Data Mining: Neural Network Applications

by Louise Francis

CAS Annual Meeting, Nov 11, 2002

Francis Analytics and Actuarial Data Mining, Inc.
louise_francis@msn.com

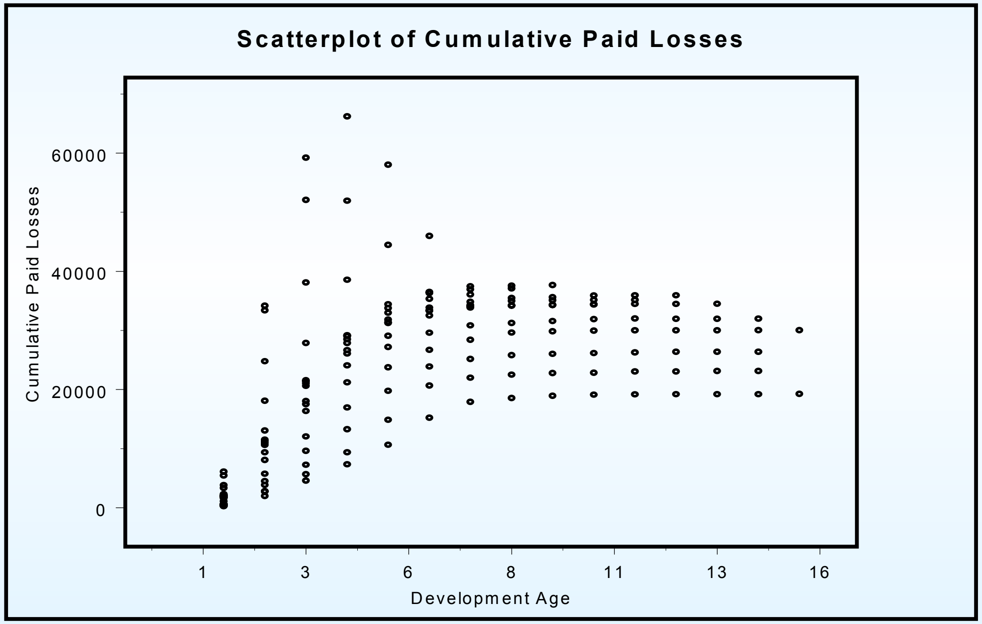
Objectives of Presentation

- Introduce insurance professionals to neural networks
- Show that neural networks are a lot like some conventional statistics
- Indicate where use of neural networks might be helpful
- Show practical examples of using neural networks
- Show how to interpret neural network models

A Common Actuarial Application: Loss Development

Accident	Months of Development														
Year	12	24	36	48	60	72	84	96	108	120	132	144	156	168	180
1980	267	1,975	4,587	7,375	10,661	15,232	17,888	18,541	18,937	19,130	19,189	19,209	19,234	19,234	19,246
1981	310	2,809	5,686	9,386	14,884	20,654	22,017	22,529	22,772	22,821	23,042	23,060	23,127	23,127	23,127
1982	370	2,744	7,281	13,287	19,773	23,888	25,174	25,819	26,049	26,180	26,268	26,364	26,371	26,379	26,397
1983	577	3,877	9,612	16,962	23,764	26,712	28,393	29,656	29,839	29,944	29,997	29,999	29,999	30,049	30,049
1984	509	4,518	12,067	21,218	27,194	29,617	30,854	31,240	31,598	31,889	32,002	31,947	31,965	31,986	
1985	630	5,763	16,372	24,105	29,091	32,531	33,878	34,185	34,290	34,420	34,479	34,498	34,524		
1986	1,078	8,066	17,518	26,091	31,807	33,883	34,820	35,482	35,607	35,937	35,957	35,962			
1987	1,646	9,378	18,034	26,652	31,253	33,376	34,287	34,985	35,122	35,161	35,172				
1988	1,754	11,256	20,624	27,857	31,360	33,331	34,061	34,227	34,317	34,378					
1989	1,997	10,628	21,015	29,014	33,788	36,329	37,446	37,571	37,681						
1990	2,164	11,538	21,549	29,167	34,440	36,528	36,950	37,099							
1991	1,922	10,939	21,357	28,488	32,982	35,330	36,059								
1992	1,962	13,053	27,869	38,560	44,461	45,988									
1993	2,329	18,086	38,099	51,953	58,029										
1994	3,343	24,806	52,054	66,203											
1995	3,847	34,171	59,232												
1996	6,090	33,392													
1997	5,451														

An Example of a Nonlinear Function



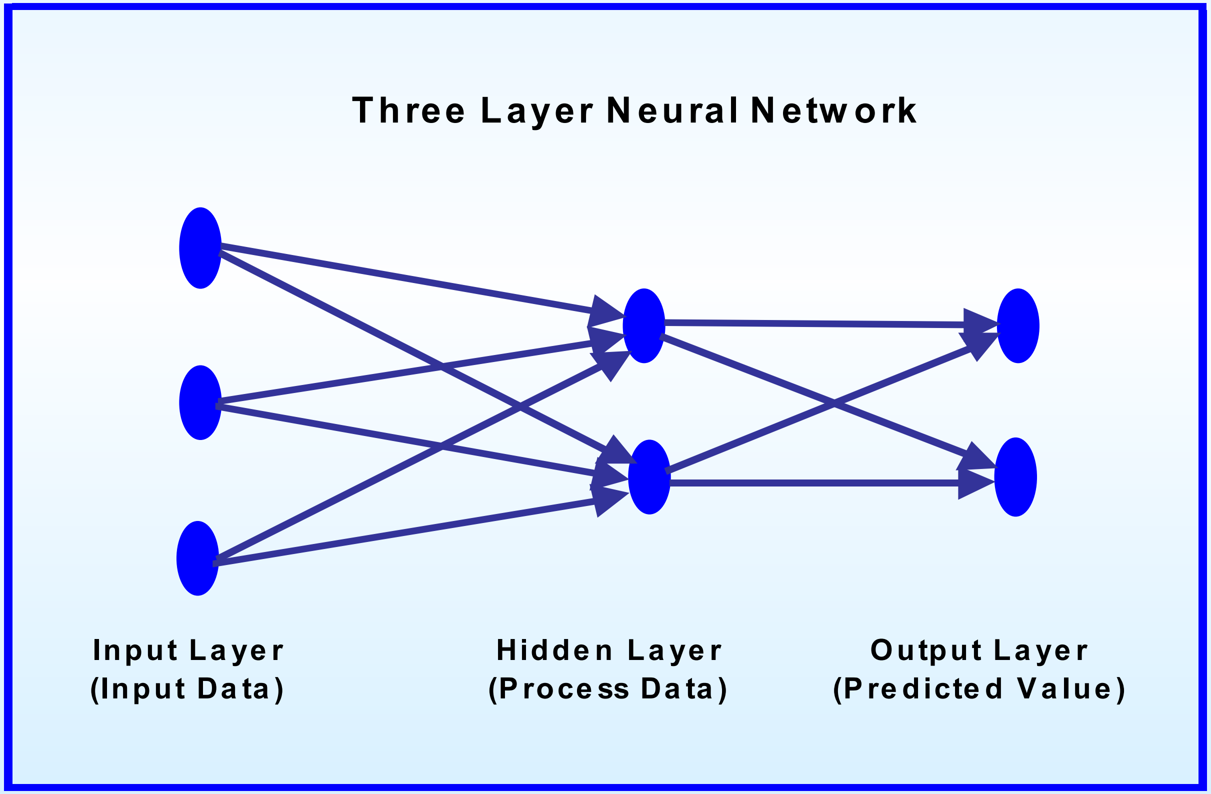
Conventional Statistics: Regression

- One of the most common methods of fitting a function is linear regression
- Models a relationship between two variables by fitting a straight line through points
- Minimizes a squared deviation between an observed and fitted value

