Article can be accessed online at http://www.publishingindia.com

Application of Generalised Additive Logistic Model for Targeted Marketing

K. V. N. K. Prasad*, G.V.S.R. Anjaneyulu**

Abstract

This study focuses on how to support marketing decision makers better in identifying better prospective customers by using generalised additive models (GAMs). Compared to logistic regression, GAM relaxes the linearity constraint which allows for complex non-linear fits to the data. In this paper, we examine how GAM-based logistic models perform compared to traditional logistic regression model and also provide some implications.

Keywords: Additive Logistics Model, Targeted Marketing

Introduction

Last few years have seen an enormous emphasis on developing non-linear modeling approaches to deal with the problem of rapidly increasing variance for higher dimensional modeling problems because non-linear modeling approach provides information about the relationship between the dependent and independent variable that are not quite revealed by the standard modeling approaches. Secondly, a wide variety of model specifications can be tested so that model misspecification can be largely avoided (Hastie & Tibshirani, 1990).

Stone (1985) proposed the concept of additive models; these models estimate an additive approximation to the multivariate regression; in additive models each term is estimated by smoother – which helps in dealing with curse of dimensionality. Secondly, each estimate for each individual predictor in the model will explain how the dependent variable change in accordance with the corresponding individual predictors.

The extension of additive models across a wide range of distribution families (exponential family) by Hastie and Tibshirani (1990) called generalised additive models (GAMs). In GAM the mean of the dependent variable depends on an additive independent variable through a non-linear link function with a flexibility of response probability distribution being any member from the exponential family of distributions.

Generalised additive model proposed by Hastie and Tibshirani (1990) is a class of generalised linear model with a linear predictor involving a sum of smooth functions of covariates. The generalised additive model replaces $\sum_{j=1}^{p}\beta_j x_j$ (linear predictor) in generalised linear model architecture with $\sum_{j=1}^{p} f_j x_j$ (additive predictors) where $f_{j's}$ are unspecified non-parametric function. The generic form of generalised additive model can be represented as follows

$$\mu(x) = E\,[Y|X_1\,X_2\,....\,X_p\,] = (X_1\,)+(X_2\,).....(X_p\,) + \xi\,;$$

$$E\,[Y|X_1\,X_2\,....\,X_p\,] = + \xi\,;$$

where $y^{\sim}$ some exponential family of distributions

Generalised Additive Logistic Model

The generic form of the logistic model in generalised linear model is as follows:

$$\text{Let } y = \begin{cases} 1 \text{ with probability } p(x) \\ 0 \text{ with probability } 1-p(x) \end{cases}$$

where $x = (X_1\,X_2\,....\,X_p\,)$ be a vector of covariates

then

$$\text{Logit } (p(x)) = \log\frac{p(X)}{1-p(X)}\,\beta_0\sum_{j=1}^{p}\beta_j x_j$$

and $$p(x) = \frac{Exp\,(\beta_0 + \sum_{j=1}^{p}\beta_j\left(x_{jj}\right)}{1+Exp\,(\beta_0 + \sum_{j=1}^{p}\beta_j\left(x_{jj}\right)}$$

* Department of Statistics, Acharya Nagarjuna University, Guntur, Andhra Pradesh, India. Email: kota.prasad.krishna@gmail.com
** Department of Statistics, Acharya Nagarjuna University, Guntur, Andhra Pradesh, India.

The generalised additive logistic model assumes the functional form

$$\log \frac{p(Y \mid xi1,\ldots\ldots xip)}{1-p(Y \mid xi1,\ldots\ldots xip)} \beta_0 f_1(x_{i1}) + \ldots. + f_p(x_{ip})$$

$$= + f_1(x_{i1}) + \ldots. + f_p(x_{ip})$$

and $p(x) = \frac{Exp\,(f_0 + \sum_{j=1}^{p} f_j(x_{ij})}{1 + Exp\,(f_0 + \sum_{j=1}^{p} f_j(x_{ij})}$

$$\eta(x) = \log \frac{p(x)}{1-p(x)}$$

where η is a function of p variables.

Assume that $\eta(x) + \xi$ $Y =$ is an initial estimate of $\eta(x)$, the adjusted dependent variable is

$$Z_i = \eta_i + (y_i + \mu_i)\left(\frac{\partial \eta_i}{\partial \mu_i}\right)$$

We fit additive model to z_i 's, treating it as the response variable Y in E(Y|X) =+ $\alpha + \sum_{J=1}^{p} f_j x_j \xi$. The function $f_{\hat{j}} x_j$ is estimated by smoothing (where $f_{\hat{j}}$ are smooth functions of).

Fitting Generalised Additive Models

The estimation of generalised additive models (GAMs) is a bi-folded iterative process. In each step, local scoring and back fitting algorithms are used until the convergence criteria for Backfitting algorithm is satisfied. If change in deviance of the estimates is below certain threshold, the algorithm stops, else based on estimates obtained from the Backfitting algorithm new weights are computed and iterative process is continued until the deviance of the estimates is below certain threshold.

