från andra onsdagar i december.
Nästa fråga var hur mycket hänsyn vi skulle ta till det s.k. pålägget. En första tanke var att subtrahera pålägget från all försäljning för att enbart få den s.k. grundförsäljningen. Men hur bestämmer vi den verkliga grund- och påläggsdelen ur antalet sålda tidningar? Finns det över huvud taget en grund för ett sådant tänkande? För att lösa dessa frågor måste det genomföras undersökningar om och varför folk köper tidningen. Vi valde att inte alls manipulera försäljningsdata.

2.5.1 Slutförsäljning

Ett annat stort och svårdefinierat problem är slutförsäljning. Vi har använt Expressens definition på slutförsäljning:

> Det är slutsålt när det är 0 eller 1 tidning kvar.

Den avgörande frågan är: Hur många tidningar hade kunnat säljas om tillgången på tidningar hade varit obegränsad? I data från Expressen är antalet sålda exemplar trunkerat. Om de har sålt 40 tidningar och levererat 41 st. så kunde de kanske i själva verket ha sålt 50 tidningar. Om vår modell har valt att skicka ut 55 tidningar kommer vi att räkna det som om vi har fått en retur på 15 tidningar fastän den i verkligheten kanske varit 5 tidningar. Det är viktigt att redan nu poängtera att alla jämförelser mellan oss och Expressen har varit till vår nackdel. Detta har vi valt för att det inte skall finnas någon som helst risk att vi överdriver våra resultat. Nedanstående bild visar hur vi reglertekniskt definierade problemet med slutförsäljning.



Figur 4. Principiell bild av problemet med slutförsäljning.

2.5.2 Skall data delas upp efter typ av veckodag?

Spelar måndagens försäljning roll för tisdagens? Eller beror tisdagens efterfråga snarare på föregående tisdags försäljning? Det finns två olika principiella sätt att se på tidsserien, antingen som en dagstidning eller som 7 oberoende veckotidningar.

DagsExpressen: Även om försäljningen på tisdagar i genomsnitt ligger på samma nivå, kan den bero på måndagens försäljning. Om det är extrapris på konserverad gröt i Konsum-butiken hela veckan, kommer försäljningen att ligga högt. Då är dagsberoendet stort.

MåndagsExpressen, TisdagsExpressen, Onsd... : Att detta verkar vara en trolig modell styrs av att medelvärdet de olika dagarna varierar.

I kapitel 4.2.1 kommer vi att diskutera beroendet mera ingående.

2.6 Hur kan vi jämföra oss med Expressen?

Hur skall vi kunna jämföra olika resultat med varandra?
Om tidningen säljs slut, är detta inte bra. Det är bättre att ha två tidningar över än att ha en tidning för lite. I dagens läge tjänar Expressen c:a fyra kronor på varje tidning som säljs. Den rörliga kostnaden är ungefär en krona per tidning. Här kommer även andra aspekter än de rent ekonomiska in. Om Expressen kan hålla en hög upplaga kommer det att öka efterfrågan på annonser i tidningen. För att få med dessa aspekter valde vi att använda följande två jämförelsekriterium:

1. Hur många procent av tidningarna som kommer i retur.

2. Hur många procent av dagarna tidningen säljs slut.

Som det nämndes ovan anses det vara slutsålt om det är 0 eller 1 tidning kvar. Expressen funderar på att ändra den definitionen, till att istället gälla om det är mindre än c:a 3% av tidningen som går i retur. Detta för att klara fallen med stora köpcentra, där tidningarna kan ta slut vid vissa kassor.
För att senare kunna jämföra oss med Expressen kan vi berätta att deras genomsnittliga värden idag ligger på c:a 20 % slutförsäljning och 23 % retur. D.v.s. en normal dag säljer 20 % av försäljarna slut på tidningen. Samtidigt får Expressen 23 % av antalet ivägskickade tidningar i retur. Expressens statistik visar att de uppnår bäst resultat hos de återförsäljare som säljer mest tidningar.

3 Standardmetoden.

Nu hade vi kommit så långt att det var dags att börja med identifieringen. Vi har en utsignal som är antalet sålda tidningar. Vi har ingen insignal till systemet (Om vi inte skapar egna sensationer!). Vi följde helt enkelt ett tillvägagångssätt som rekommenderas av Johansson [1991]

3.1 ARMA metoden

En första spektralanalys visade att beroendet är störst mellan liknande veckodagar. Vi gjorde därefter ett försök med att modellera en ARMA-process [Johansson, 1991]. ARMA står för Auto Regressive Moving Average och har formen

$$A(q)Y(t) = B(q)\, e(t)$$

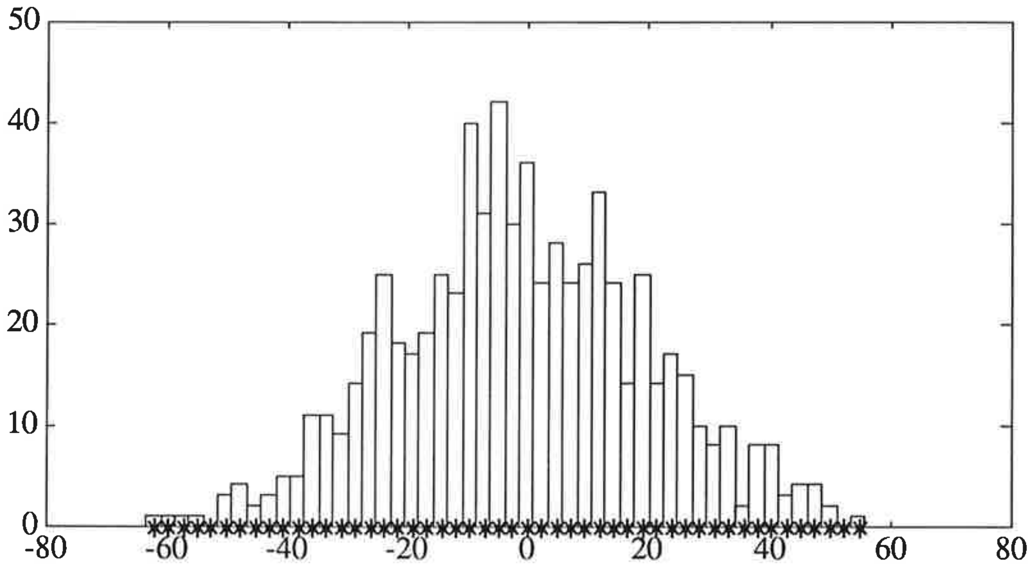
Den bästa ARMA-modell vi fick fram med hjälp av Akaikefunktionen [Johansson,1991] var en (2,2)-modell

$$Y_t = a_1 * Y_{t-1} + a_2 * Y_{t-2} + e_t + b_1 * e_{t-1} + b_2 * e_{t-2}$$

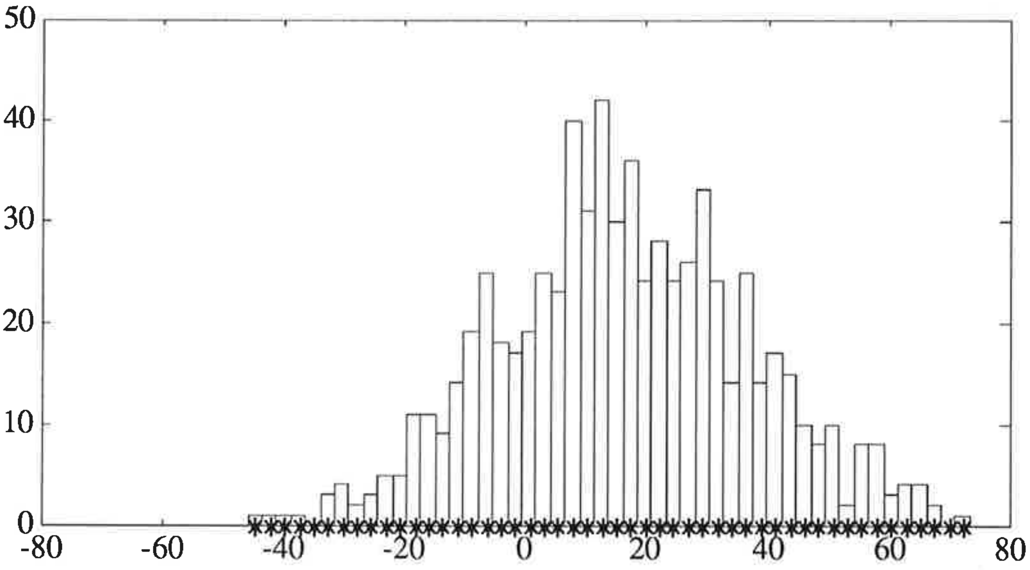
Expressen hade i genomsnitt c:a 16 tidningar över och sålde slut i 8 % av fallen på den återförsäljare, 2113, vi valde att koncentrera oss på. För att vår ARMA-modell skulle nå ett jämförbart resultat, d.v.s. ha samma slutförsäljnings-frekvens blev vi tvungna att öka alla prediktioner med 17.5 tidningar. Detta gjorde att vår modell fick 17.4 tidningar över i genomsnitt. Vår modell var alltså sämre än Expressens.

3.1.1 Fel fördelning

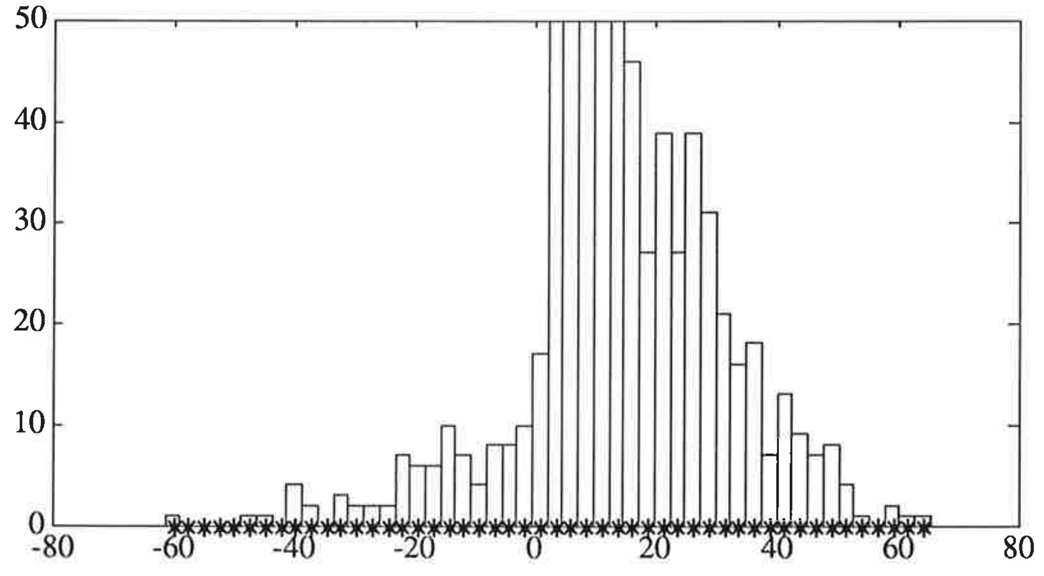
Vad beror det på att vi är sämre än Expressen? Lösningen är enkel, men svår att komma på. Funktionen ARMAX i MATLAB utnyttjar en ml-skattning [Blom,1984] vilken bygger på ett antagande om normalfördelning. Detta ger ett fel som oftast är normalfördelat.
Problemet är att vi inte vill ha ett normalfördelat fel. Återigen, det kostar mer att sälja slut tidningen än att ha tidningar över. Positiva och negativa fel är inte lika mycket värda.



Figur 5. ARMAX ger ett normalfördelat fel. På x-axeln syns antal dagar, på y-axeln prediktionsfelet.



Figur 6. När vi flyttar normalfördelningen i sidled får vi medelvärdet på fel ställe. På x-axeln syns antal dagar, på y-axeln prediktionsfelet.



Figur 7. Ungefär så här skulle vi vilja att felfördelningen skulle se ut. På x-axeln syns antal dagar, på y-axeln prediktionsfelet.

I figur 7 syns vad för typ av felfördelning vi tror är realistiskt att uppnå. Vi vill att medelvärdet av antalet tidningar som går i retur skall vara lågt, samtidigt som vi inte vill sälja slut för ofta.

3.2 Två enkla modeller

Det finns några olika sätt att skaffa sig ett icke normalfördelat fel. Ett av sätten är att använda sig av en listig transformering av datavärdena [Olbjer,1985], och ett annat är att hitta på en modell med en annorlunda straffunktion.

3.2.1 Transformation

Vi försöker först lösa problemet med en transformation. Problemet med detta är att alla återförsäljare har olika beteenden. Expressen skulle ha stora svårigheter att bygga ett system där alla 16500 återförsäljare har var sin transformation. Dessutom är det olika beteende beroende på vilken dag det handlar om, vilket skulle medföra 7*16500 transformationer. Alla transformationer skulle inte ha samma utseende.