Neural Networks

- Also minimizes squared deviation between fitted and actual values
- Can be viewed as a non-parametric, non-linear regression

The Feedforward Neural Network



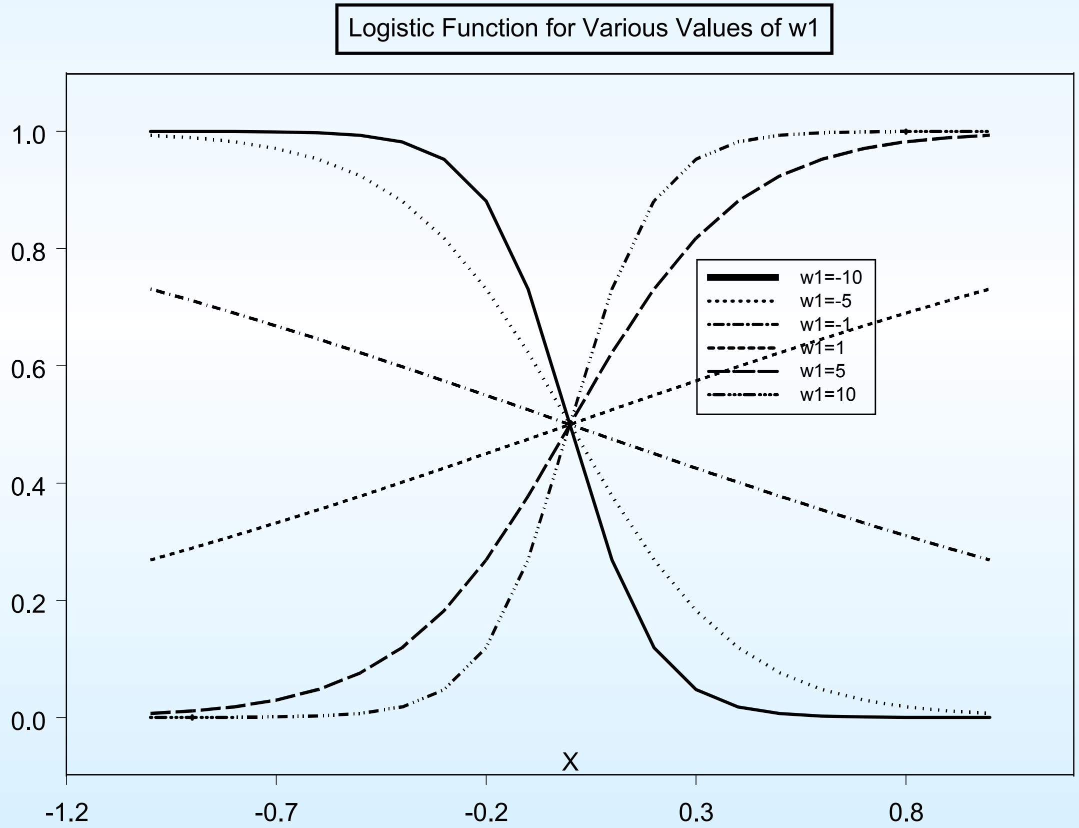
The Activation Function

- The sigmoid logistic function

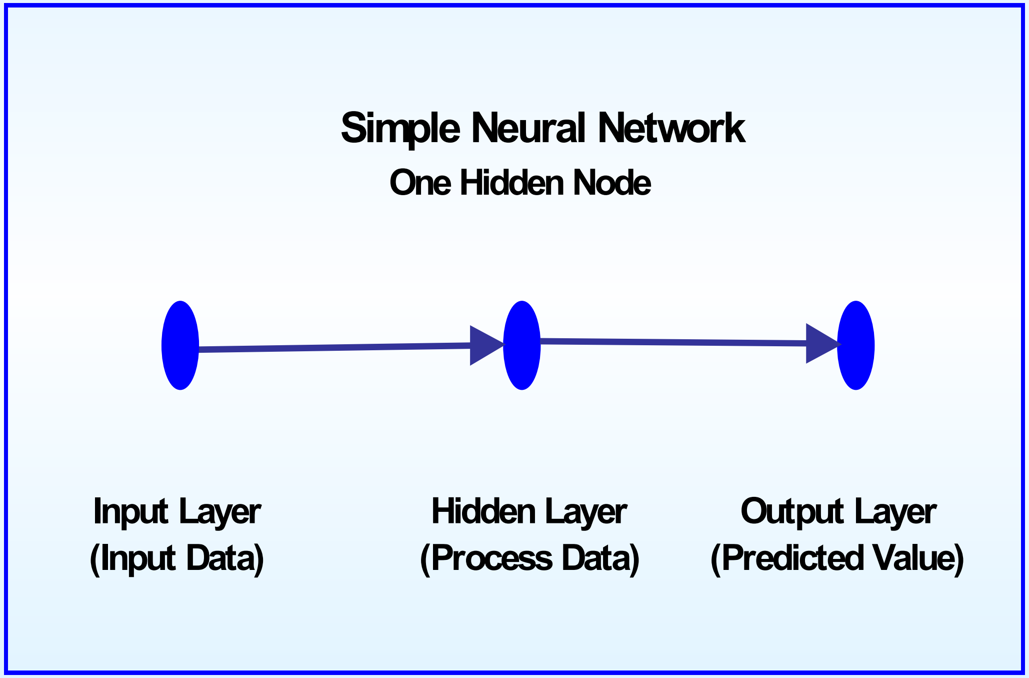
$$f(Y) = \frac{1}{1 + e^{-Y}}$$

$$Y = w_0 + w_1 * X_1 + w_2 X_2 ... + w_n X_n$$

The Logistic Function



Simple Example: One Hidden Node

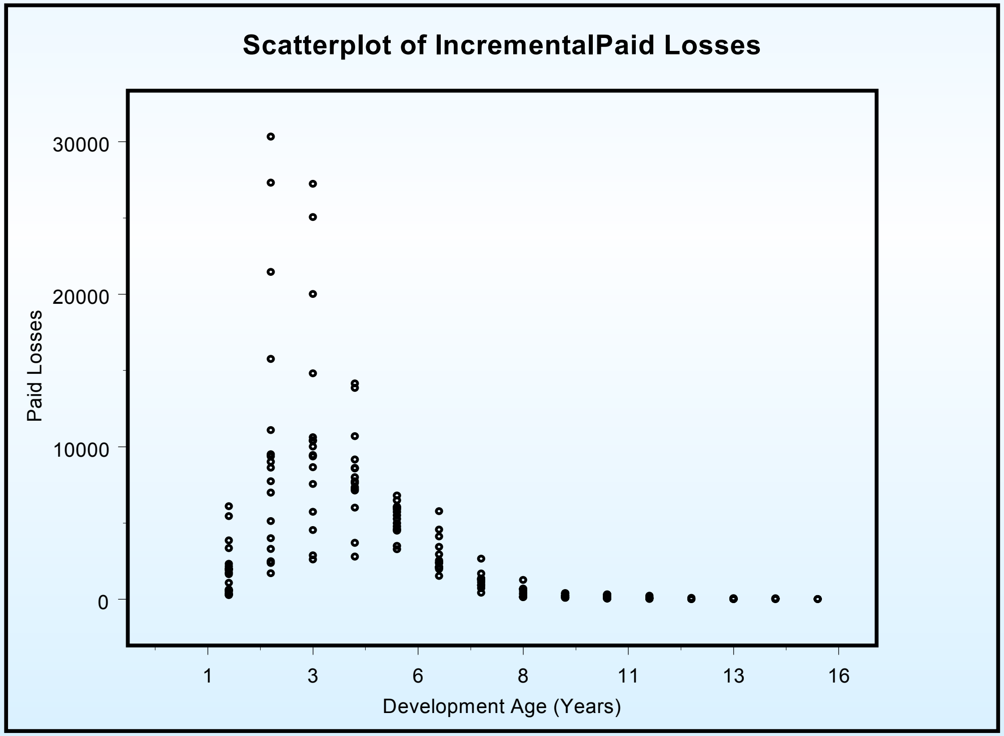


Function if Network has One Hidden Node

$$h = f(X; w_0, w_1) = f(w_0 + w_1 X) = \frac{1}{1 + e^{-(w_0 + w_1 X)}}$$

$$f(f(X; w_0, w_1); w_2, w_3) = \frac{1}{1 + e^{-\left(w_2 + w_3 \frac{1}{1 + e^{-w_0 + w_1 X}}\right)}}$$

Development Example: Incremental Payments Used for Fitting



Two Methods for Fitting Development Curve

- Neural Networks
  - Simpler model using only development age for prediction
  - More complex model using development age and accident year
- GLM model
  - Example uses Poisson regression
  - Like OLS regression, but does not require normality
  - Fits some nonlinear relationships
  - See England and Verrall, PCAS 2001

The Chain Ladder Model

Cumulative paid:

$$D_{ij} = \sum_{k=1}^{j} C_{ik}$$

Age to age factor:

$$\lambda_{ij} = \frac{D_{i,j+1}}{D_{ij}}$$

Estimate of age to age factor using mean:

$$\lambda_j = \frac{\sum_{i=1}^{n} \lambda_{ij}}{n}$$

Common Approach: The Deterministic Chain Ladder

Estimate of paid at 24 months:

$$C_{24} = D_{12}\lambda_{12} - D_{12}$$

Estimate of Ultimate Paid:

$$D_{iu} = D_{ij} \prod_{k=j}^{u} \lambda_{ik}$$

GLM Model

A Stochastic Chain Ladder Model

Poisson Model:

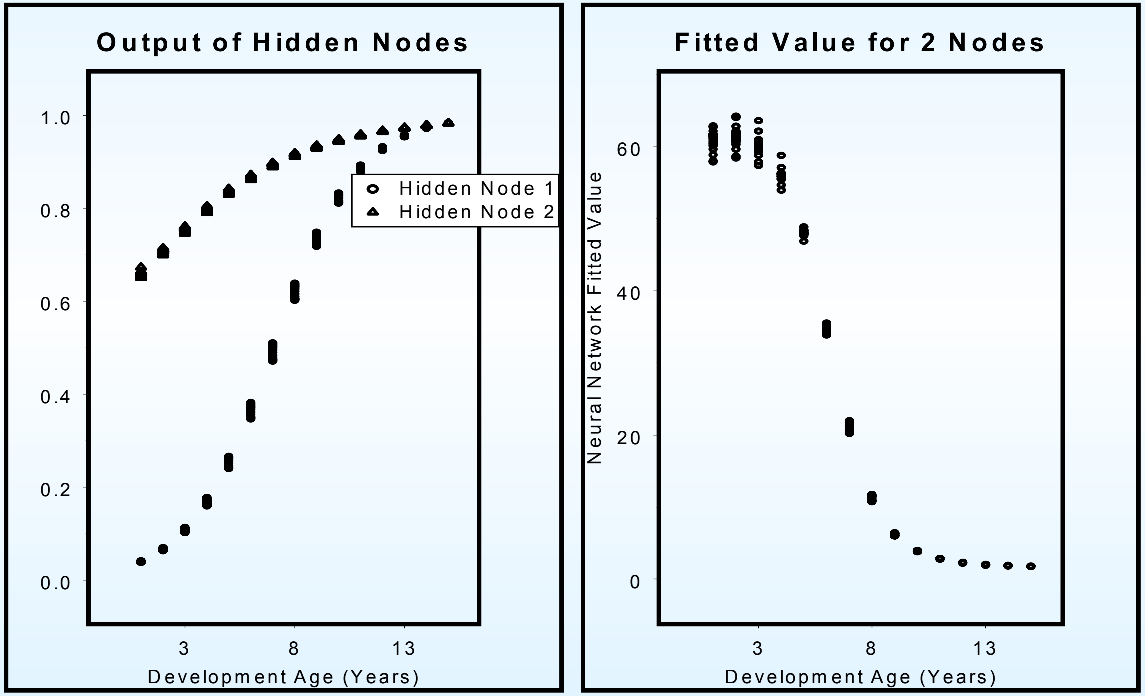
$$E(C_{ij}) = m_{ij} = x_i y_j$$

$$Var[C_{ij}] = \phi x_i y_j$$

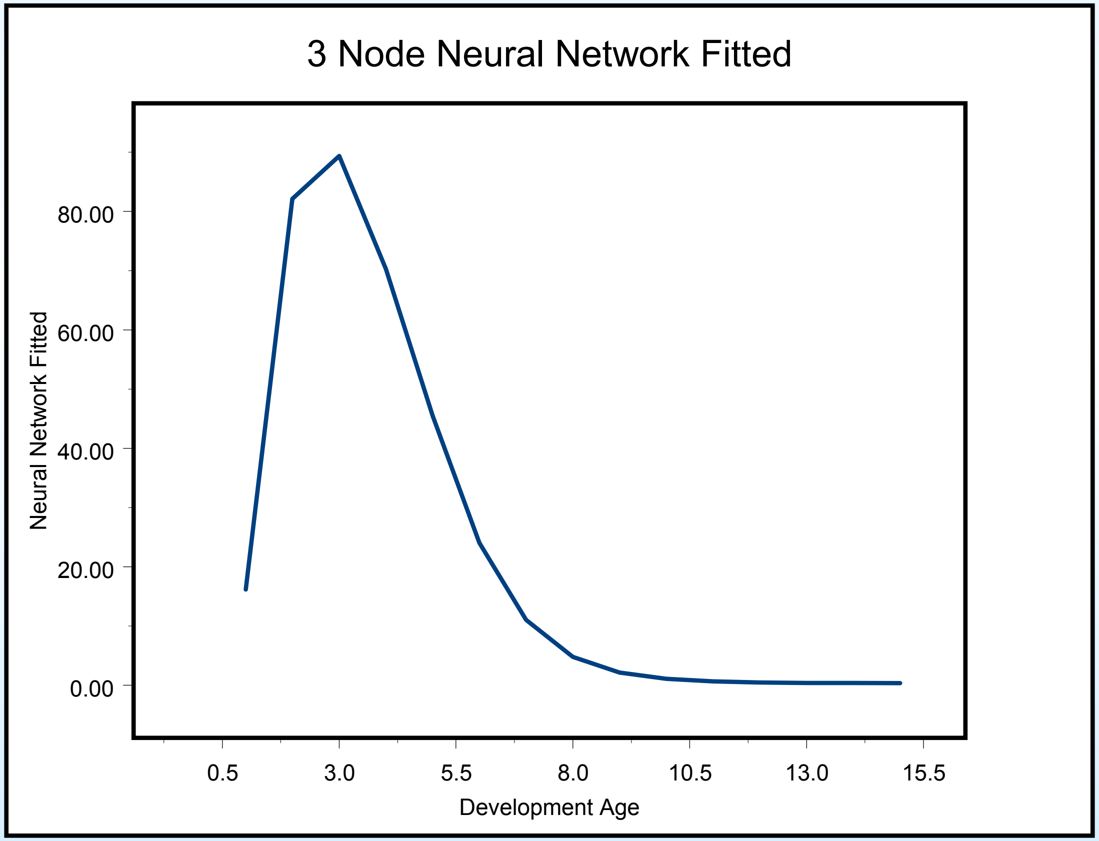
$$\sum_{k=1}^{n} y_k = 1$$

Data often normalized by dividing by an exposure base

Hidden Nodes for Paid Chain Ladder Example



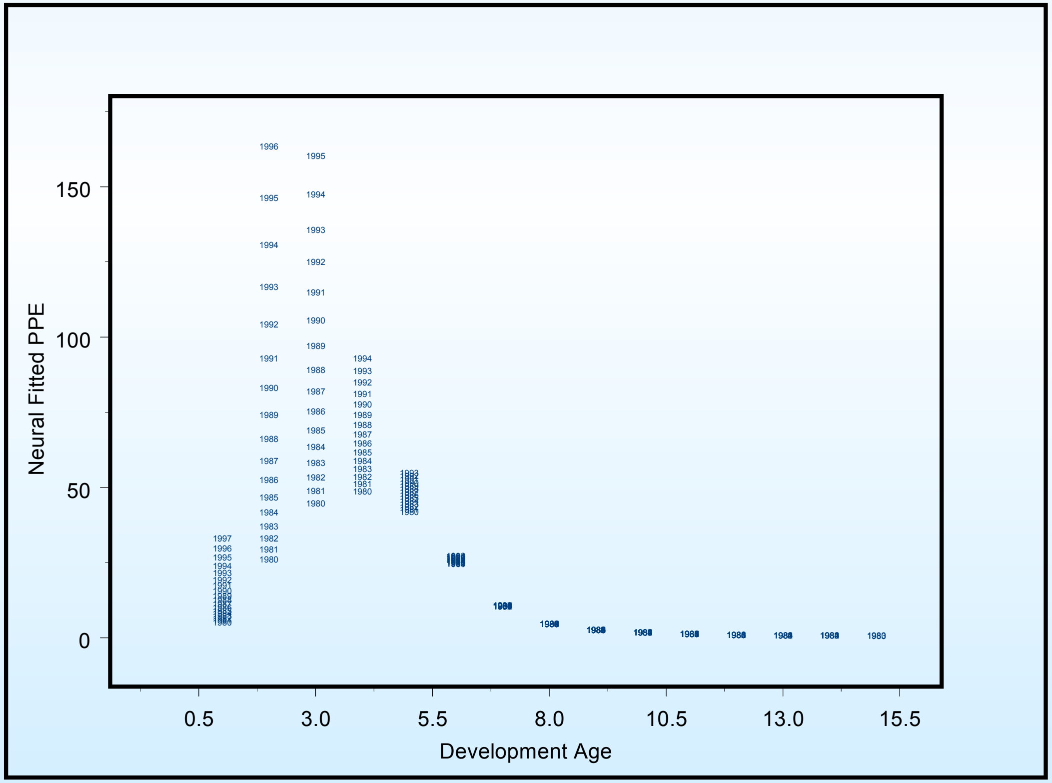
NN Chain Ladder Model with 3 Nodes



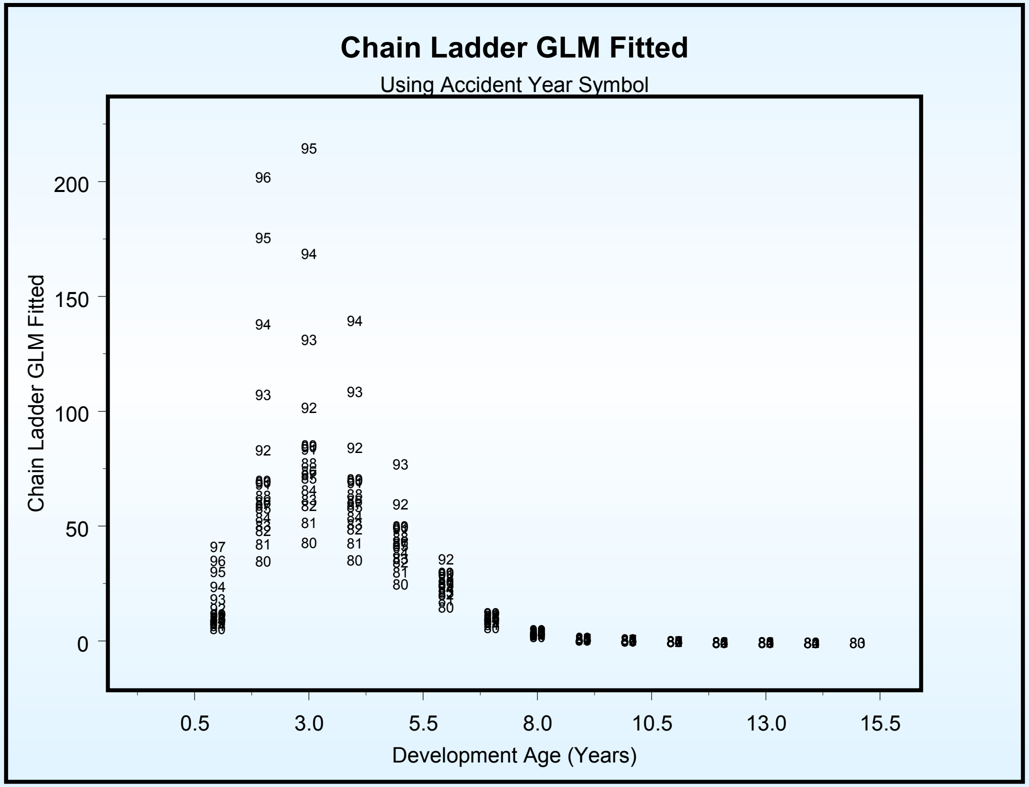
Universal Function Approximator

- The feedforward neural network with one hidden layer is a universal function approximator
- Theoretically, with a sufficient number of nodes in the hidden layer, any continuous nonlinear function can be approximated

Neural Network Curve with Dev Age and Accident Year



GLM Poisson Regression Curve



How Many Hidden Nodes for Neural Network?