Backfitting and Local Scoring Algorithm

The Backfitting algorithm is a generic algorithm that can fit an additive model using any regression mechanism. The principal ideal behind Backfitting algorithm is to find the j^{th} smoother of the $(k+1)^{th}$ iteration by smoothing the partial residuals defined by

$$R_j = Y - f_0 - \sum_{k \neq j} f_k(x_k)$$

The partial residuals remove the effects of all the other variables from y, thus they can be used to model the individual effect against x_j. The iterative mechanism in Backfitting algorithm constructs a smooth curve f(x) that summarises the dependency of y on x until the change in each effect is sufficiently small. This is achieved by starting with initial function and looping iteration cycle through partial residuals, fitting the individual smoothing components to its partial residuals. Hastie and Tibshirani (1990) proved that with many smoothers, the residual sum of squares (RSS) will never increase at any step; this implies that there are no convergence issues with the algorithm.

The general local scoring algorithm is an extension to the Backfitting Algorithm and is also an iterative algorithm. Since the linear predictors are replaced by additive predictors in generalised additive models, thus the Fisher scoring procedure is replaced by the local scoring algorithm as the predictions for the adjusted dependent variable are localised by nonparametric smoothers.

Smoothing and Methods of Choosing Smoothing Parameter

Each smoother specified in generalised additive model has a single smoothing parameter. Smoothing is a mechanism of summarising the trend (non-parametric) in dependent variable Y as a function of one or more independent variables X_1, X_2, …. X_p. Since smoothing is just a summarisation of trend, it will not assume any rigid form of dependency/ relationship between Y and X_1, X_2, …. X_p. There are two methods for choosing smoothing parameters:

(a) Cross-Validation: In this method a data point (x_i, y_i) is left out at a time as testing set and the smoother is estimated at xi based on remaining n-1 points to minimise the sum of those squared residuals.

(b) Generalised Cross-Validation: It is a weighted cross-validation method which was proposed by Craven and Wahba (1979). The generalised cross-validation method provides an alternative and convenient approximation to the leave-one-out cross-validation with lesser computing cost. Minimising the generalised cross-validation criterion often yields a similar smoothing parameter to that obtained by the leave-one-out cross-validation.

Introduction to Logistic Regression

Consider the following simple linear regression setting with 'r' predictor and binary response variable

$$y_i = \beta_0 + \beta_1 x_i + \ldots\ldots + \beta_r x_r + \int_i, i = 1, 2, \ldots . n$$

where y_i is the binary response variable, $\varepsilon_i \sim N(0, \sigma_\varepsilon^2)$), and are independent.

Let P_i denote the probability that $y_i = 1$ and $x_i = x$

$$P_i = P(Y_i = 1 | X_i = X) = \frac{1}{\left(1 + e^{-z}\right)}$$

where $Z = \beta_0 + \beta_1 x_i + \ldots\ldots + \beta_r x_r$

Or

$$\text{Logit}(p) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \ldots + \beta_r x_r$$

The above equitation is called logistic regression, a statistical method in which we model the logit (p) in terms of explanatory variables that are available to modeler. It is non-linear in the parameters,……. The response probabilities are modeled by logistic distribution and estimating the parameters of the model constitutes fitting a logistic regression.

Maximum Likelihood Method of Estimation

In maximum likelihood estimation, we search over all possible sets of parameter values for a specified model to find the set of values for which the observed sample was most likely. That is, we find the set of parameter values that, given a model, were most likely to have given us the data that we have in hand.

Model Performance Measures

The traditional model fit evaluation based on AIC and BIC can't be used for evaluation as - the AIC for generalised additive model is computed as deviance + 2p, where p are the model degrees of freedom; whereas the AIC for generalised linear models is computed as 2LL + 2p, where LL is the log likelihood of the fitted model and p are the model degrees of freedom.

Thus to evaluate the performance of the model we rely on the significance of the independent variables in both models, along with significance of the smoothing terms/ effects in the generalised additive models along with the following measures that are usually used to evaluate the performance of the model.

Kolmogorov-Smirnov (KS)

This measures the maximum vertical separation (deviation) between the cumulative distributions of goods and bads and is defined as follows

$$KS = MAX\ |F_G^{(s)} - F_B^{(s)}|$$

The higher the KS value, the better is the model's ability for separation.

Accuracy Ratio

Typically, the assessment of a marketing response model is evaluated by discriminatory power is commonly measured by the CAP or Lorenz curve (a variant of the ROC curve) and the accuracy ratio AR derived from this curve. The difference in comparison with the ROC curve consists in plotting the cumulative distribution function of all scores against that of the default score (instead of plotting the cumulative distribution functions of the non-default and default scores against each other). The accuracy ratio calculated from the CAP curve is however linearly related to the area under curve AUC of the ROC curve:

$$AR = 2AUC - 1$$

Data

The data is related with direct marketing campaigns of a Portuguese banking institution. The marketing campaigns were based on phone calls. The classification goal is to predict if the client will subscribe a term deposit. The dataset consist of 45211 instances.