För att visa att det är fullt möjligt att använda sig av en transformation har vi gjort det för en återförsäljare, 2113, på onsdagar. Här följer en beskrivning av tillvägagångssättet:

Vid en första snabb titt på fördelningen av antalet sålda, såg det ut som en normalfördelning. Vi normerade så att antalet sålda blev N(0,1) [Blom, 1984]. Denna normalfördelning transformerade vi om till en rektangelfördelning m.h.a. funktionen erf [Matlab,1991]. Nästa steg var valet av en sned fördelning.

Vi fastnade för exponentialfördelningen. Täthetsfunktionen har den egenskapen att den bara är större än noll för positiva tal. Formeln för transformationen mellan rektangelfördelningen och exponentialfördelningen togs direkt ur fördelningsfunktionen. Lägesparametern [Blom, 1984] valdes till 10. Nu hade vi en tidningsförsäljning som var exponentialfördelad.

Efter detta fortsatte vi med en ARMAX-modellering.Här fastnade vi återigen för en (2,2) modell efter användande av Akaikes felfunktion. Modellen utnyttjades sedan för att göra en prediktion av de transformerade värdena.

För att få en bra prediktion av tidningsförsäljningen transformerade vi tillbaka vår prediktion av de transformerade värdena. Först till rektangelfördelning, och sen slutligen tillbaka till den ursprungliga fördelningen. Nu hade vi en prediktion av tidningsförsäljningen.

Detta gav ett resultat som hade en genomsnittlig returprocent på 16.7 % och en slutprocent på 15.4 %. Detta kan jämföras med vanlig ARMAX som utan transformation gav en returprocent på 17.4 % (vid samma slutprocent 15.4 %).

En transformation kan alltså leda till förbättringar i storleksklassen 1-2 procentenheter jämfört med vanlig ARMAX. Men p.g.a. att många återförsäljare har olika fördelningar för sina antal sålda tidningar kommer detta att leda till problem. Även om en hel del återförsäljare har liknande fördelning, kommer detta att leda till ett ohanterligt problem för Expressen. Vi beslöt oss därför att lämna denna väg.

3.2.2 Modell med anpassad straffunktion

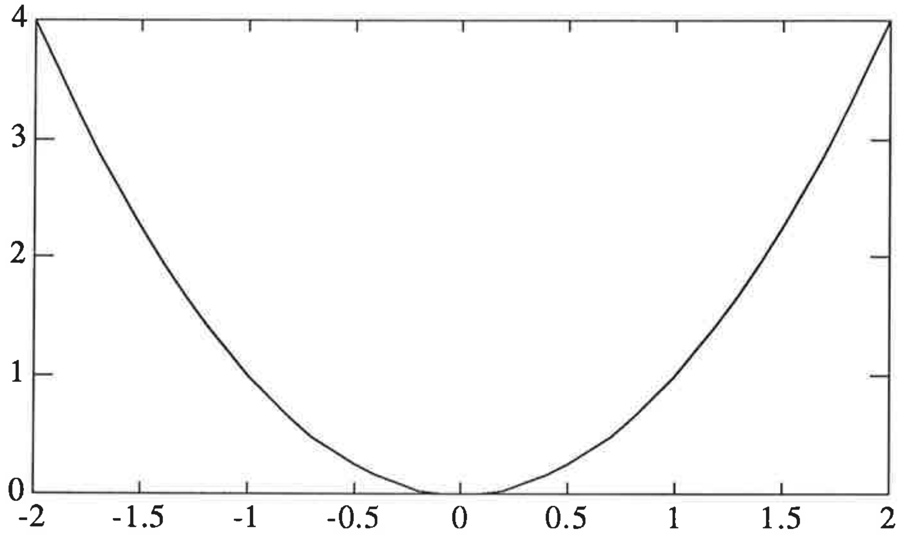
En annan metod är att påverka straffunktionen. Vi antog att dagens förmodade försäljning kunde räknas ut som en linjärkombination av föregående dagars försäljning. D.v.s. en AR-modell [Olbjer,1985].

$$A(q)Y(t) = e(t)$$

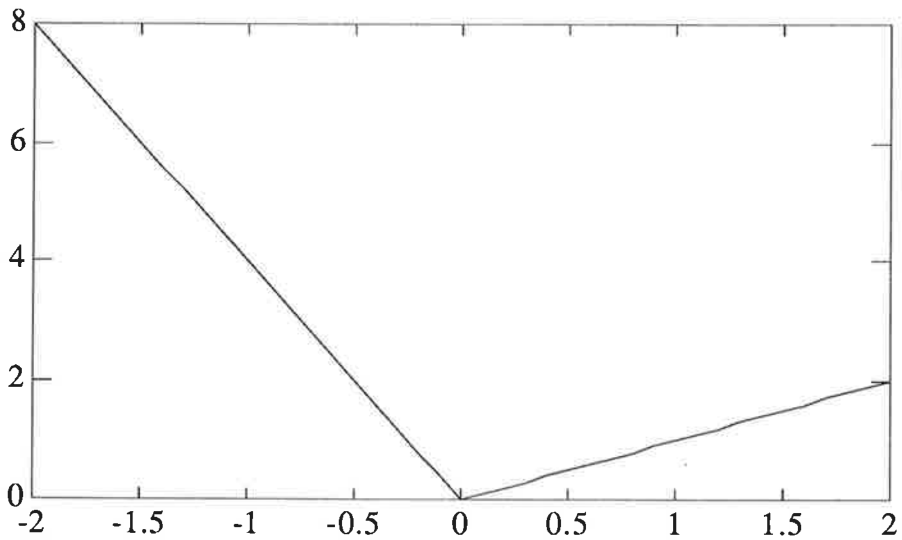
1-stegs prediktorn fås då enligt :

$$\widehat{Y}_t = \alpha_1 * Y_{t-1} + \alpha_2 * Y_{t-2} + \cdots + \alpha_n * Y_{t-n}$$

För att vi skulle få en modell som tog hänsyn till hur prediktionsfelet bör vara, beslöt vi oss för att prova med en annan straffunktion än den kvadratiska. Vi använde oss av en linjär funktion med olika lutningar för positiva och negativa prediktionsfel. Lutningen på kurvan bör bestämmas av Expressen, då den har avgörande betydelse för i hur många fall tidningen blir slutsåld. Här är det som vanligt ekonomiska aspekter som kommer att spela in. Vilken vikt lägger Expressen vid att kunden alltid skall ha tillgång till en tidning, även om det är fem minuter innan stängningsdags? När vi provade funktionen, hade vi ett straff som var 4 gånger så stort när tidningen sålde slut, som när det blev tidningar över.



Figur 8. En straffunktion som straffar lika mycket för negativa som för positiva fel.



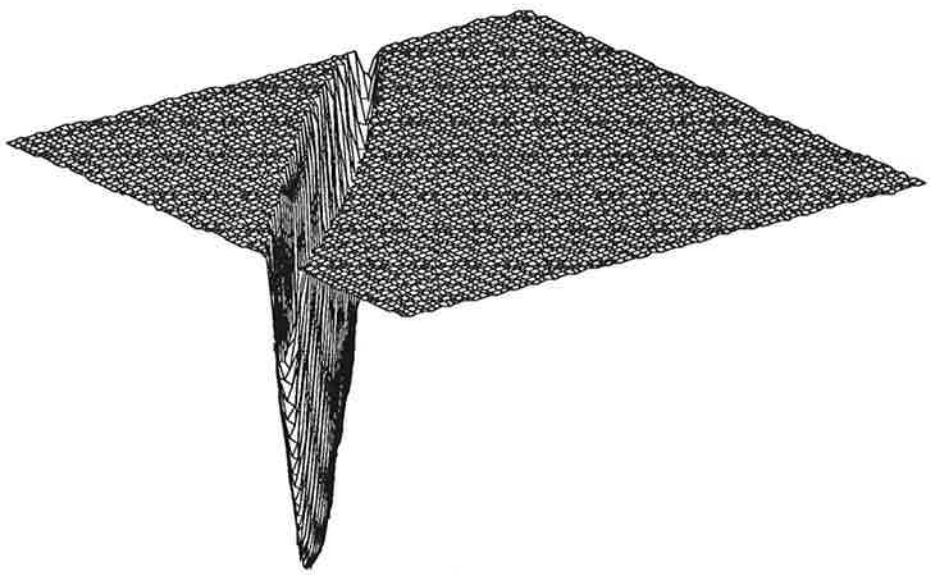
Figur 9. Ett exempel på en annorlunda straffunktion.

3.2.3 Realisering.

Parametrarna α_k beräknade vi med hjälp av optimeringspaketet i MATLAB. Funktionen som användes heter FMINS och utnyttjar Nelder-Meads algoritm [Grace, 1990] för att hitta de rätta värdena. Vad det handlar om är helt enkelt en minimering av en k:te gradens funktion.Vi provade bl.a. att göra en torsdagsprediktor som använde sig av de 5 föregående torsdagarna. Den såg ut på följande sätt:

$$\hat{Y}_t = 0.1854*Y_{t-1} + 0.4564*Y_{t-2} + 0.3696*Y_{t-3} + 0.0240*Y_{t-4} + 0.2212*Y_{t-5}$$

Om vi ville ha ungefär samma slutförsäljningsfrekvens som Expressen, fick vi ner medelvärdet av antalet överblivna tidningar från 19.7 till 17.9 stycken för återförsäljare 2113. Då hade Expressen sålt slut i 4.5 % av fallen och vi i 5.6 %. Här är svårt att göra några relevanta bedömningar om vi är bättre eller sämre än Expressen p.g.a att vi ej har provat denna metod på ett signifikant antal återförsäljare. Tydligt är att vi ligger c:a 2% bättre än vad Expressen gör.



Figur 10. I figuren ser vi hur felfunktionen beror av parametrarna a_1 och a_2. Nelder -Meads algoritm hittar enkelt minimum. Vi har skurit bort vissa bitar ur figuren.

Vad kan detta bero på? Om man studerar parametrarna α_k noggrant, ser man att det är något som är konstigt. Hur kan dagens försäljning bero mer på försäljningen för fem veckor sedan än på förra veckans, när detta inte syns i ett korrelationstest ? Svaret på detta är att minimeringen är svår att genomföra p.g.a. att det finns många lokala minimipunkter.
När man däremot väljer en funktion av andra ordningen är det lätt att hitta ett minimum. I fig. 10 ser vi hur felfunktionen beror på a_1 och a_2. Nelder-Meads algoritm hittar här hela tiden samma minimum oberoende av startvärden. Men när vi

utvidgar modellen till fjärde ordningens modell blir a_1, a_2, a_3 och a_4 beroende av startvärdena. Vi har problem med att få algoritmen att lägga in några "intelligenta" aspekter vid val av modell.Slutsatsen av resultatet från den enkla modell blir att vi måste komma på ett helt nytt sätt att lösa problemet.
Det enda skälet till att vi hittills har varit bättre än Expressen, är att vi har varit systematiska i vårt tillvägagångssätt.

4 Artificiella Neurala Nätverk

För att få en bra investeringsrutin för Expressen behöver vi en prediktor som bland annat måste ha följande egenskaper:

Snabb adaptering till mätdata
Olinjär
Något som ger liknande fördelning som mätdata

Vi fastnade till slut för något som kallas för Artificiella Neurala Nätverk (ANN). Denna metodik är inte på något sätt ny, utan har funnits ganska länge. Under sjuttiotalet blev det impopulärt att använda sig av ANN p.g.a. svårigheten att utföra massiva beräkningar och att det ansågs ha allvarliga begränsningar. Det saknas fortfarande en del teoretiska bevis. På senare tid har dock användandet börjat öka, och på Lunds Universitet bedrivs numera forskning inom området [Peterson,1989, 1991,1992].

4.1 Vad är ett neuralt nätverk?

Vad är nu artificiella neural nätverk? Vi väljer här att försöka beskriva vad det är, på ett lite mer populärvetenskapligt sätt [Peterson,1991]:

Kortfattat kan ANN beskrivas som ett försök att kopiera algoritmer från den mänskliga hjärnan. Varför försöker vi kopiera hjärnan?

