A Flexible Generalized Link Function for Credit Scoring

André de Waal, Tiny du Toit and Tanja de la Rey

NORTH-WEST UNIVERSITY
YUNIBESITI YA BOKONE-BOPHIRIMA
NOORDWES-UNIVERSITEIT
POTCHEFSTROOMKAMPUS

Agenda

- Introduction
- Logistic Regression
- Generalized Additive Neural Network
- A Flexible Link Function
- Example
- Two Scorecards
- Conclusions

Introduction

- **Generalized additive models**

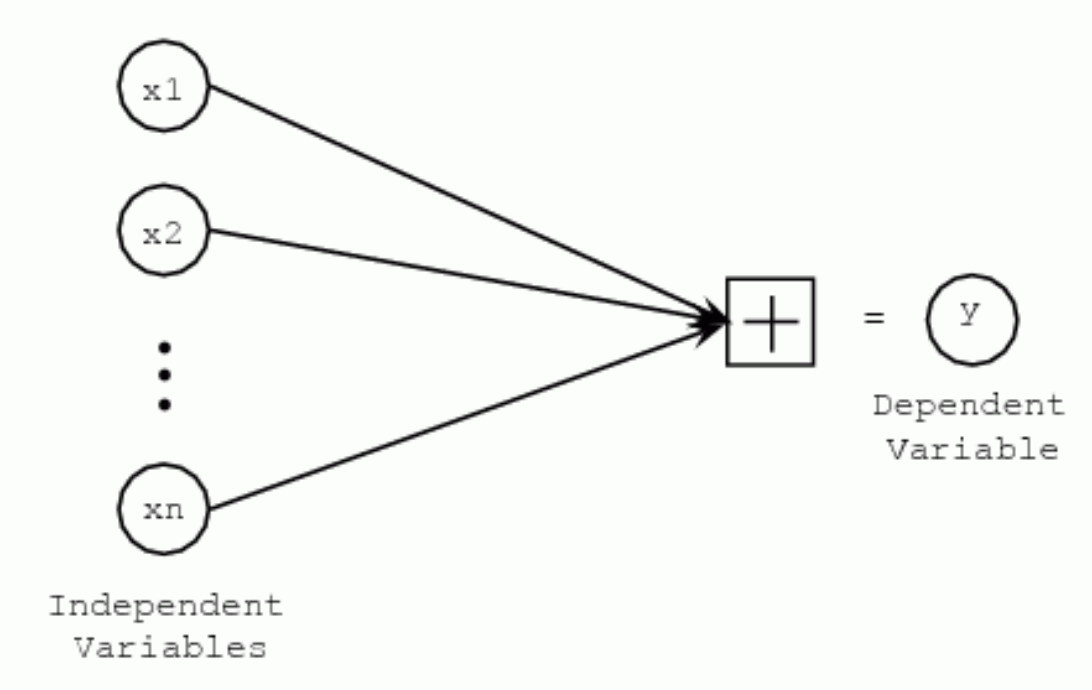
$$g_0^{-1}(E(y_i)) = \beta_0 + f_1(x_{1i}) + f_2(x_{2i}) + \cdots + f_k(x_{ki})$$

- **Approximated univariate functions with neural networks (mlp's)**
- **AutoGANN modelling node in SAS Enterprise Miner**
- **ACE (Alternating Conditional Expectation)**
- **Interval target**
  - link function is also a univariate function
- **Binary case**
  - ??