- Too few nodes: Don’t fit the curve very well
- Too many nodes: Over parameterization
  - May fit noise as well as pattern

How Do We Determine the Number of Hidden Nodes?

- Use methods that assess goodness of fit
- Hold out part of the sample
- Resampling
  - Bootstrapping
  - Jacknifing
- Algebraic formula
  - Uses gradient and Hessian matrices

Hold Out Part of Sample

- Fit model on 1/2 to 2/3 of data
- Test fit of model on remaining data
- Need a large sample

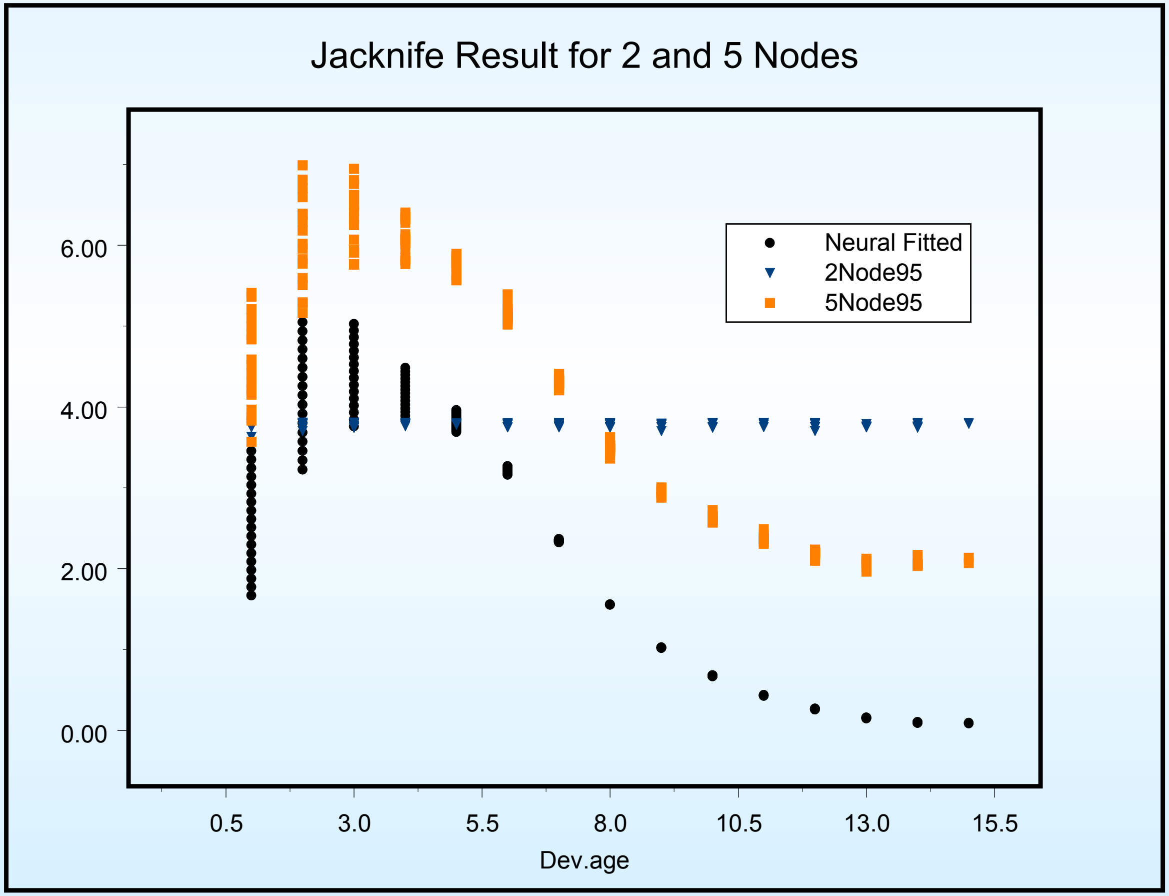
Cross-Validation

- Hold out 1/n (say 1/10) of data
- Fit model to remaining data
- Test on portion of sample held out
- Do this n (say 10) times and average the results
- Used for moderate sample sizes
- Jacknifing similar to cross-validation