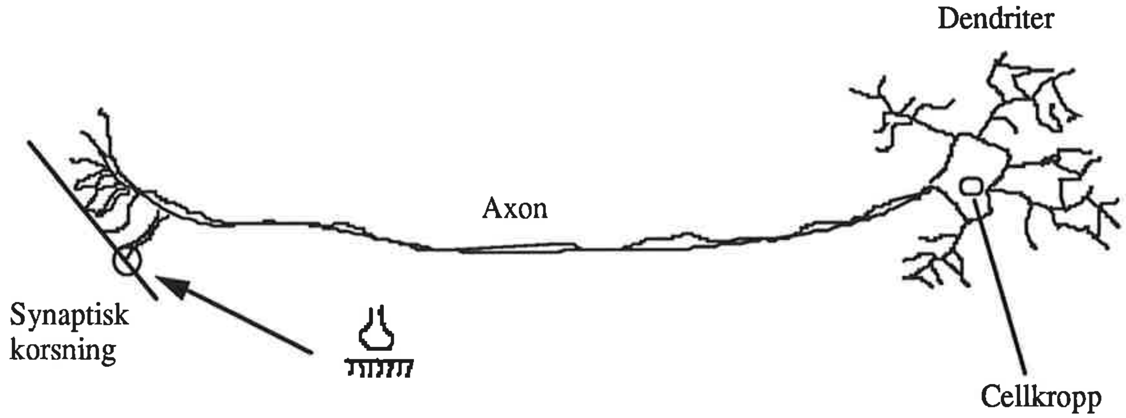
En dator är mycket bra på att genomföra enorma beräkningar på pikosekunder, lagra ofantliga mängder data och allt detta utan att göra några slarvfel. Människan kan inte alls utföra samma saker som en dator, men den kan göra andra saker som vid en första betraktelse verkar mycket enkla. Vi kan gå, prata, lyssna och känna igen Lasse Holmqvist fastän han har peruk på sig. Vi kan dessutom resonera med andra människor. Om någon ber om lagom mycket kaffe,så vet vi hur mycket vi skall hälla upp. Dagens datorer klarar inte detta. Men varför är det på detta sättet?

Det beror kanske på att en hjärna är uppbyggd för enkla men abstrakta saker, medan en dator är byggd för att klara aritmetiska operationer. I den mänskliga hjärnan finns uppemot 100000000000000 (10^{14}) neuroner. Det finns många olika typer, en schematiska bild ges i fig. 11. För den intresserade finns även en kort beskrivning av den medicinska bakgrunden.[1]

---

[1] Cellkroppen mottager elektriska insignaler till dendriterna med hjälp av jontransporter. Insidan av cellkroppen är negativt laddad gentemot en omgivande extracellular vätska. Signaler som kommer till dendriterna depolariserar den vilande potentialen, d.v.s. Na+ joner strömmar in i cellen genom jonkanaler i cellmembranet. Detta resulterar i en elektrisk urladdning från neuronen. Den ackumulerade effekten av många signaler från dendriter blir någorlunda linjär, men utsignalen från neuronen blir i högsta grad olinjär. Den elektriska urladdningen, utsignalen, fortsätter längs axonen till en synaptisk korsning, där neuronladdningen tar sig fram över en synaptisk klyfta för att komma vidare till de postsynaptiska dentriterna.

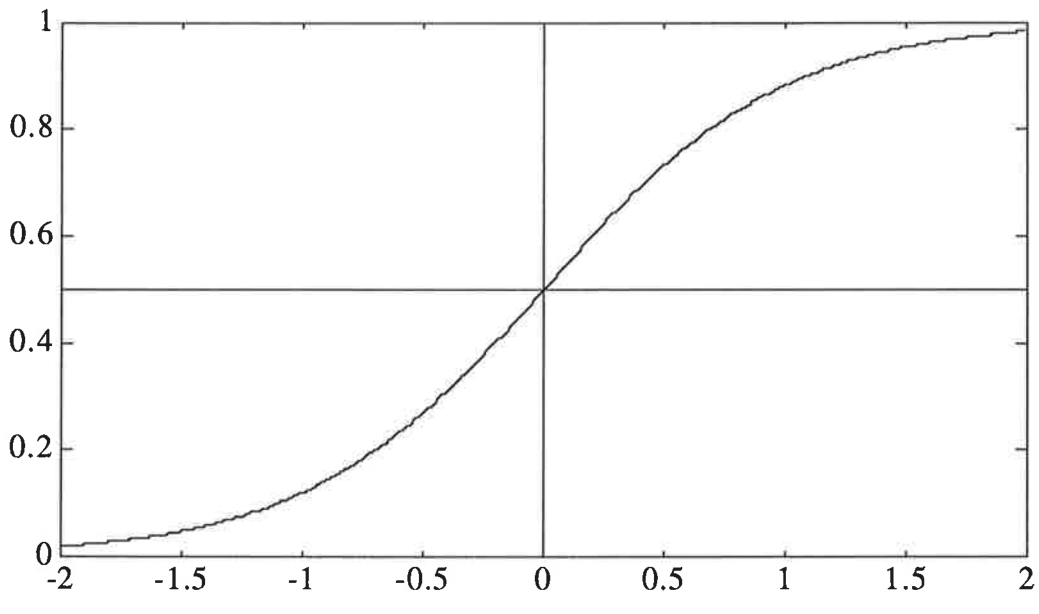
En synaps som ofta används kommer att bli starkare, medan en synaps som aldrig används kommer att bli svagare. Det är detta fenomen som spelar en nyckelroll i inlärning.



Figur 11. Schematisk bild av en neuron.

När vi bygger upp ett artificiellt nätverk så försöker vi efterlikna strukturen på ett matematiskt sätt. En av de viktigaste funktionerna i den biologiska bakgrunden är ett olinjärt beteende. För att efterlikna detta använder vi oss av en s.k. sigmoidfunktion i det artificiella neurala nätverket.

$$g(a_i) = \frac{1}{2}[1 + \tanh(a_i/T)]$$



Figur 12. En sigmoid kurva.

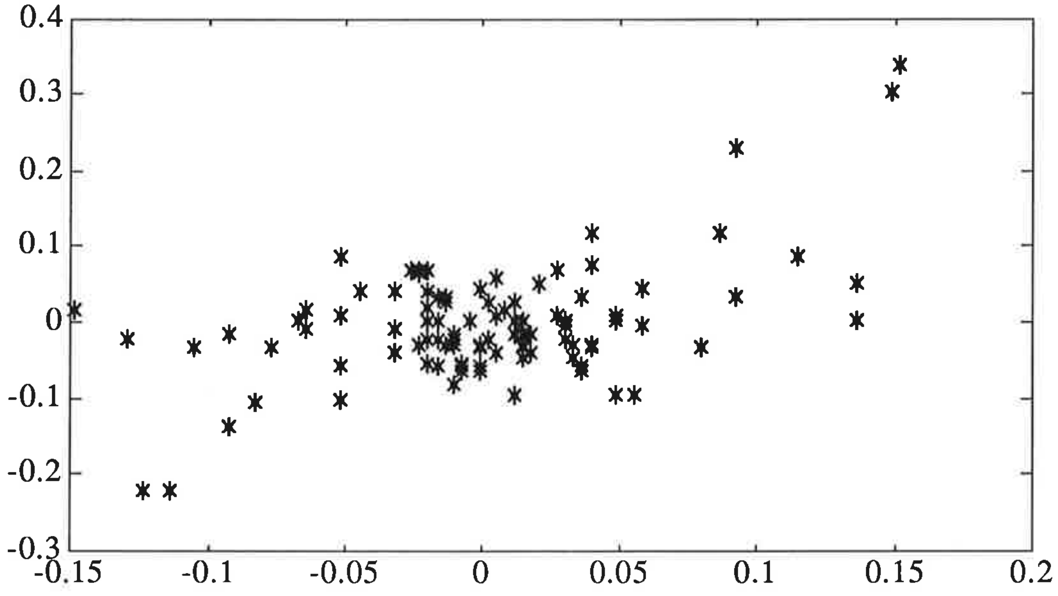
4.2 Vad gör ANN lämpligt för oss?

Vi har tidigare sett att antalet sålda och levererade tidningar inte har normalfördelade täthetsfunktioner. Vi kan också anta att fördelningarna varierar mellan återförsäljarna. Ett problem som uppstår är att anpassa en modell till var och en av de 16 000 återförsäljarna som skall få rätt antal tidningar varje dag.
Det bästa sättet vore om vi hittade en metod där varje återförsäljare fick en egen process som ställde in sig efter dess egenskaper. Modellens fördelning och övriga egenskaper skulle då kunna anpassas till återförsäljarens specifika egenskaper. Dessutom måste det ske en uppdatering av modellerna som skall kunna klaras av utan mänsklig inblandning. Ett neuralt nätverk för varje återförsäljare som uppdaterades med jämna mellanrum skulle kanske lösa våra problemen?

4.2.1 Vad har vi för beroenden ?

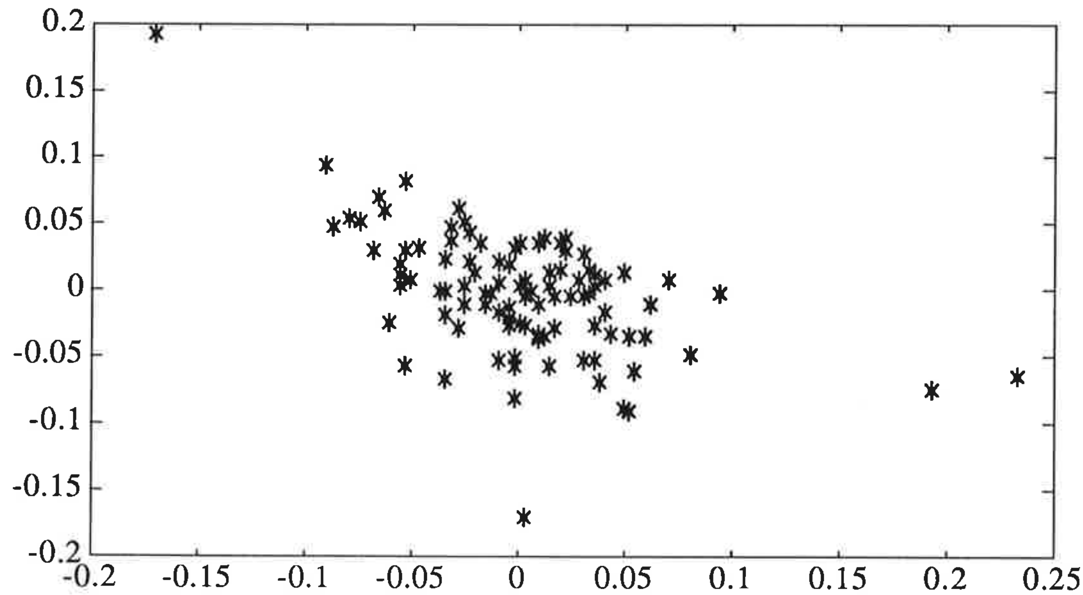
Prediktioners grundförutsättning är beroenden i tiden. Genom att se hur en tidsserie varierar i tiden kan man dra slutsatser om dess fortsatta utseende. Ett mått på beroendet mellan olika data är alltså ett mått på dess lämplighet som insignal till ett neuralt nätverk. För mätning av linjära beroenden finns flera metoder att tillgå. Några sätt är att skatta kovariansfunktionen eller korskovariansen [Johansson, 1991] vilka ger beroendet i tiden resp. mellan olika tidsserier. Med dessa metoder kommer vi fram till att det finns ett litet beroende en dag tillbaka i tiden resp. en vecka tillbaka i tiden. Ett litet beroende finns även två veckor tillbaka i tiden. Men neurala nätverk utnyttjar inte bara linjär information som AR-modeller, utan också olinjär information. Forskningsresultat visar också på svårigheten att använda sig av ovanstående grova mått [Rignér, 1991]. Icke-signifikanta beroende kan tillsammans bilda ett signifikant beroende [Bermin och Gyllerup, 1991].

Hur gör man för att skatta storleken på olinjära beroenden? Ett enkelt sätt som prof. K. J. Åström rekommenderade och som bl.a. psykologer använder, är att rita upp spridnings-diagram. Två tidsserier kan jämföras genom att en punkt plottas med den ena tidsseriens värde i x-led och den andras värde i y-led för varje observation i tiden . Detta kan ses som en del av linjär regression.



Figur 13. Som x-koordinat har vi normerad måndagsförsäljningen, och som y-koordinat normerad tisdagsförsäljningen. Dessa punkter ritar vi sedan ut i diagrammet. I figuren ovan har vi använt data från en Konsumbutik i Södertälje. Vi tycker att bilden visar på positivt beroende, d.v.s. om det säljs många tidningar på måndag, kommer det att öka sannolikheten att det säljs många tidningar denna tisdag.

I figuren ovan kan med lite fantasi en cigarr-form skönjas med positiv lutning. För en vecka som har liknande värde måndag och tisdag kommer punkten att hamna vid en linje med lutning 1. Om däremot värdena är motsatta blir lutningen -1. Vid betraktelse av samtliga punkter gäller, att om de ligger formade längs en kurva med en viss funktion f(x), finns ett beroende mellan tidsserierna som har att göra med f(x):s utseende. I figur 13 ses ett positivt beroende mellan dagarna, medan ett negativt beroende ses i figur 14.



Figur 14. Plott mellan lördagar med en veckas förskjutning, för en Konsumbutik i Södertälje. Vi tycker att bilden visar på negativt beroende, d.v.s. om det såldes många tidningar förra lördagen, kommer det att öka sannolikheten att det säljs mindre antal tidningar denna lördag

Vid granskning av spridnings-diagram för flera återförsäljare visar det sig att det finns beroenden en dag och en vecka tillbaka i tiden. Mellan pålägget (se ovan) och antal sålda finns för vissa dagar ett svagt beroende. I övrigt hittades ett litet beroende två veckor tillbaka.