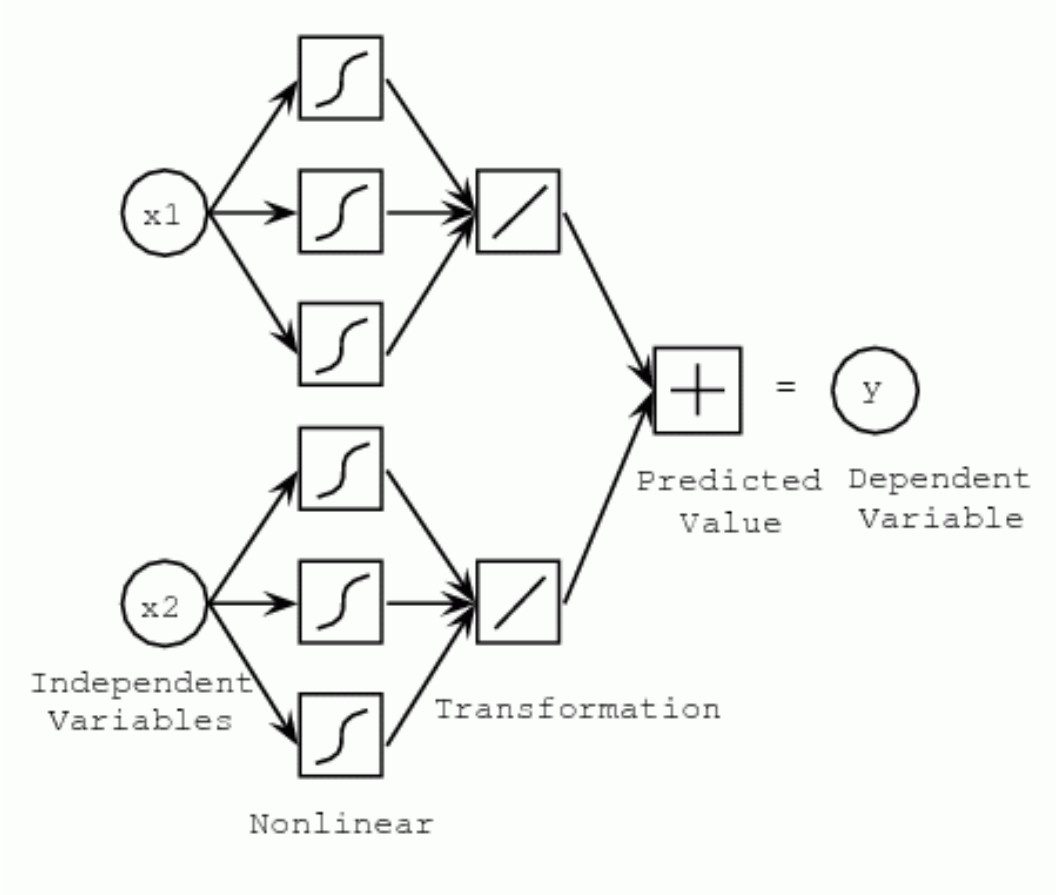
Logistic Regression

- **Logistic regression**
  - most popular tool used for constructing scorecards
  - de facto standard
    - building a good scorecard
    - judge any other technique
  - easy to interpret
  - widely understood
  - accepted by regulatory authorities
- **Alternative**
  - Probit
  - Complementary log-log

Logistic Regression (continued)