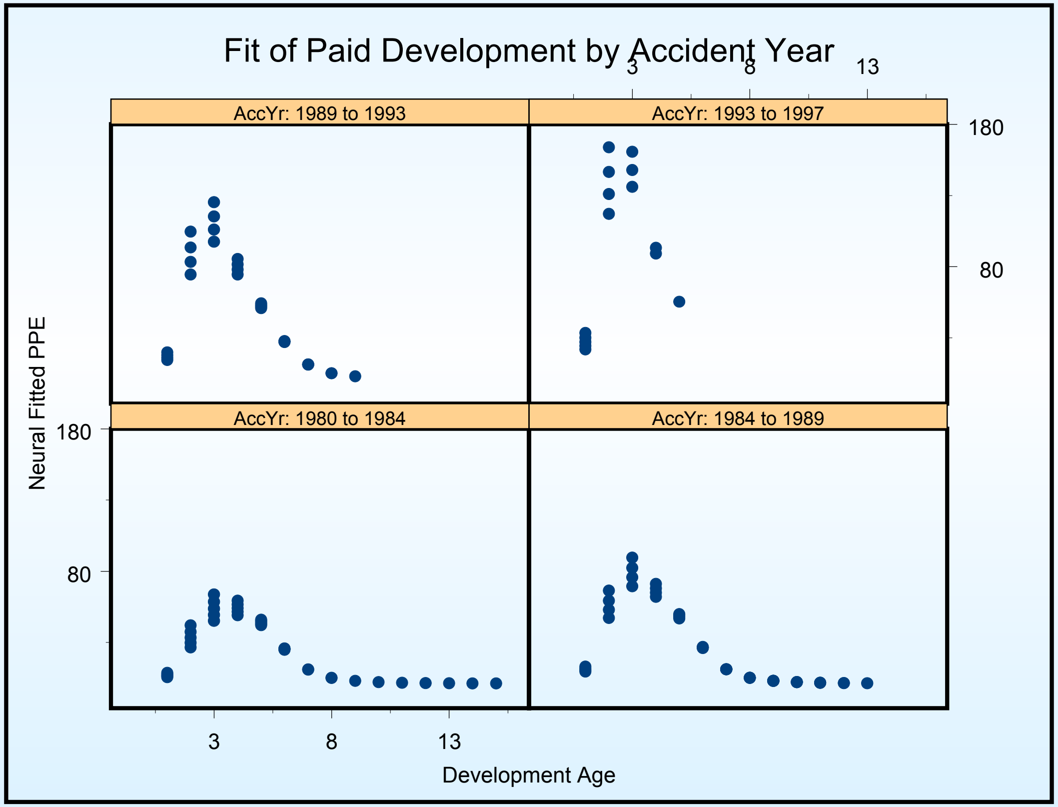
Bootstrapping

- Create many samples by drawing samples, with replacement, from the original data
- Fit the model to each of the samples
- Measure overall goodness of fit and create distribution of results
- Used for small and moderate sample sizes

Jackknife of 95% CI for 2 and 5 Nodes



Another Complexity of Data: Interactions



Technical Predictors of Stock Price

A Complex Multivariate Example

Stock Prediction: Which Indicator is Best?

- Moving Averages
- Measures of Volatility
- Seasonal Indicators
  - The January effect
- Oscillators

The Data

- S&P 500 Index since 1930
  - Open
  - High
  - Low
  - Close

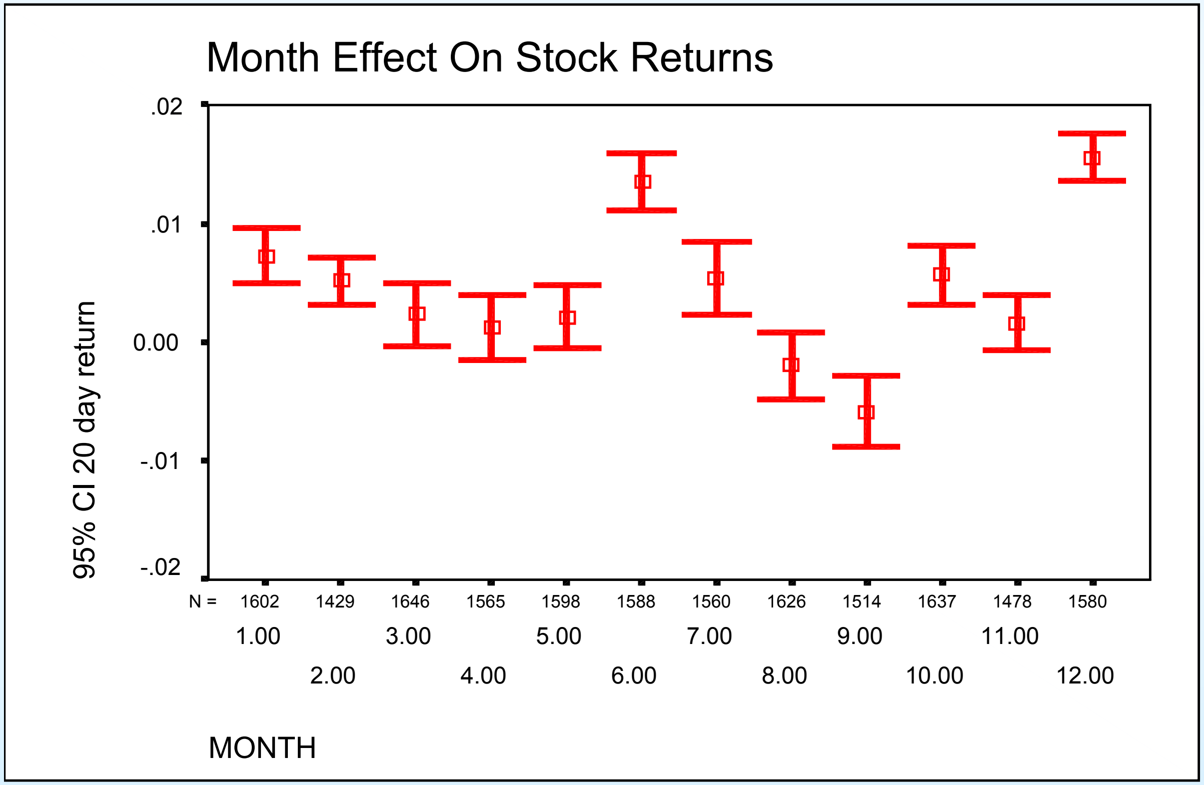
Moving Averages

- A very commonly used technical indicator
  - 1 week MA of returns
  - 2 week MA of returns
  - 1 month MA of returns
- These are trend following indicators
- A more complicated time series smoother based on running medians called T4253H

Volatility Measures

- Finance literature suggests volatility of market changes over time
- More turbulent market -> higher volatility
- Measures
  - Standard deviation of returns
  - Range of returns
  - Moving averages of above

Seasonal Effects



Oscillators

- May indicate that market is overbought or oversold
- May indicate that a trend is nearing completion
- Some oscillators
  - Moving average differences
  - Stochastic

Stochastic and Relative Strength Index

- Stochastic based on observation that as prices increase closing prices tend to be closer to upper end of range
  - %K = (C – L5)/(H5 – L5)
    - C is closing prince, L5 is 5 day low, H5 is 5 day high
  - %D = 3 day moving average of %K
- RS =(Average x day's up closes)/(Avg of x day's down Closes)
  - RSI = 100 – 100/(1+RS)