Vår slutsats blir att följande insignaler kan anses motiverade för att prediktera dagens försäljning:

X1 = Antalet sålda för en vecka sedan.
X2 = Antalet sålda föregående dag.
X3 = Hur många procent Expressen tror att försäljningen kommer att öka eller minska, d.v.s. en "löpsedels-faktor".

Beroendena är av varierande slag, och vi vågar oss inte på någon djupare analys av dem. Det framgår dock, att de beroende som finns är både linjära och olinjära.

4.2.2 Signalernas förändring i tiden

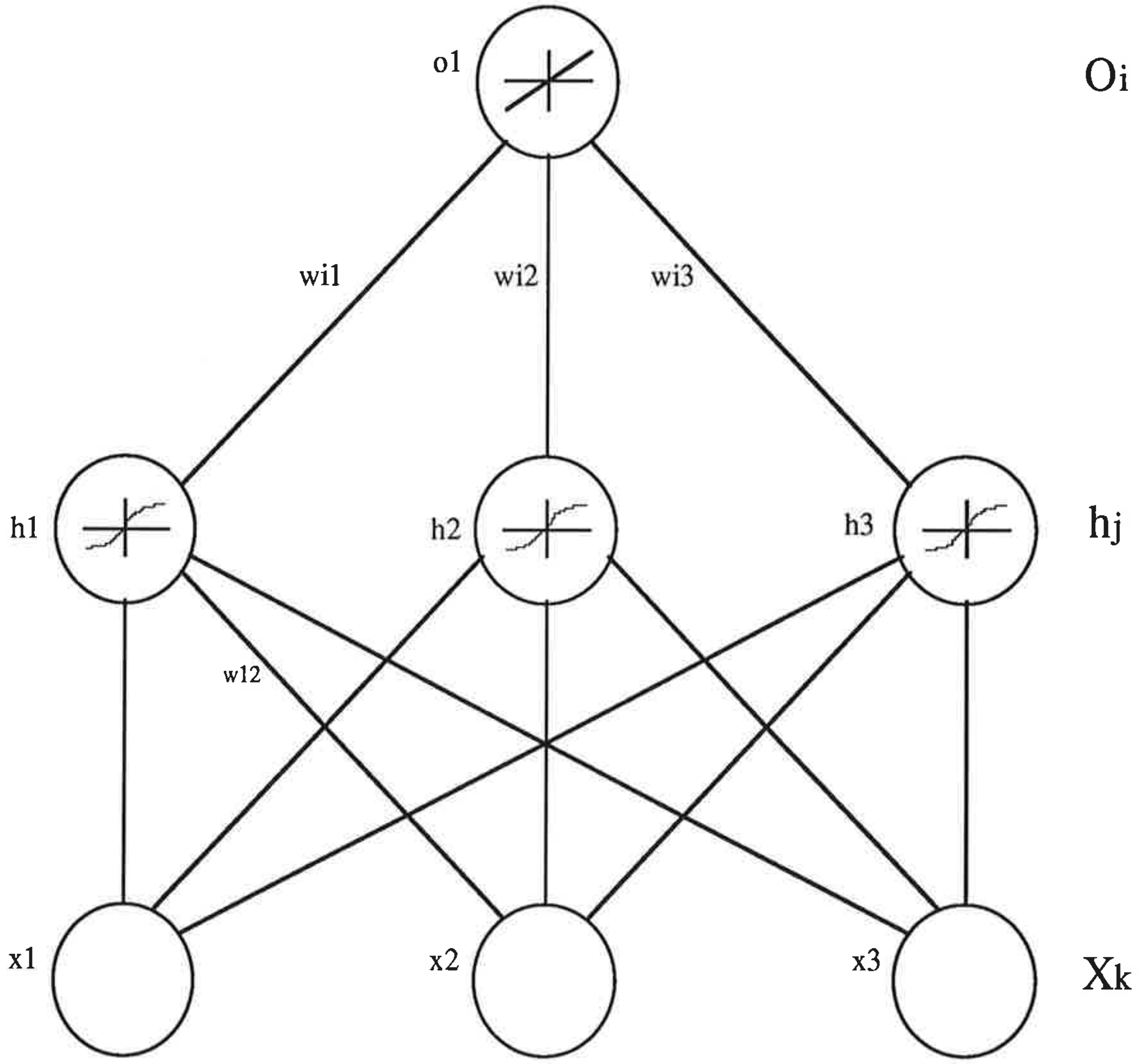
Om man granskar tidsserierna för olika återförsäljare får man intrycket av att ingen är den andra lik. Likaledes verkar de variera sitt utseende i tiden. Antag att en affär plötsligt börjar sälja sämre än tidigare p.g.a. vägarbete eller ägarbyte. Att manuellt variera parametrarna till en reglerande modell skulle i stort sett vara omöjligt för den stora mängden av återförsäljare. Det skulle därför vara smidigt för Expressen om upplagan reglerades adaptivt för varje process. I ANN är det lätt att lägga in en snabb och generell uppdatering för att klara av ändrade köpbeteenden.

4.3 Utformning av ett ANN

Det står nu klart att neurala nätverks-tekniken skulle vara lämplig i vårt fall. Nu gäller det att utforma nätet på ett snyggt och funktionellt sätt.

4.3.1 Val av modell

Vi har valt att ha tre insignaler till nätverket, och det gäller nu att bestämma storlek på nätet. I litteraturen finns en tumregel som säger att antalet parametrar i nätet skall vara i storleksordningen ≈ antalet datapunkter i insignalen dividerat med tio. Våra tidsserier är av storleksordningen hundra punkter och vi väljer därför ett nät med drygt tio parametrar. Ett gömt lager med tre noder leder till att vi får tolv parametrar,ω, vilket är acceptabelt. Att utformningen av nätet har skett på detta icke-teoretiska sätt beror på att teorin för val av nätets utseende och storlek är ofullständig. Nedan visas en bild på nätets principiella utseende



Figur 15. Utseendet på det valda neurala nätverket.

4.3.2 En beskrivning av hur vi tror att nätet fungerar

Nätet kan i princip ses som ett flertal MA-processer [Johansson, 1991] som är kopplade både parallellt och seriellt. Den avgörande skillnaden är att det finns olinjäriteter kopplade till varje MA-process i form av sigmoider. Detta gör att nätet klarar av att avbilda alla olinjära funktioner [Chen et al, 1989].
Vid inlärning av nätet förändras vikterna för att uppnå ett minimalt fel. Storleken på felet i utsignalen bestäms och utnyttjas, för att med hjälp av backpropagation [Peterson, 1991] ställa om parametrarna så att felet minskas. När felet har blivit tillräckligt litet avslutas träningen. De värden som finns på parametrarna behålls och används för att köra på verklig data.

4.3.3 Matematisk beskrivning

Här följer en kort matematisk beskrivning av hur det neurala nätverket är uppbyggt.

Som ovan nämndes bygger det neurala nätverket till en stor del på den icke-linjära sigmoiden

$$g(a_i) = \frac{1}{2}[1 + \tanh(a_i/T)]$$

som visades i kapitel 4.1. I vårt nätverk används sigmoidfunktionen bara i de mittersta noderna. Värdena på h ges av

$$h_j = g\left(\sum_j w_{jk} x_k\right)$$

I den översta noden används en linjär funktion, och o räknas ut med hjälp av formeln

$$o_i = k\, w_{ij}\, h_j$$

Felfunktionen definierar vi enligt nedan för att kunna använda oss av en mk-funktion med olika vikter beroende på om vi har sålt slut tidningen eller om det blivit tidningar över.

$$E = \begin{Bmatrix} \left(O_i^{(p)} - \text{straff} * t_i^{(p)}\right)^2 & \text{om } O_i^{(p)} > t_i^{(p)} \\ \phi\left(\text{straff} * t_i^{(p)} - O_i^{(p)}\right)^2 & \text{om } O_i^{(p)} < t_i^{(p)} \end{Bmatrix}$$

Index p står för träningsrunda nr (p). Enligt figuren ovan ges O av

$$O_i = k\begin{bmatrix} \omega_{i1} & \omega_{i2} & \omega_{i3} \end{bmatrix}\begin{bmatrix} h_1 \\ h_2 \\ h_3 \end{bmatrix}$$

Om vi sätter in värdena i funktionen g(a) ges formeln för h uttryckt i x enligt

$$\begin{bmatrix} h_1 \\ h_2 \\ h_3 \end{bmatrix} = \frac{1}{2}\left(1 + \tanh\left(\frac{1}{T}\begin{bmatrix} \omega_{11} & \omega_{12} & \omega_{13} \\ \omega_{21} & \omega_{22} & \omega_{23} \\ \omega_{31} & \omega_{32} & \omega_{33} \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}\right)\right)$$

Vid uppdatering av ω-värdena använder vi oss av back-propagation. Gradienten för ω_{ij} beräknas enkelt till

$$\begin{bmatrix} \frac{\partial E}{\partial \omega_{i1}} \\ \frac{\partial E}{\partial \omega_{i2}} \\ \frac{\partial E}{\partial \omega_{i3}} \end{bmatrix} = 2\left(O_i^{(p)}\text{-straff} * t_i^{(p)}\right)\Phi(\text{fel})\, k \begin{bmatrix} h_1 \\ h_2 \\ h_3 \end{bmatrix}$$

För uträkning av utsignalen, o, med insignalerna, x, används

$$O_i = k\begin{bmatrix} \omega_{i1} & \omega_{i2} & \omega_{i3} \end{bmatrix}\frac{1}{2}\left(1 + \tanh\left(\frac{1}{T}\begin{bmatrix} \omega_{11} & \omega_{12} & \omega_{13} \\ \omega_{21} & \omega_{22} & \omega_{23} \\ \omega_{31} & \omega_{32} & \omega_{33} \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}\right)\right)$$

Gradienten för ω_{jk} fås av

$$\begin{bmatrix} \frac{\partial E}{\partial \omega 11} & \frac{\partial E}{\partial \omega 12} & \frac{\partial E}{\partial \omega 13} \\ \frac{\partial E}{\partial \omega 21} & \frac{\partial E}{\partial \omega 22} & \frac{\partial E}{\partial \omega 23} \\ \frac{\partial E}{\partial \omega 31} & \frac{\partial E}{\partial \omega 32} & \frac{\partial E}{\partial \omega 33} \end{bmatrix} =$$

$$= \frac{\left(O_i^{(p)}\text{-straff} * t_i^{(p)}\right)\Phi(\text{fel})\, k}{T}\begin{bmatrix} \omega_{i1} & 0 & 0 \\ 0 & \omega_{i2} & 0 \\ 0 & 0 & \omega_{i3} \end{bmatrix}\cosh^{-2}\left(\frac{1}{T}\begin{bmatrix} \omega_{11} & \omega_{12} & \omega_{13} \\ \omega_{21} & \omega_{22} & \omega_{23} \\ \omega_{31} & \omega_{32} & \omega_{33} \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}\right)\begin{bmatrix} x_1 & x_2 & x_3 \end{bmatrix}$$

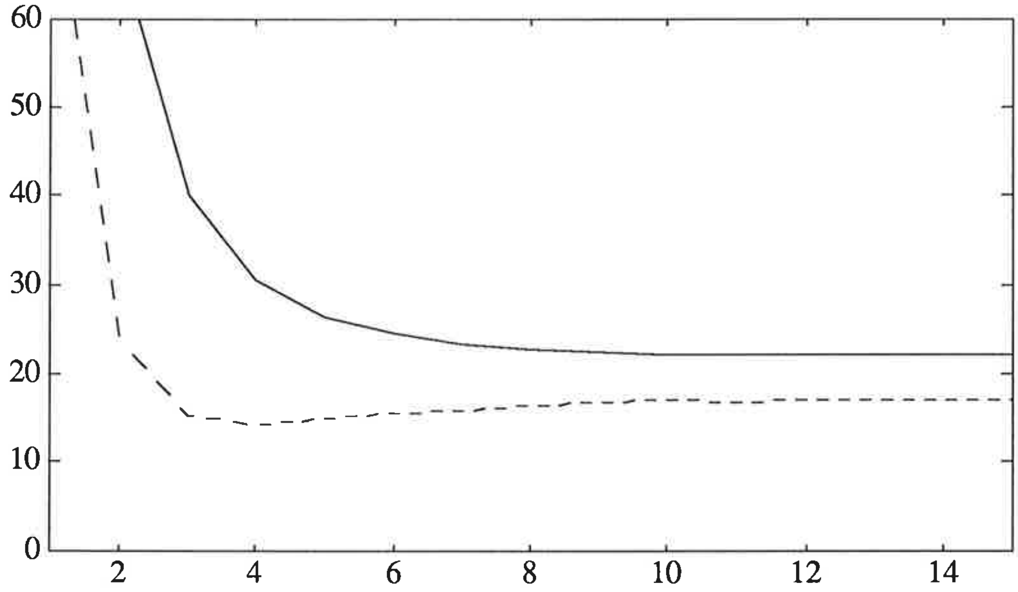
Vi har nu samtliga formler som behövs för att kunna köra igång vårt nätverk.