Generalized Additive Neural Network

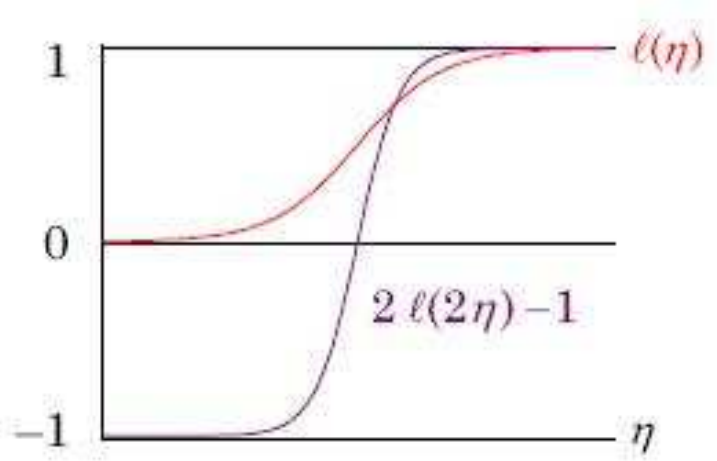


A Flexible Link Function

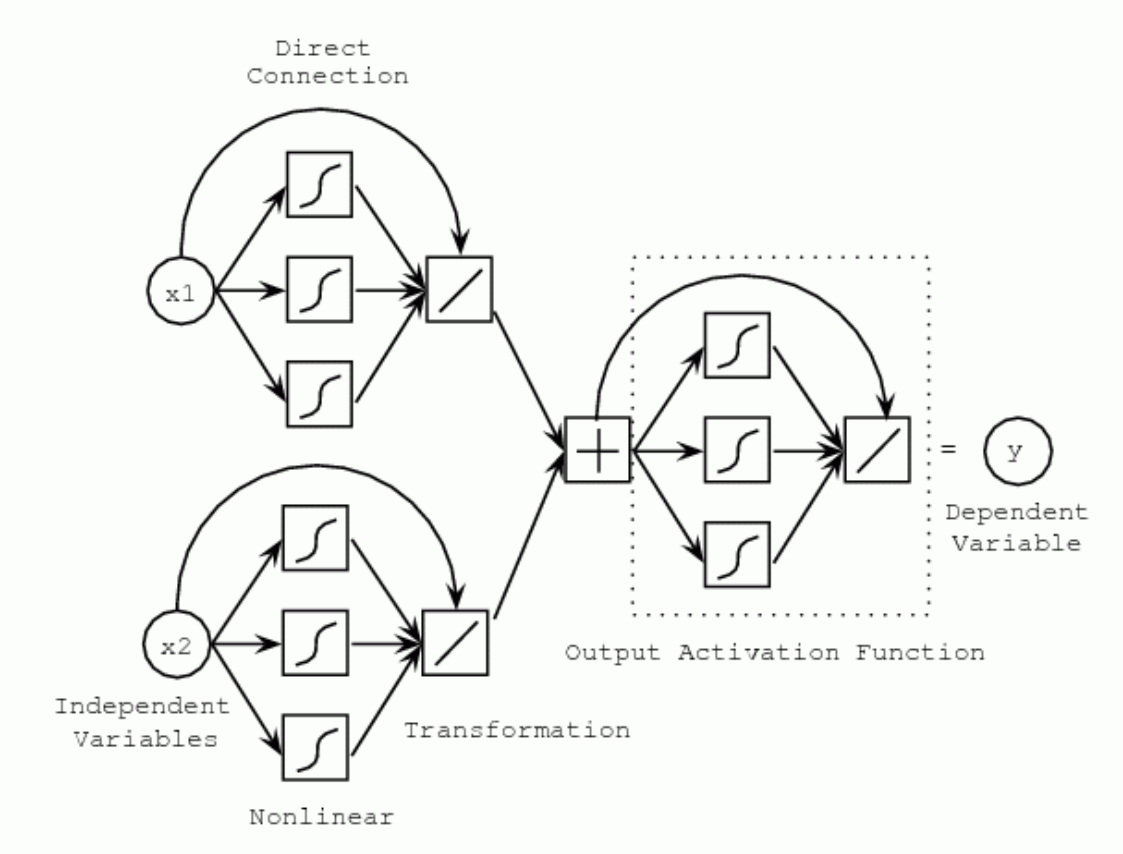
- **The logit link function not optimal in all cases**
  - non-symmetrical link function
  - one response is much more frequent than the other
- **Want a data driven approach**
- **Compute the optimal link function**
  - using a neural network
- **Improved Gini coefficient**
- **Lower misclassification rate**
- **“Holy Grail of Credit Scoring”**

A Flexible Link Function (continued)

- logistic(η) = $1/(1+e^{-\eta})$
- tanh(η) = $(e^{\eta}-e^{-\eta})/(e^{\eta}+e^{-\eta}) = 2 * \text{logistic}(2\eta) - 1$



A Flexible Link Function (continued)



A Flexible Link Function (continued)

- **Choice of activation function for hidden nodes**
  - want probabilities
  - logistic activation function
  - outputs of hidden nodes restricted to (0,1)
- **Constraints**
  - $c_1 \geq 0, c_2 \geq 0, \ldots, c_p \geq 0$
  - $c_1 + c_2 + \ldots + c_p = 1$
- **Linear combination**
  - $g_o(z) = c_1/(1+e^{-\alpha z}) + c_2/(1+e^{-\beta z}) + \ldots + c_p/(1+e^{-\rho z})$
  - $z = \beta_0 + f_1(x_1) + f_2(x_2) + \ldots + f_k(x_k)$

A Flexible Link Function (continued)