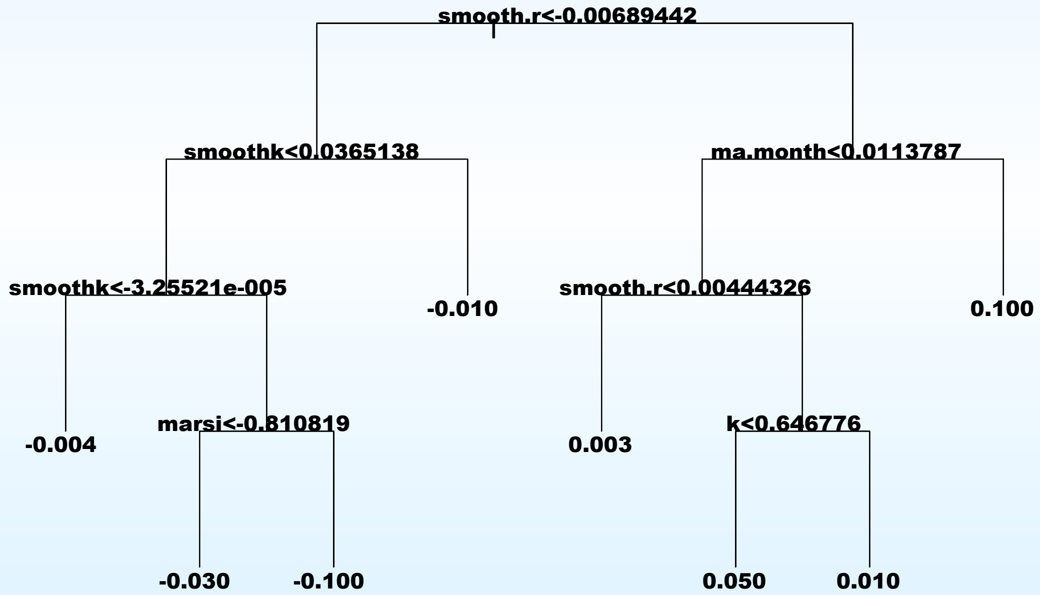
Measuring Variable Importance

- Look at weights to hidden layer
- Compute sensitivities:
  - a measure of how much the predicted value's error increases when the variables are excluded from the model one at a time

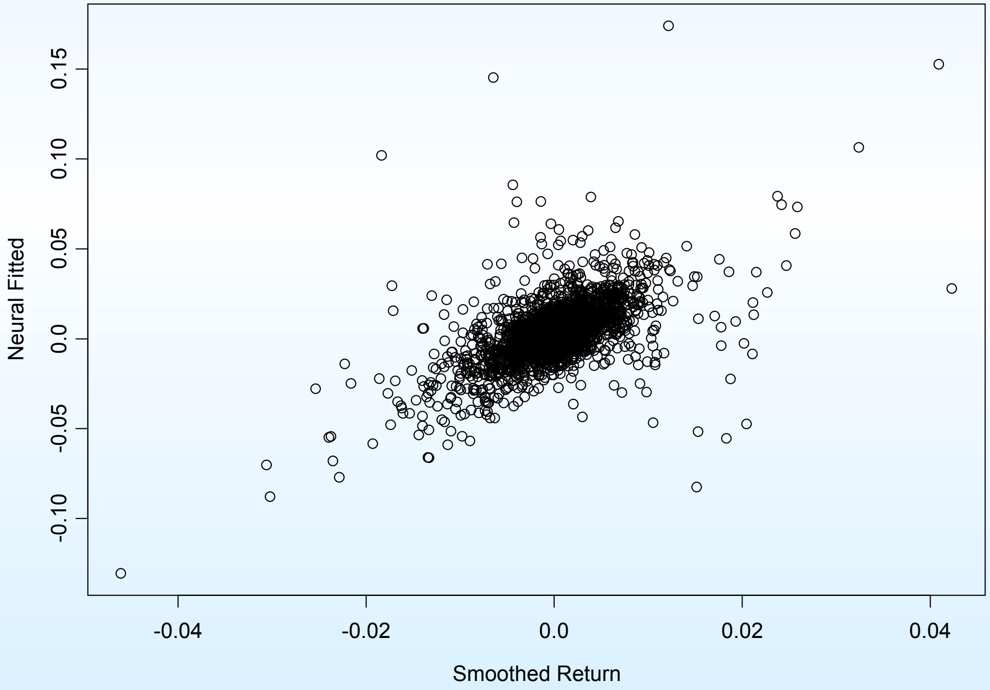
Neural Network Result

- Variable Importance
  - Smoothed return
  - %K (from stochastic)
  - Smoothed %K
  - 2 Week %D
  - 1 Week range of returns
  - Smoothed standard deviation
  - Month
- R^2 was .13 or 13% of variance explained

Understanding Relationships Between Predictor and Target Variables: A Tree Example



What are the Relationships between the Variables?



Visualization Method for Understanding Neural Network Functions

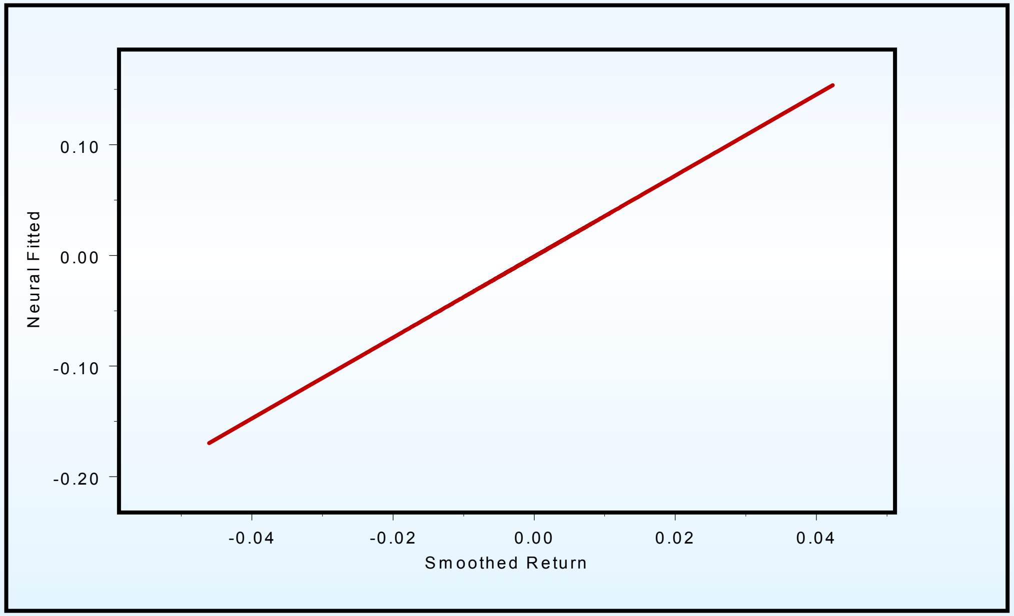
- Method was published by Plate *et al.* in Neural Computation, 2000
  - Based on Generalized Additive Models
  - Detailed description by Francis in "Neural Networks Demystified", 2001
- Hastie, Tibshirini and Friedman present a similar method