4.3.4 Extravaganser

Vid implementering av neurala nätverk används en hel del finesser för att få fram bra resultat. I detta kapitel räknar vi upp några av de "extravaganser" vi använt oss av. Det bör tilläggas att detta kapitlet endast är skrivet för den som är intresserad av att närmare gå in i utformningen av ett ANN. Eftersom litteraturen oftast är på engelska har vi valt att behålla en del av de engelska namnen trots vår motvilja mot detta språkliga kaos. Samtliga råd har kommit från Carsten Peterson [1991]

TrainingSet, ValidationSet och ReferenceSet

Vid inlärning av ett neuralt nätverk delar man upp den tillgängliga dataserien i tre separata delar. För inlärning av nätets parametrar behövs en dataserie av storleksordningen tio gånger antalet modell-parametrar. I början på serien definieras ett s.k. TrainingSet (TS). Den används enbart till inlärning. I insvängningskurvan nedan ses prediktionsfelet för TS, heldragen kurva, som en funktion av antal itereringar.

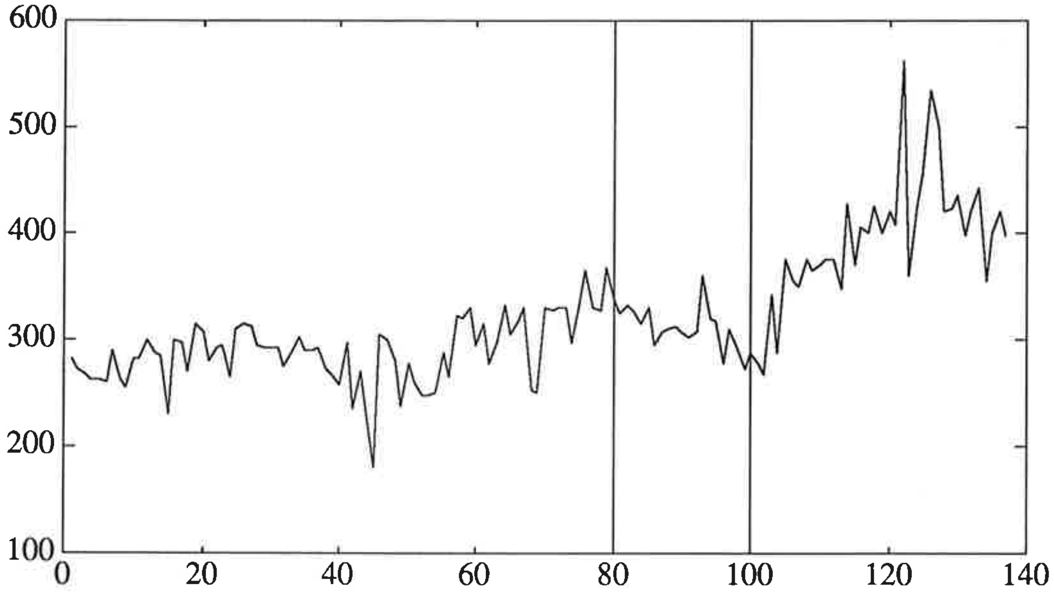


Figur 16. Här syns insvängningsförloppet när felet minskar. I diagrammet ses felet som funktion av antalet itereringar. Streckad linje är ValidationSet medan heldragen är TrainingSet.

ValidationSet (VS) är den mittersta delen av dataserien. Den används för att bestämma antalet itereringar. Prediktionsfelet för VS är den streckade linjen i figuren ovan. Som synes har de båda kurvorna ett liknande utseende i början av inlärningen. Efter ett tag blir felen för VS växande medan felet för TS fortsätter att minska. Detta beror på att vid för lång inlärning av nätet tappar det sin generalitet och blir för mycket anpassat till utseendet på TS. Därför skall man avbryta itereringen när VS har nått sitt minimum.

Den sista delen av dataserien kallas för ReferenceSet (RS) och används för att mäta resultatet av nätets prediktion.

Storleken på TS bör som ovan sagts, vara c:a tio gånger antal parametrar i nätet. VS bör ungefär ha längden 20% av TS, och RS längden 30% av TS.



Figur 17. Här ovan ses indelningen i TrainingSet, ValidationSet och ReferenceSet.

Firing unit

Nästa ingrediens i häxbrygden är en extra insignal till varje nod som hela tiden har värdet 1. Den kallas "Firing unit", eftersom den hela tiden skjuter iväg en etta. Den har som uppgift att hålla summan av insignaler till varje nod på en bra arbetsnivå i sigmoiden. Utan "Firing unit" skulle summan till varje nod bli för låg. Den skulle då arbeta i nedre delen av sigmoiden, där funktionen i princip bara utgörs av en rät linje med liten lutning (se fig. 12).

Lyft

Ett problem som möter oss är att utsignalen från nätet blir för låg. D.v.s. att tidningarna tar slut för ofta. Detta löser vi genom att höja nivån på verkliga data som jämförs med nätets utsignal vid beräkning av prediktionsfelet. Nätet kommer nu konsekvent att gissa på något för stora värden. För att få ett bra värden på retur och slutprocent bör parametern Lyft ha värdet ≈ 1.2. Lyft har också en annan viktig funktion, nämligen upprätthållandet av systemets stabilitet. Eftersom nätet är återkopplat med antal sålda tidigare dagar, finns risk för att systemet stryper sig genom att börja leverera för lite tidningar.

Straff

Straffet är ett kompliment till Lyftet. Prediktionsfelet straffas olika vid inlärning, beroende på om man predikterar för mycket eller för lite tidningar. Straffet varierar fördelningen se resonemanget i kap. 3.1.1. Enligt samma resonemang påverkar lyftet väntevärdet av prediktionen.

Olja

Under inlärning av nätet ställer man parametrarna för varje punkt i dataserien med hjälp av back-propagation. Då felen har stor variation kommer insvängningen av nätet att gå "hackigt". Detta löser vi genom att olja nätet, d.v.s genom att medelvärdesbilda gradienten [Peterson, 1991].

4.3.5 Programvara och realisering

När vi skulle utforma nätet valde vi mellan olika typer av programvara. Vi förkastade idéen med att använda speciella programpaket för ANN p.g.a. nedanstående skäl.

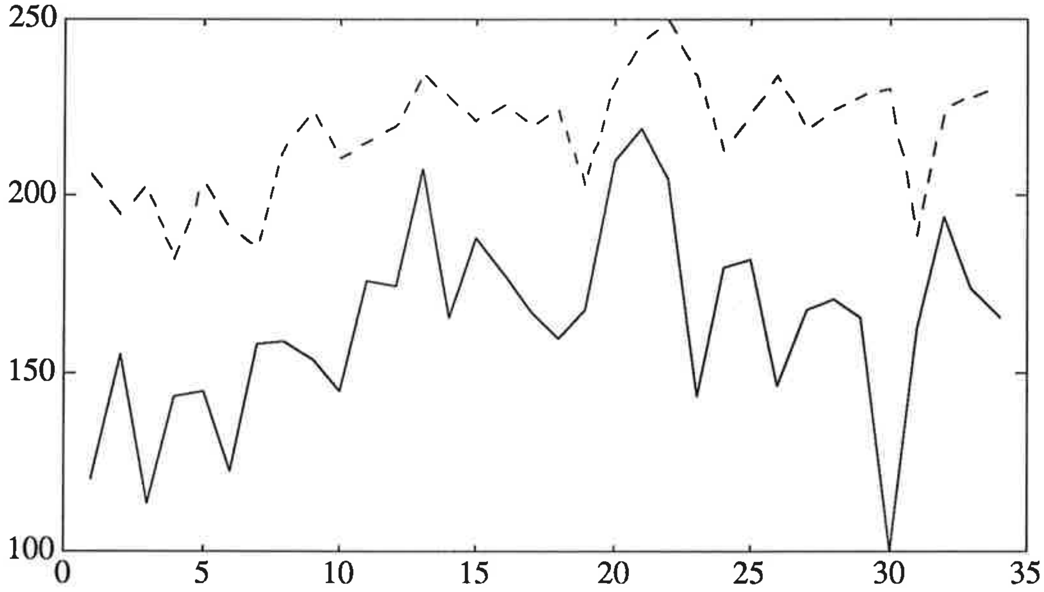
Vi ville begripa vad som hände, och se hur nätet uppför sig.

Det skulle vara billigt att realisera nätet i en större skala.

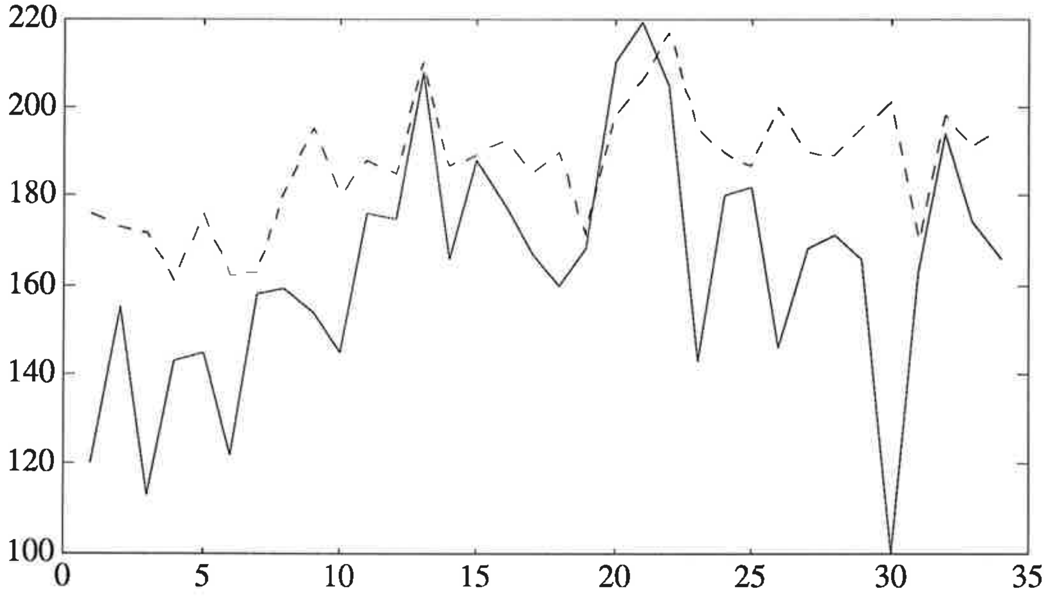
Det programspråk vi fastnade för var MATLAB, vilket har både för- och nackdelar. Den stora fördelen är att det går snabbt och lätt att rita bra bilder. Tyvärr är det ganska långsamt vid skalära beräkningar. När man skriver ett MATLAB program handlar det mycket om att arbeta med matriser, för att på så sätt utnyttja dess styrka att utföra matrisberäkningar. Vi har skrivit programmet så att det snabbt och enkelt skall kunna skrivas om till Fortran eller C++. Detta för att Expressen lätt och billigt skall kunna realisera nätet.

4.4 Prediktioner av Expressen med hjälp av ANN

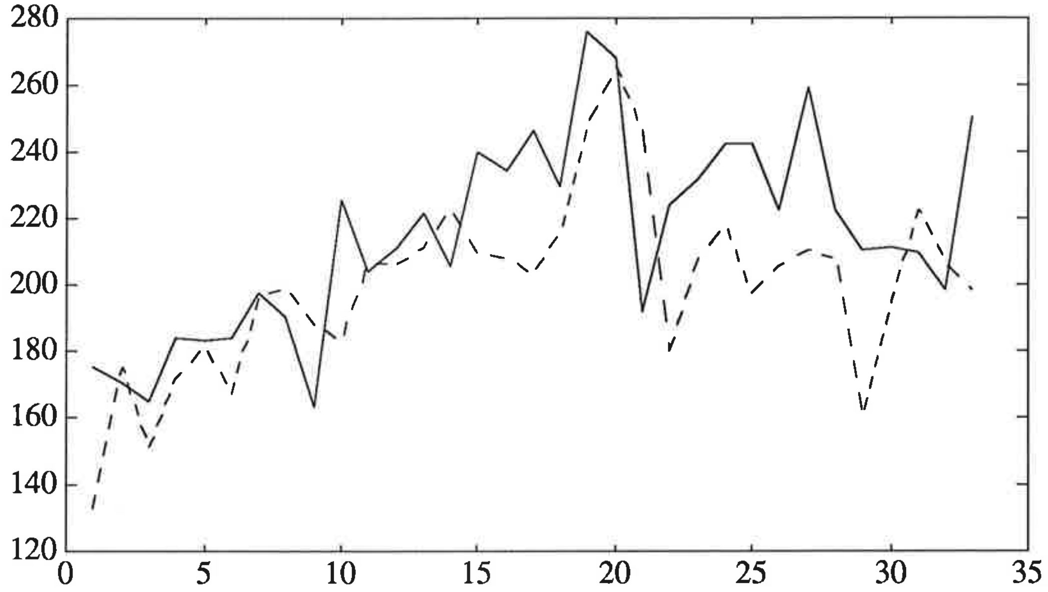
Vi kan nu visa hur prediktionen av Expressens försäljning har utfallit. I detta kapitel visar vi inte enbart de bästa prediktionerna, utan även några exempel på mindre bra prediktioner. Det är svårt att göra tvärsäkra uttalande om hur bra nätet är enbart på dessa bilder och därför presenterar vi resultat i nästa kapitel. I alla bilder är streckade linjer predikterade värden, medan heldragna linjer är försäljningsdata. På x-axeln: tid (dagar), y-axeln: antalet tidningar.