- **Restrictions**
  - combination of error function and activation function
  - inability to specify constraints
- **Implemented output activation function**
  - $g_o(\eta) = 1/(1+e^{-\eta})$
  - $\eta = (\beta_1 + c_1/(1+e^{-z}) + c_2/(1+e^{-\beta z}) + \ldots + c_p/(1+e^{-\rho z}))$
  - $z = \beta_0 + f_1(x_1) + f_2(x_2) + \ldots + f_k(x_k)$
- **η represents the linear combination**
- **α set equal to 1**
- **β_1 required for correct scaling**

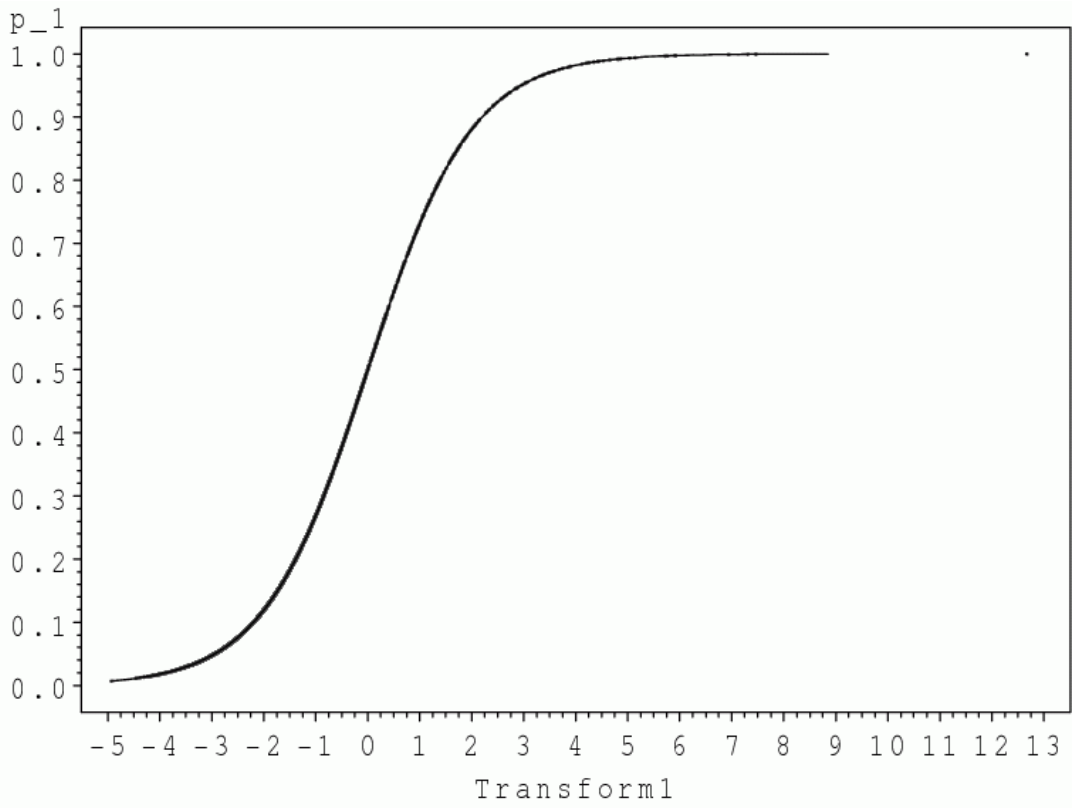
Home Equity

- 5960 recent home equity loans
- 1189 defaulted or seriously delinquent (±20%)
- Target
  - BAD
- 12 input variables
  - REASON, JOB, LOAN, MORTDUE, VALUE, DEBTINC, YOJ, DEROG, CLNO, DELINQ, CLAGE, NINQ