Visualization

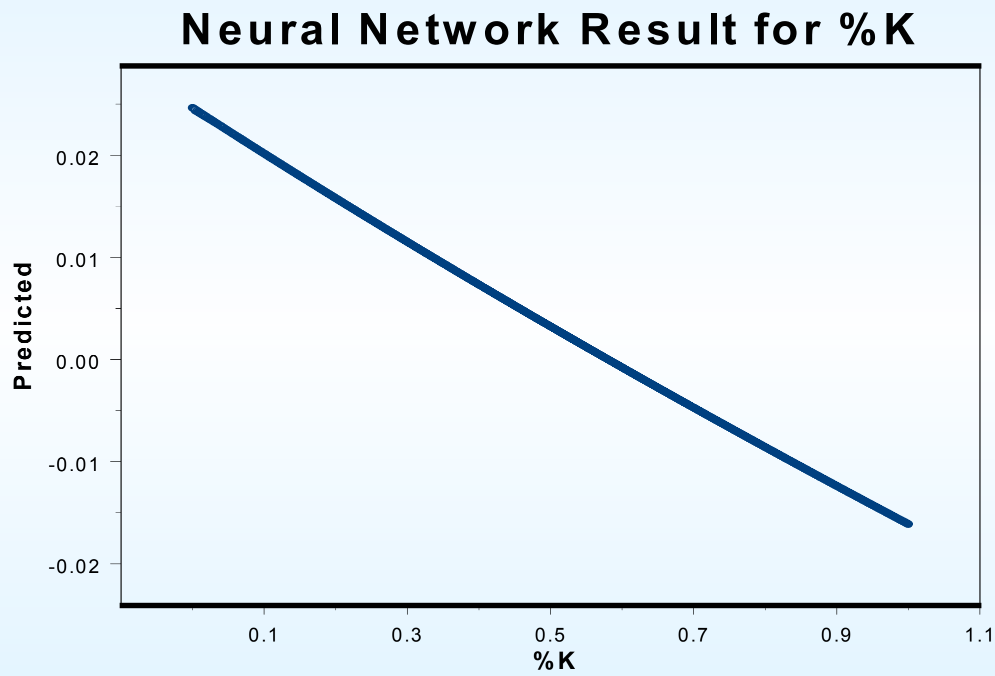
- Method is essentially a way to approximate the derivatives of the neural network model

$$\Delta_i f(x_1 \mid x_2 \ldots x_n) \approx f(x_{1,i} \mid x_2 \ldots x_n) -$$
$$f(x_{1,i+1} \mid x_2 \ldots x_n)$$

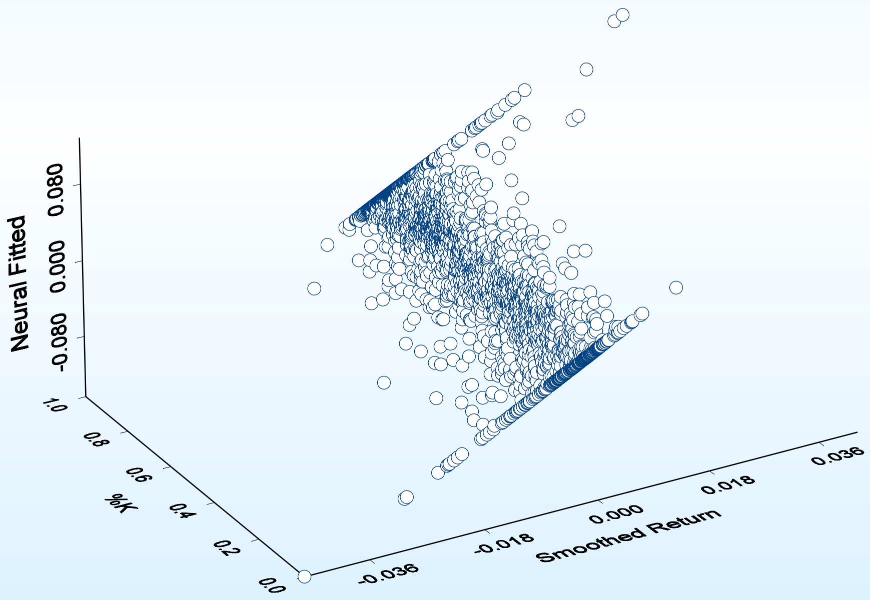
Neural Network Result for Smoothed Return



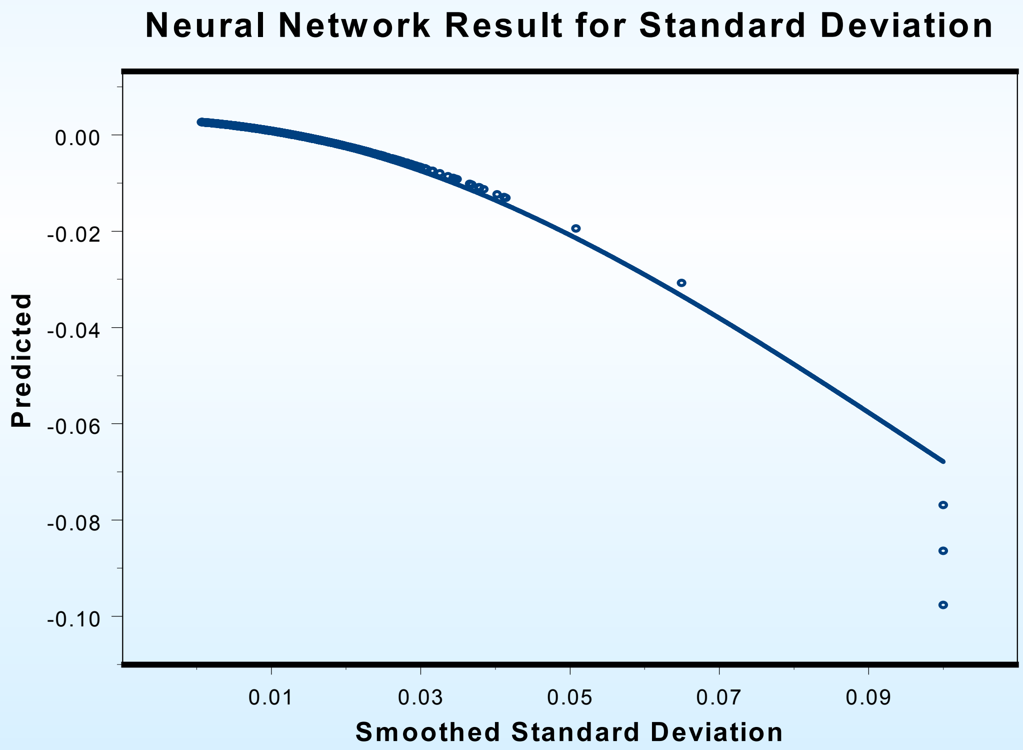
Neural Network Result for Oscillator



Neural Network Result for Smoothed Return and Oscillator



Neural Network Result for Standard Deviation



Conclusions

- Neural Networks are a lot like conventional statistics
- They address some problems of conventional statistics: nonlinear relationships, correlated variables and interactions
- Despite black block aspect, we now can interpret them
- Find further information, including many references, at www.casact.org/aboutcas/mdiprize.htm
  - *Neural Networks for Statistical Modeling*, M. Smith
- Reference for Kohonen Networks:
  - *Visual Explorations in Finance*, Guido Debok and Teuvo Kohone (Eds), Springer, 1998