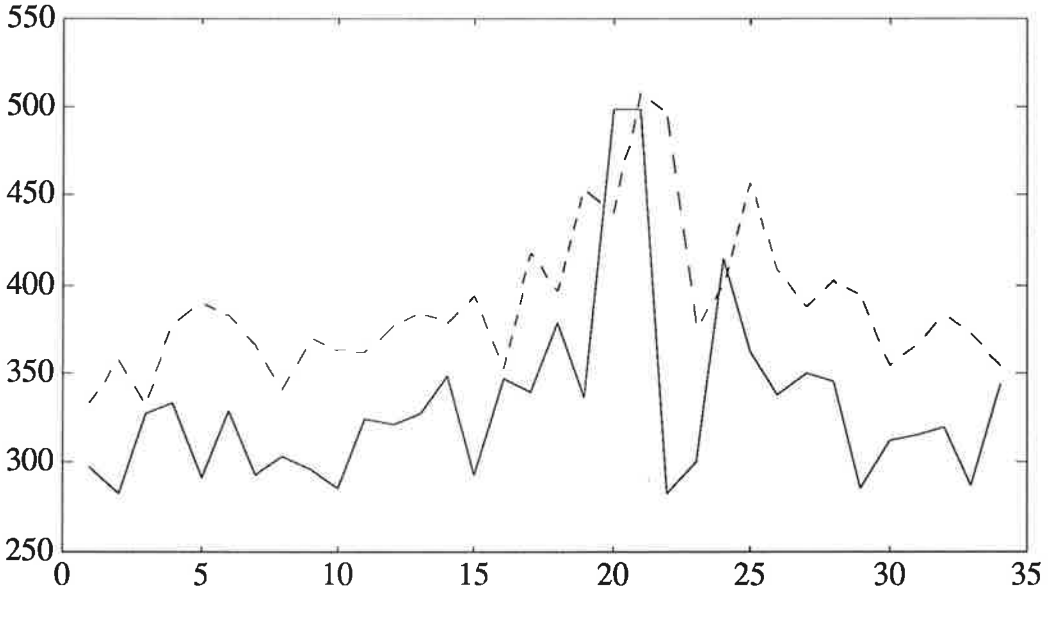
Figur 18. Återförsäljare 2140 fredagar. Här ligger nätet lite väl högt, och det skulle kunna sänkas 2-3 %. Jämför gärna ovanstående prediktion med om vi hade haft en rak linje på c:a 220 tidningar. Då hade resultatet blivit sämre.



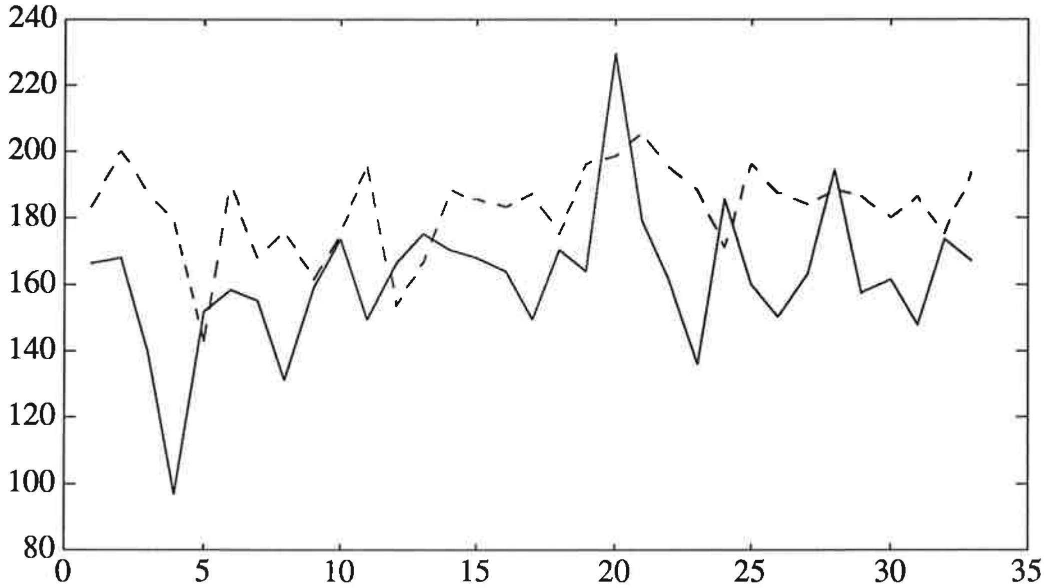
Figur 19. Återförsäljare 2140 fredagar. Här har vi ändrat "lyftet" så att vi kommer längre ner. Det ger en bättre returprocent men här säljs tidningen slut 2-3 gånger.



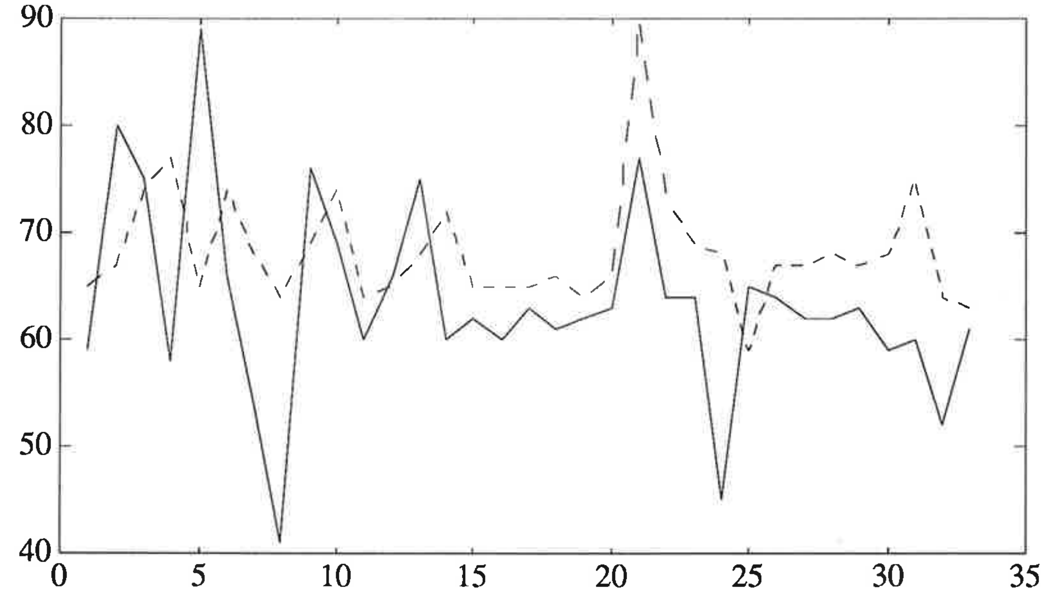
Figur 20. Återförsäljare 2140 lördagar. Här har vi ett exempel på en bra prediktion, men vi ligger under det rätta värdet. Värdet på "lyftet" är för litet. Det blir slutsålt för ofta.



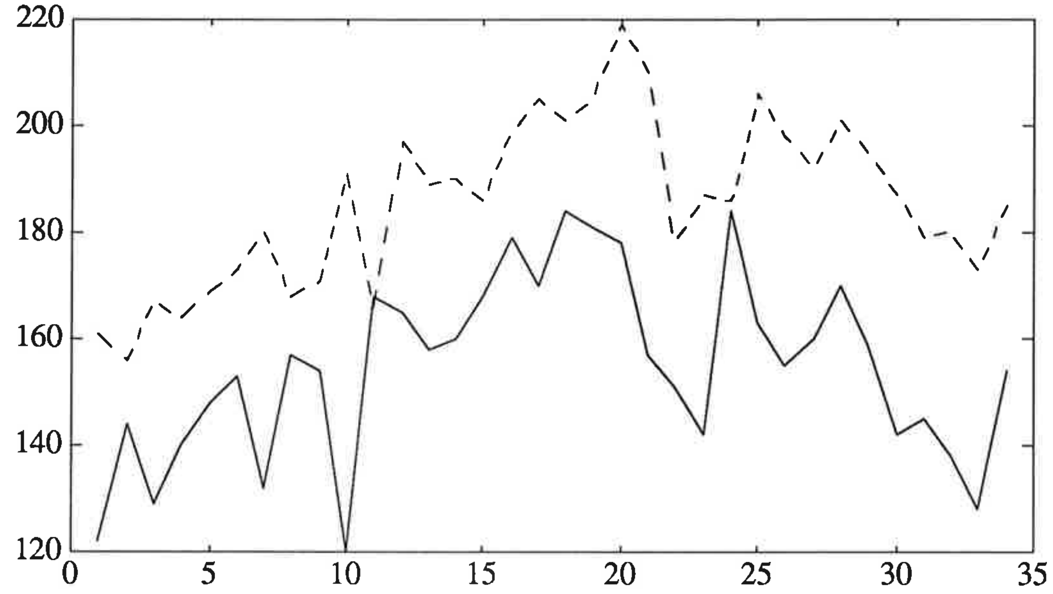
Figur 21. Återförsäljare 2141 måndagar. Nätet hittar de flesta av topparna, men det blir slutsålt i c:a 5-10 % av fallen.



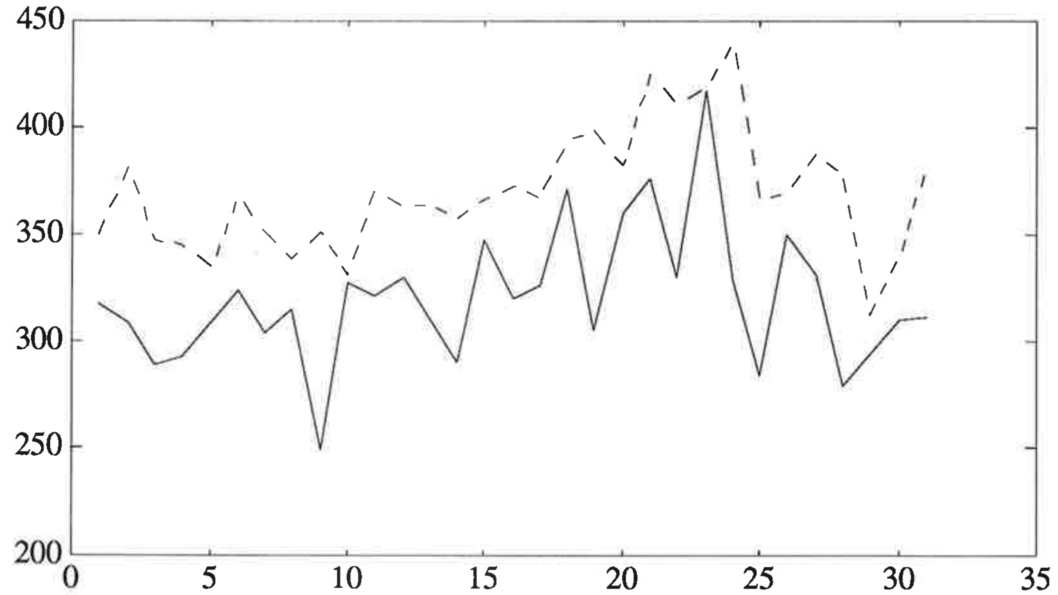
Figur 22. Återförsäljare 2128 onsdagar. Här har vi valt för låg returprocent. Det medför att slutförsäljningsprocenten blir för hög. Ett exempel på en dålig prediktion.



Figur 23. Återförsäljare 2113 torsdagar. Det speciella med den här bilden är inte att vi ligger lite för lågt med leverans, utan att nätet hittar spiken.



Figur 24. Återförsäljare 2140 söndagar.



Figur 25. Återförsäljare 2141 onsdagar.

4.5 Resultat

Vi provade vårt nätverk på 4 återförsäljare 2113, 2128, 2140 och 2141. Dessa återförsäljare är ganska stora och tillhör de 3000 största i landet. Vi har koncentrerat oss på dessa eftersom Expressen har bett oss att göra detta, då de största vinsterna finns att hämta här.

Följande förutsättningar gällde:
Vi körde på data som ej använts vid träningen av nätet. Antalet datapunkter var 33 (antal veckor) * 7 (antal veckodagar) * 4 (antalet återförsäljare) ≈ 920.