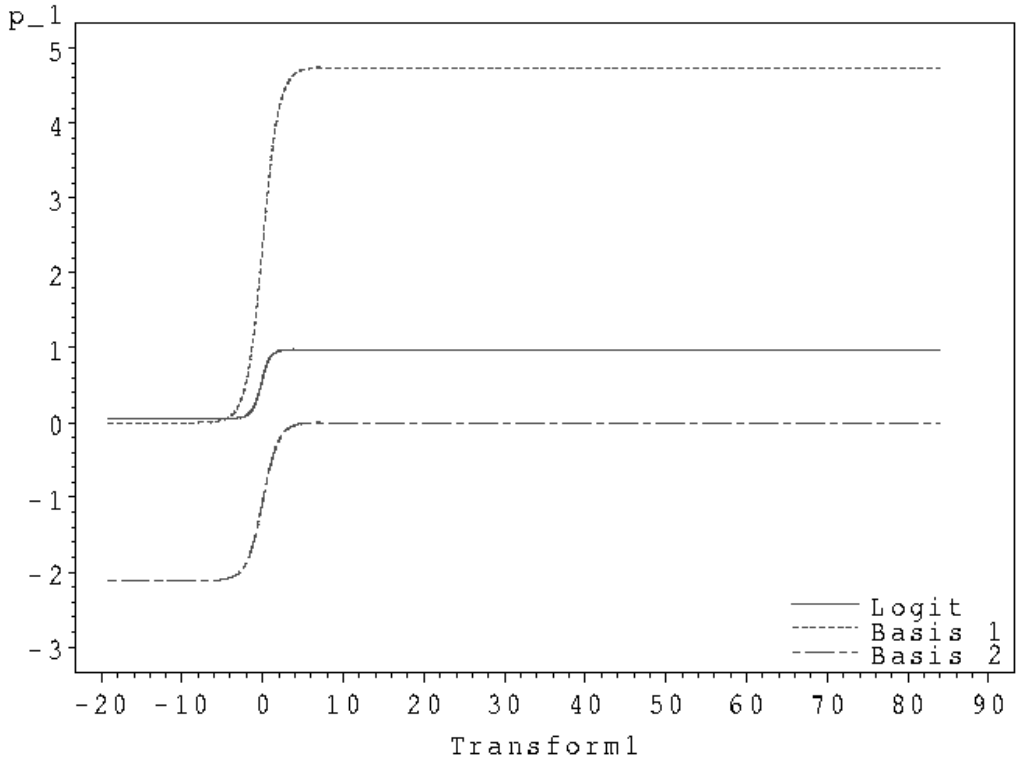
Home Equity (continued)

Variable	GANN architecture	Target relationship
LOAN	MLP with a skip layer and a hidden node	Nonlinear
JOB	No MLP (variable removed from model)	None
REASON	No MLP (variable removed from model)	None
CLAGE	MLP with a skip layer and a hidden node	Nonlinear
CLNO	MLP with a skip layer	Linear
DEBTINC	MLP with a skip layer and a hidden node	Nonlinear
DELINQ	MLP with a skip layer	Linear
DEROG	MLP with a skip layer and a hidden node	Nonlinear
NINQ	MLP with a skip layer and a hidden node	Nonlinear
MORTDUE	MLP with a skip layer	Linear
VALUE	MLP with a skip layer and a hidden node	Nonlinear
YOJ	No MLP (variable removed from model)	None