Vi provade nätet två gånger med olika "lyft" och fick följande:

Vårt resultat

Returprocent 13.0 % Slutprocent 13.6 %

eller

Returprocent 11.9 % Slutprocent 16.6 %

Expressens resultat

Returprocent 13.9 % Slutprocent 16.7 %

Vid en första betraktelse verkar nätet inte vara mycket bättre än Expressens förutsägelser. Men det finns några punkter som måste beaktas vid en jämförelse.

På de här tre återförsäljarna ligger Expressen betydligt bättre än vad de l igger på riket i genomsnitt. Som nämndes i inledningen är medelvärdet på deras slutprocent omkring 20 % och på returprocenten omkring 23 %.

En minskning av returprocenten med endast 2% innebär en besparing på c:a 10000 tidningar per dag (omkring 3 miljoner tidningar på ett år!).

Ett nätverk behöver mycket mindre tillsyn än ett manuellt system. Det är dessutom mycket lätt att ställa retur- och slutprocenten med ett nätverk så att Expressen kan maximera sin vinst utifrån marginalintäkterna.

Ett automatiskt nätverk kommer medföra att alla återförsäljare kommer att ligga på önskvärd nivå. Det finns ingen chans för någon att dra ner resultaten.

Eftersom datavärdena vi har är trunkerade (återförsäljaren kan ha sålt slut sina tidningar) är det troligt att vi fått ett ännu bättre resultat om vårt nätverk hade använts i verkligheten.

En sista detalj, som kan vara värd att nämna, är att nätet bara blir bättre och bättre med tiden. När nätet har stått och kört något år kommer det att ha lärt sig ännu mer om de olika försäljningsställenas karakteristik.

5 Tillämpning av ANN på andra ekonometriska problem

I det här kapitlet vill vi prova vårt nät på andra typer av prediktioner.Vi har valt ett helt annat område, nämligen fastighetspriser. Vi hämtade data från SCB [1976 - 1990], och tittade på köpeskillingsmedelvärdet av hyreshus, månadsvis i hela riket. Det sker ett par tusen försäljningar av hyreshus i månaden.

5.1 Vad skiljer fastigheter från tidningar

Det skall gå att använda samma metod för så skilda saker som tidnings- och fastighetspris-prediktion. Med fastigheter finns det dock två svårigheter som kan vara värda att nämna.

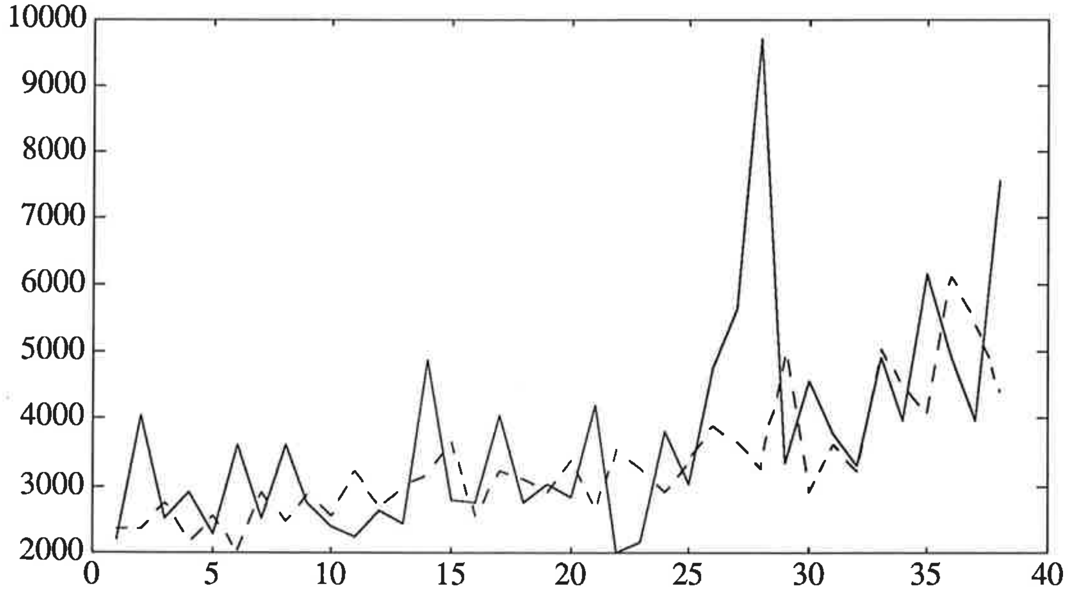
Skillnaden mellan två olika fastigheter. En tidning är en mycket homogen produkt oavsett om den säljs i Skultorp eller Göteborg. En fastighet är däremot en heterogen vara. Det är stor skillnad på en fastighet i utkanten av Nossebro och en fastighet på Östermalm i Stockholm.

Kausalitetsproblem. Ett fastighetsköp behöver inte äga rum den dag då köpet registreras. Köparen och säljaren kan ha kommit överens långt innan.

I tidningsfallet predikterar vi antalet sålda tidningar, men i detta fall predikterar vi priset i kronor vilket beror av inflation och andra ekonomiska faktorer. Dessutom har fastighets marknaden varit utsatt för ett antal störningar i form av statlig inblandning, såsom valuta- och kreditavregleringen.

5.2 Lineär regression

Innan vi provade nätet valde vi att använda oss av enkel lineär regression [Johansson,1991]. Detta gjorde vi för att kunna jämföra mot vår prediktion med ANN. Vi använde oss av femton månaders tidigare fastighetspriser.

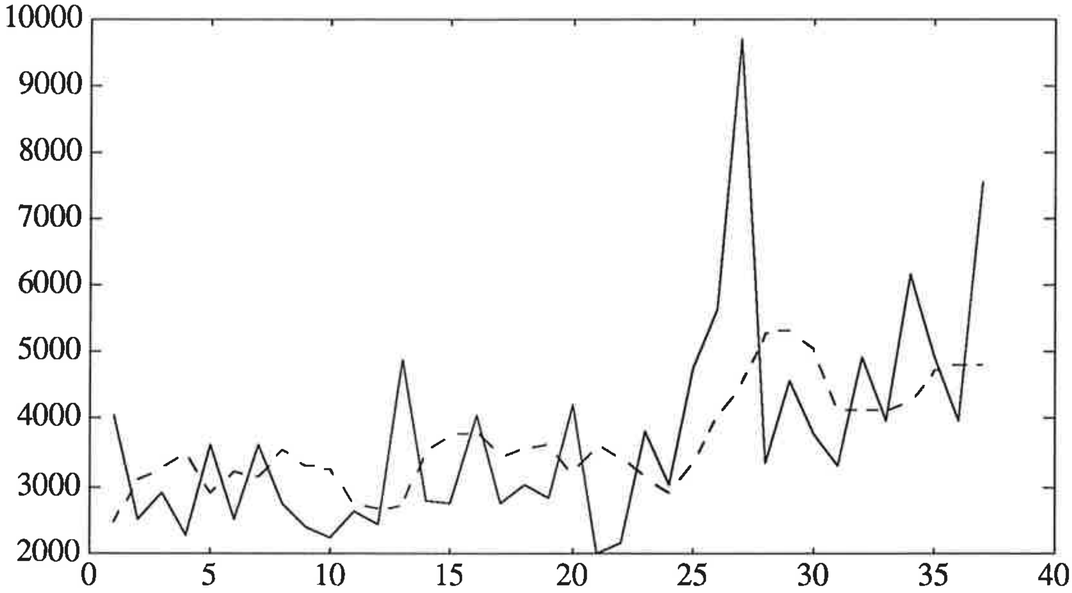


Figur 26. Prediktion med lineär regression. På x-axeln: medelpriset i kkr. På y-axeln: tid (månader). Den heldragna linjen är verkligt fastighetspris. Streckade linjen är predikterat fastighetspris.

Som synes i figur 26 är prediktionen ganska dålig, särskilt om man tar i beaktelse att det bara rör sig om enstegs prediktion.

5.3 Nätverkets prediktion

Vi valde här att använda ett nät med de tre tidigare månaderna som insignaler. Nätet hade dock en lite annorlunda topografi jämfört med nätverket för tidningsprediktion. Vi utökade antalet mittennoder från tre till fem. Prediktionen sker på samma tidsserie som vid linjär regression.

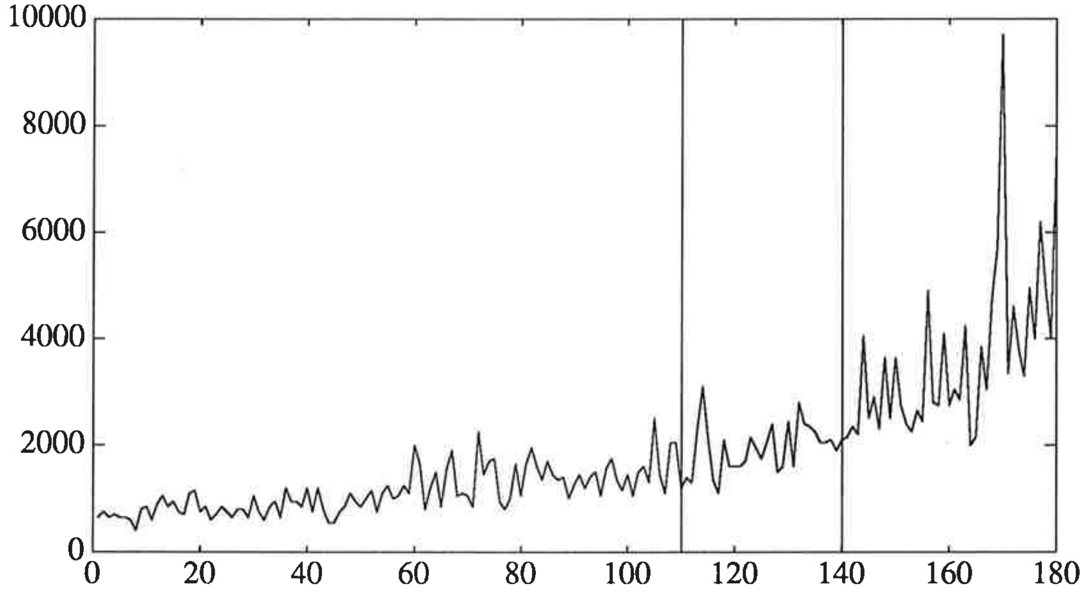


Figur 27. Prediktion av fastighetspriser med ANN. Heldragen linje är verklig data, streckad linje är predikterad.

I figuren ovan syns att prediktionen är bättre än vid användandet av lineär regression. Den kvadratiska summan av residualerna är något mindre jämfört med lineär regression (c:a 30% bättre). Men nätverket klarar inte av att ge en tillfredsställande prediktion.

Varför är nätet så dåligt? Vi tror att det avgörande skälet är bristen på bra insignaler. Ett problem är också att data är "manipulerade". Som ovan nämndes är det först när fastighetsförsäljningen registreras hos myndigheten, som den införs i statistiken.

En annan förklaring kan vara att prisbildningen de tre senaste åren (1988,1989,1990) i vår data var markant annorlunda än föregående år. Det visar sig tydligt i fig. 29, där man ser att tendensen de tre senaste åren är skild från övriga år. Kort sagt, inget beroende - ingen bra prediktion.



Figur 28. Här syns hur fastighetspriset varierar mellan 1976 och 1991. På x-axeln: tid (månader), på y-axeln: priset i kkr. Uppdelningen i TrainingSet,ValSet och RefSet syns som streck i figuren.

6 Slutord

Utvärdering och kommersiell användning av prediktionspaketet.
Utan tvivel finns det en kommersiell användning av ANN för Expressen Till den känsliga frågan "Hur mycket vinst kan ANN ge?" har vi däremot inget entydigt svar.

Att förverkliga metoden kommer inte att medföra några stora kostnader eftersom vi har byggt programmen för att de skall vara lättanpassade. Med t. ex. C++, som programspråk, kommer datautrustningen att vara billig. Installationstiden bör vara ca. två månader. Innan systemet är i full drift kommer det att behövas en inkörningstid på omkring två månader.

Ett av de större problemen är att Expressen i dagens datum inte har något system för inhämtning av försäljningsdata från föregående dag. Detta kan lösas på två olika sätt:

- Ökat användande av telenätet för inhämtning av data.

- Användande av nätverk där gårdagens försäljning ej finns med. Försök på neuronnät som istället utnyttjar försäljningen två veckor tillbaka har givit bra resultat som pekar på en eller två procentenheters sämre resultat jämfört med tidigare.

Ett ANN kan vara till hjälp för många andra företag med liknande problem. Några förslag på detta är övrig tidningsverksamhet och livsmedelsindustrin (framför allt för mejeri- och bröd-produkter).

Användandet av ANN på andra ekonometriska data .
När vi ovan provade nätet på fastighetspriserna är det bara att erkänna att vi misslyckades. Visst blev det c:a 30% bättre än vid lineär regression, men resultatet kan ej användas. Att det blev så berodde nog snarare på ett olyckligt val av problem än på att nätet ej är dugligt för sådana problem. Vi tror att det finns stora möjligheter för ekonomer att använda sig av ANN. Vi ser därför med glädje på att ekonomihögskolan i Lund har börjat seminarier med inbjudna gäster från Stanford University, USA, om användandet av neurala nätverk inom ekonometrin. Utsikterna att få goda resultat vid prediktion av andra ekonometriska data verkar goda.