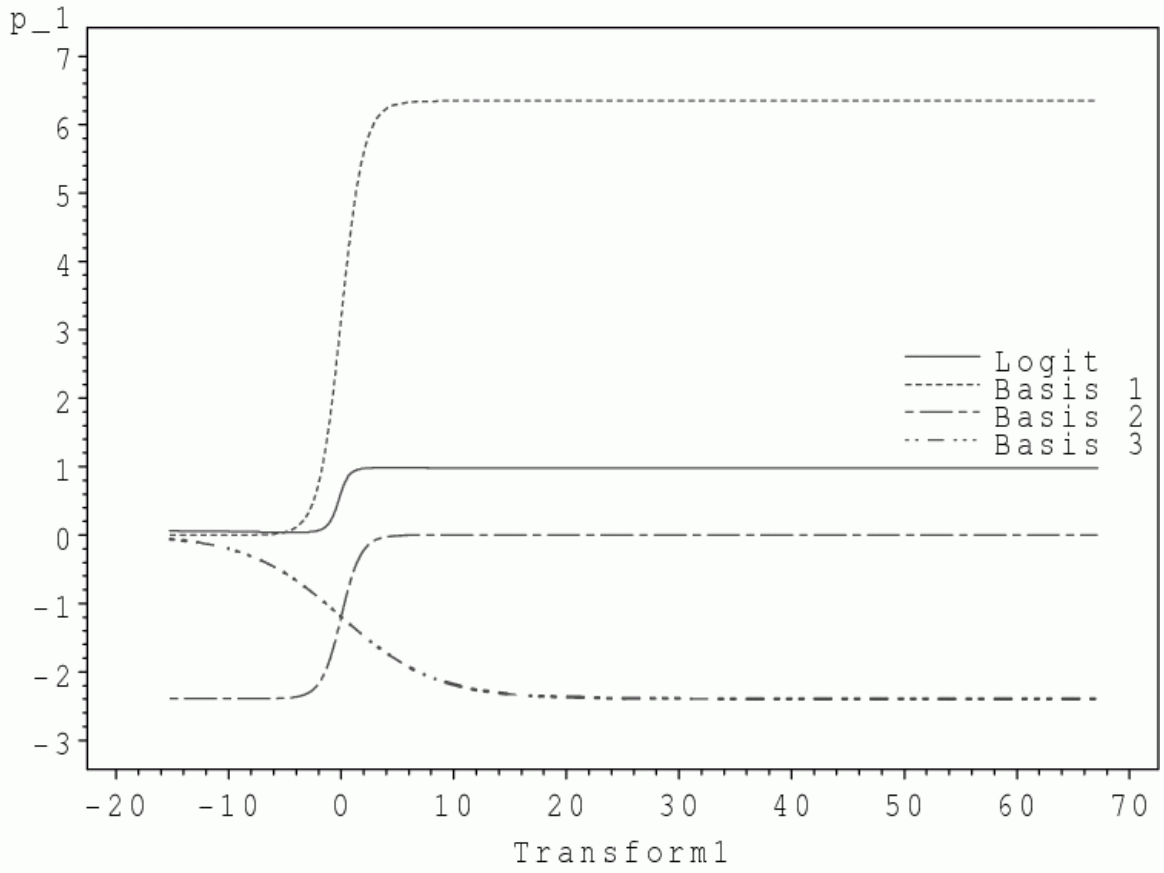
Home Equity (continued)



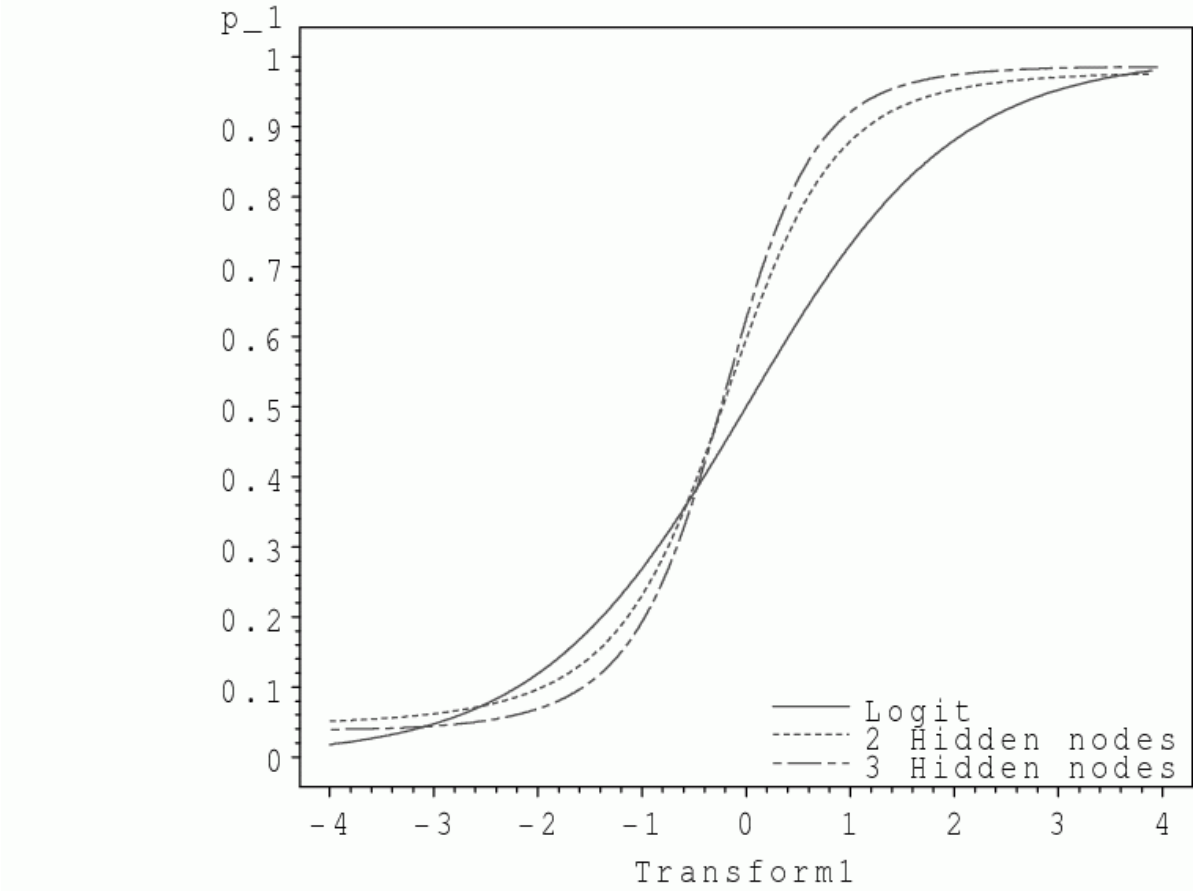
Home Equity (continued)



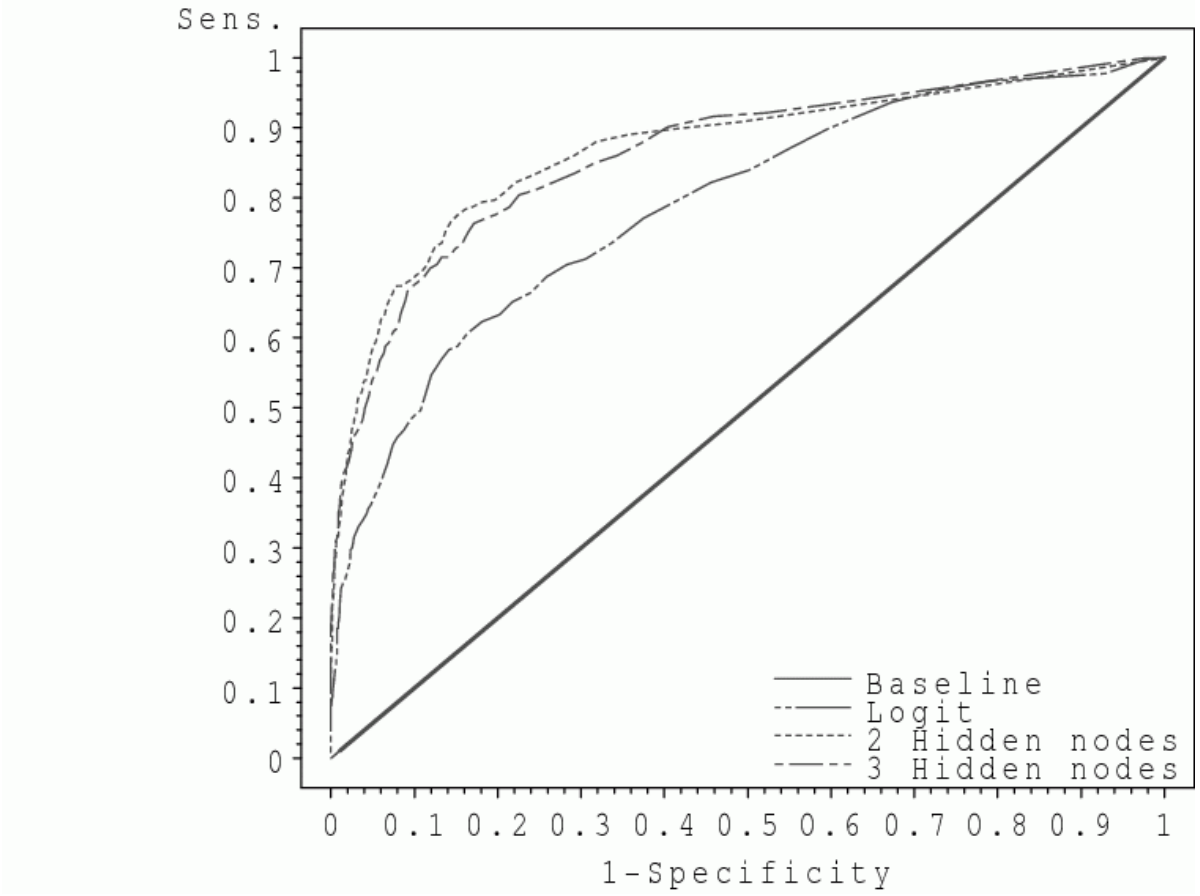
Home Equity (continued)