6.1 Tack

Vi vill framför allt tacka vår handledare Rolf Johansson, Institutionen för Reglerteknik (LTH), för hans stöd. Vi uppskattar verkligen att han gav oss chansen att få kombinera ekonomi och reglerteknik i vårt examensarbete.

Vi vill också tacka Carsten Peterson, Institutionen för Teoretisk Fysik (LU), för hjälp med nätverken, Curt Wells, Nationalekonomiska Institutionen (LU), för hjälp med det ekonomiska och Joakim Claesson, Upplagechef Expressen, för hjälp med problem rörande tidningsvärlden.

Följande personer har givit värdefulla råd och synpunkter:
professor Karl Johan Åström Institutionen för Reglerteknik , professor Ingemar Ståhl, Nationalekonomiska Institutionen (LU) och professor Georg Lindgren, Avdelningen för Matematisk Statistik (LTH).

Vi har även fått hjälp av Bo Bernhardsson, Anders Blomdell, Leif Braun, Eva Dagnegård, Mats Graffman, Jan Holst, Ulla Holst och av alla som kan UNIX och ftp. Tack!

Till slut måste vi tacka två personer som har fått stå ut med många sena kvällar och tidiga morgnar. Tack Emma och Nina !

Lund den 29/5 1992

Jens Henriksson　　　　Björn Törnqvist

Referenser

Bermin, H.P och Gyllerup, M. (1991): "Cointegration och Tillståndsmodellering av Icke-Stationära Tidsserier - Tillämpning på svenska aktier," Avd. för Matematisk Statistik, LTH, Lund

Blom, G. (1984): "Statistikteori med tillämpningar," Studentlitteratur, Lund

Chen, S. ,Billings, S.A. och Grant, P.M.(1989), "Non linear system identification using neural networks," Int. J. Control, 1990 vol 51, no 6

Expressen (1992): "Årsredovisning 1991," Stockholm

Forsythe, G. ,Malcolm, M. och Moler, C. (1977): "Computer Methods for Mathematical Computations," Prentice-Hall, USA

Grace, A. (1990): "Optimization toolbox," MathWorks Inc., MA, USA

Hederstierna, A. (1981): "Decisions under uncertainty," Economic Research Institute,Stockholm

Johansson, R. (1991): "Processidentifiering," Inst. för Reglerteknik, LTH, Lund

Matlab (1991): "Users guide," MathWorks Inc.

Ny Teknik (1991), artikel om Neurala Nätverk nr 20 1991

Olbjer, L. (1985): "Tidsserieanalys," Avd. för Matematisk Statistik, LTH, Lund

Peterson, C. (1989): "Track Finding with Neural Networks," Nuclear Instruments and Methods A279, 537

Peterson, C. ,Ohlsson, M. och Söderberg, B. (1992): "Neural Networks for Optimization Problems with Inequality Constraints - the Knapsack Problem," Inst för Teoretisk Fysik, LU, Lund,

Peterson, C. och Rögnvaldsson, T. (1991): "An Introduction to Artificial Neural Networks," Inst för Teoretisk Fysik, LU, Lund,

Rignér, B. (1991): "Hidden Correlation and Cointegration in Time Series Analysis," Avd. för Matematisk Statistik, LTH, Lund

Rosenblatt, F. (1961): "Principles of Neurodynamics: Perceptrons and the Theory of Brain Mechanisms," Spartan Books, Washington D.C., USA

SCB (1991) "S M 1991 P18," Statistiska Central Byrån, Stockholm

Appendix

```
%         NEURALT NATVERK med önskat antal noder
%         + 2 extra noder som ständigt har värdet 1.
%
%
%    Tar det optimala w-värdena for prediktion.
%                oi          linjär
%           / / | \ \
%            / / / \ \
%         hj ... hj   1      sigmoid
%
%              X  X
%
%           xk xk xk  1      indata
%
%
%    Utsignaler:  KundNrDagwij och KundNrDagwjk
%
%
%
%
%
%
%    Copyright  Jens  Henriksson  &  Bjorn  Tornqvist
%                     920506
%
%    Lunds  Tekniska  Högskola  Inst.  för  ReglerTeknik

len = input('Ange antal noder som onskas');
KundNr = input('Ange aterforsaljarnr med fyra siffror','s');
KundNr = ['0' KundNr];
DagNr = input('Ange dagnr 1-7 ');
antal = input('Ange maximalt antal itereringar ');
ny = input('Ange aeta');


% Finns startvarden for given aterforsaljare i Matlab?

DagNr = rem(DagNr,7);
veckodagar = ['sonmantisonstorfrelor'];
Dag = veckodagar(DagNr*3+1:DagNr*3+3);
eval(['start =exist(''XKMatris',Dag,KundNr,''');'])
if start ~= 1,
  disp(' Anvander STARTVARDEN ')
  startvarden;
end;


% Slumpar ut vardena pa wij och wjk

wjk = [rand(len,3) - rand(len,1)];
wij = [rand(1,len) - rand];


% I tpvektor lagras x(k) Det används som TrainingSet.
% I xkmatris finns [ x(k-1) x(k-7) palagg(k) ]
eval(['xkmatris = XKMatris',Dag,KundNr,';']);
eval(['tpvektor = TPVektor',Dag,KundNr,';']);

% I vsvektor lagras x(k) Det används som ValidationSet
```

```
% I xkvalmatris finns [ x(k-1) x(k-7) palagg(k) ]
eval(['xkvalmatris = XKvalMatris',Dag,KundNr,';']);
eval(['vsvektor = VSVektor',Dag,KundNr,';']);

% I rsvektor lagras x(k) Det används som ReferenceSet
% I xkrefmatris finns [ x(k-1) x(k-7) palagg(k) ]
eval(['xkrefmatris = XKrefMatris',Dag,KundNr,';']);
eval(['rsvektor = RSVektor',Dag,KundNr,';']);


% Nerskalningar för att det skall passa nätet

maximum = max(max(xkmatris));
xkmatris = [xkmatris(:,1:2)/maximum xkmatris(:,3)/100 ];
xkvalmatris = [xkvalmatris(:,1:2)/maximum xkvalmatris(:,3)/100 ];
xkrefmatris = [xkrefmatris(:,1:2)/maximum xkrefmatris(:,3)/100 ];
tpvektor = tpvektor'/maximum;
vsvektor = vsvektor'/maximum;
rsvektor = rsvektor'/maximum;


% Dessa varden bestammer retur och slutprocenten
%##################################################################
straff = 0;               % MK-funktion straff pa värden<0
lyft = 1.24;              % Önskad returprocent
olja = 4;           % Medelvärdesbildning vid back-propagation
vardering = 1.1;          % Vad vi gissar att vi skulle sålt extra
T = 1,                    % Temperaturen
%##################################################################


% Initieringar

clear fellen;
iter = 1;
summa = 0;
felsumma = 0;
maxfel = 0;
maxfell = 0;


% Träning av nätet

while iter<= antal;

  smorjvarv = 0;
  dwij = 0;
  dwjk = 0;
  fel = 0;
  for dag =1:length(xkmatris),
    smorjvarv = smorjvarv + 1;
    xk = [xkmatris(dag,:)';1];
    z = exp(2*wjk*xk/T);             % z = exp(2a)
    hj = [ z./(z+1); 1];             % 1+tanh(a)=2z/(z+1)
    oi = wij*hj;
    steg =ny * (oi-lyft*tpvektor(dag));
    fi = 1 + straff*(oi<tpvektor(dag));
    dwij = dwij + 2*(steg*fi*hj)';
    coshyp =  (exp(wjk*xk/T) + exp(-wjk*xk/T));
    dwjk = dwjk + steg*fi/T*diag(wij(1:len))*(coshyp.^(-2))*xk';
    if smorjvarv == olja,
      wij = wij - dwij/olja;
      wjk = wjk - dwjk/olja;
```

```
        dwij = 0;
        dwjk = 0;
        smorjvarv = 0;
      end;
      oi = round(wij*hj*maximum);
      fel = fel+(oi-tpvektor(dag)*maximum)*(oi-tpvektor(dag)*
                                                  maximum>0);
      fel = fel-(oi-tpvektor(dag)*maximum)*10*(oi-tpvektor(dag)*
                                                  maximum<0);
    end;
    maxfel = fel*(iter==1)+ maxfel;    % ger initialvärde at maxfel
    felen(iter) = fel/length(xkmatris);% ger vardet i %


  % Följande sekvens är till for det s.k ValidationSet
  % Felen räknas ut med olika vikter pa positiva och negativa fel.
    fell = 0;
    for dag =1:length(xkvalmatris)
      xk = [xkvalmatris(dag,:)';1];
      z = exp(2*wjk*xk/T);              % z = exp(2a)
      hj = [z./(z+1); 1];               % 1+tanh(a)=2z/(z+1)
      oi = round(wij*hj*maximum);
      fell = fell+(oi-vsvektor(dag)*maximum)*(oi-vsvektor(dag)*
                                                  maximum>0);
      fell = fell-(oi-vsvektor(dag)*maximum)*10*(oi-vsvektor(dag)*
                                                  maximum<0);
    end;
    maxfell = fell*(iter==1) + maxfell;
    fellen(iter) = fell/length(xkvalmatris);     % ger vardet i %
  % Slut pa ValidationSet


  % Nu skall de optimala wij och wjk sparas.
    if iter==1,
      fellopt = fellen(iter);
    end;
    if fellen(iter)<=fellopt,
      fellopt = fellen(iter);
      wijopt = wij;
      wjkopt = wjk;
    end;
  % De optimala värdena finns nu i wijopt och wjkopt.


  % Rita upp felen for val-set och training-set
    plot([1:length(fellen)],fellen,'--',[1:length(felen)],felen);
    pause(0.05);
    iter = iter + 1
  end;


  % Nu ar wij wjk uträknade Nästa del av programmet visar enbart
  % upp vilka resultat som uppnåtts Inga beräkningar av vikt göres


  % Initieringar

  fel = 0;
  slut = 0;
  predikt = 0;
  rsvektor = rsvektor*maximum;
```

```
% Beräkning av hur prediktorn skulle sett ut.
% De optimala w-vardena anvands i beräkningen.

for dag =1:length(xkrefmatris),
  xk = [xkrefmatris(dag,:)';1];
  z = exp(2*wjkopt*xk/T);              % z = exp(2a)
  hj = [z./(z+1);1];                   % 1+tanh(a)=2z/(z+1)
  oi = round(wijopt*hj*maximum);
  if rsvektor(dag)>=oi-1,
    slut = slut + 1;
    xkrefmatris(dag+1,2) = oi*vardering/maximum;     % Hur många
                                 tidningar tror vi att vi kunde sålt?
  end;
  fel = fel+(oi-rsvektor(dag))*(oi>rsvektor(dag));
  predikt(dag) = oi;
end;


% Visa upp resultatet

disp (' Foljande uppnaddes ')
predikt = predikt';
plot([1:length(predikt)],predikt,'--',[1:length(rsvektor)],
                                        rsvektor);

% Vad vi far for retur och felprocent.

sumfel = sum((predikt-rsvektor).*((predikt-rsvektor)>0));
% summan av felen
sumpred = sum(predikt);       % summan av antal levererade
returprocent = sumfel/sumpred*100
disp('Anger hur manga procent av tidningarna som kom i retur ')
slutprocent = slut/dag*100
disp('Anger hur manga procent av dagarna Expressen salde slut
                                        eller 1 kvar)')

% Spara värdena i Matlab

eval(['kviwij',KundNr,Dag,'=wij;']);
eval(['kviwjk',KundNr,Dag,'=wjk;']);
eval(['kvipred',KundNr,Dag,'= [ rsvektor predikt];']);
eval(['kvisumpred',KundNr,Dag,'=sumpred;'])
eval(['kvisumfel',KundNr,Dag,'=sumfel;'])
eval(['kvislproc',KundNr,Dag,'=slutprocent;'])

% Släng alla variabler

clear antal; clear coshyp; clear dag; clear Dag; clear DagNr;
clear dwij; clear dwjk; clear fel; clear felen; clear fell; clear
fellen; clear fellopt; clear felsumma; clear fi; clear hj; clear
iter; clear KundNr; clear len; clear lyft; clear maxfel; clear
maxfell;  clear maximum; clear medelfel; clear ny; clear oi;
clear olja; clear predikt; clear returprocent; clear rsvektor;
clear slut; clear slutprocent; clear smorjvarv; clear start;
clear steg; clear straff; clear summa; clear sumfel; clear
sumpred; clear T; clear tpvektor; clear vardering;  clear
veckodagar; clear vsvektor; clear wij; clear wijopt; clear wjk;
clear wjkopt;  clear xk; clear xkmatris; clear xkrefmatris; clear
xkvalmatris; clear z;

% Slut på programmet
```