Home Equity (continued)



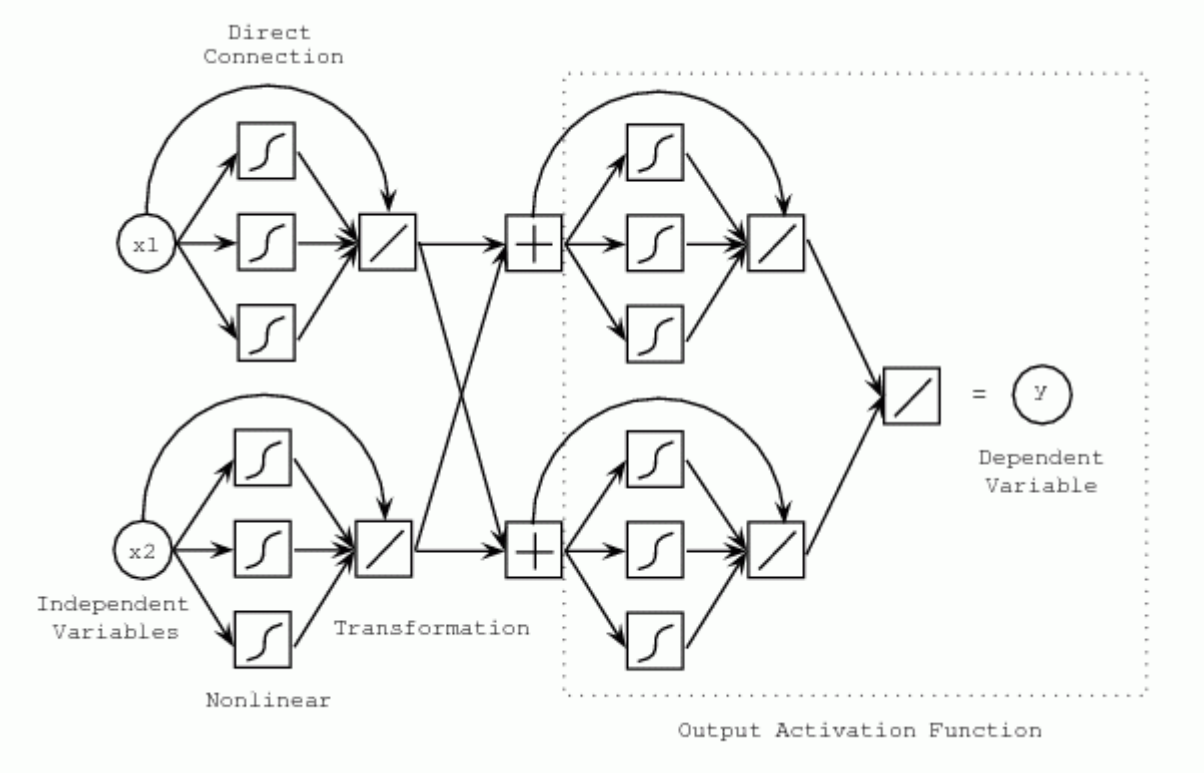
Home Equity (continued)



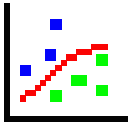
Home Equity (continued)

Model	ASE	Misclassification rate	Gini coefficient
Baseline GANN	0.1210	15.95%	0.5771
Baseline with 2 hidden nodes	0.0926	12.50%	0.7371
Baseline with 3 hidden nodes	0.0966	13.26%	0.7251

Two Scorecards



Conclusions

- **Proposed a flexible generalized link function**
  - based on neural networks
- **Implemented in AutoGANN modelling node**
- **Search over different architectures using Gini**
- **Future work**
  - detailed evaluation on more complex data sets
  - consequences when building a scorecard
  - interpretation of results
- **AutoGANN demo** 

THE END