Basic Statistics and Correlation

The basic statistics for the numerical variables used in the model are reported below.

Table 1: Statistics for Numerical Variables used in the Model

Variable	N	N Miss	Minimum	Maximum	Variance	Std Dev
age	45211	0	18	95	112.76	10.62
balance	45211	0	-8019	102127	9270598.95	3044.77
day	45211	0	1	31	69.26	8.32
duration	45211	0	0	4918	66320.57	257.53
campaign	45211	0	1	63	9.60	3.10
pdays	45211	0	-1	871	10025.77	100.13
previous	45211	0	0	275	5.31	2.30
target	45211	0	0	1	0.10	0.32
housing_n	45211	0	0	1	0.25	0.50
loan_n	45211	0	0	1	0.13	0.37
default_n	45211	0	0	1	0.02	0.13
job_n	45211	0	1	2	0.25	0.50
marital_n	45211	0	1	2	0.24	0.49
education_n	45211	0	1	2	0.22	0.47
contact_n	45211	0	1	3	0.81	0.90
poutcome_n	45211	0	1	2	0.03	0.18
_n_ variables are characterstic variables that are binned						

The following correlation analysis among the numeric variables indicates that there is no problem of co-linearity in the modeling dataset.

Table 2: Correlation Analysis among the Numeric Variables

	age	balance	day	duration	campaign	pdays	previous	target	housing_n	loan_n	default_n	job_n	marital_n	education_n	contact_n	poutcome_n
age	1	0.098	-0.009	-0.005	0.005	-0.024	0.001	0.025	-0.186	-0.016	-0.018	0.102	0.286	-0.050	0.026	0.036
balance		1.000	0.005	0.022	-0.015	0.003	0.017	0.053	-0.069	-0.084	-0.067	-0.038	0.026	0.086	-0.027	0.035
day			1.000	-0.030	0.162	-0.093	-0.052	-0.028	-0.028	0.011	0.009	-0.031	0.007	0.021	-0.028	-0.030
duration				1.000	-0.085	-0.002	0.001	0.395	0.005	-0.012	-0.010	0.015	-0.023	0.001	-0.021	0.042
campaign					1.000	-0.089	-0.033	-0.073	-0.024	0.010	0.017	-0.013	0.031	0.015	0.020	-0.057
pdays						1.000	0.455	0.104	0.124	-0.023	-0.030	0.009	-0.028	-0.010	-0.245	0.229
previous							1.000	0.093	0.037	-0.011	-0.018	-0.019	-0.013	0.019	-0.148	0.201
rand								-0.005	-0.001	0.000	0.005	-0.013	-0.007	0.009	-0.003	0.000
target								1.000	-0.139	-0.068	-0.022	-0.031	-0.060	0.069	-0.148	0.307
housing_n									1.000	0.041	-0.006	0.072	0.018	-0.116	0.188	-0.091
loan_n										1.000	0.077	0.021	0.037	-0.065	-0.011	-0.054
default_n											1.000	0.008	-0.014	-0.015	0.015	-0.023
job_n												1.000	0.111	-0.362	0.125	-0.029
marital_n													1.000	-0.089	0.033	-0.018
education_n														1.000	-0.118	0.051
contact_n															1.000	-0.114
poutcome_n																1.000
_n_ indicates characterstic variables that are binned																

Training Vs Validation

The modeling data was split into 50/50 by using simple random sampling; the first 50% of the data was used to train the data and the remaining 50% of the data is used to validate the model.

Table 3: Training vs Validation

Training Vs Validation:						
Target	Frequency	Cumulative Frequency	Train Response rate	Frequency	Cumulative Frequency	Test Response rate
0	19,920	19,920	11.88%	20,002	20,002	11.52%
1	2,685	22,605		2,604	22,606	

Logistic Vs Generalised Additive Logistic Model comparisons

Table 4: Logistic vs Generalised Additive Logistic Model Comparisons

Logistic Regression						
Analysis of Maximum Likelihood Estimates						
Parameter	Categories	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept		1	-2.9686	0.1258	556.4165	<.0001
balance		1	0.000023	6.86E-06	11.5676	0.0007
duration		1	0.00413	0.000088	2176.839	<.0001
age		1	0.00244	0.00234	1.0899	0.2965
housing	01: Yes	1	0.3764	0.0266	200.9022	<.0001
	02: No		Ref	Ref	Ref	Ref
Loan	01: Yes	1	0.319	0.041	60.5037	<.0001
	02: No		Ref	Ref	Ref	Ref
poutcome_n	01 : 'unknown','other','failure'	1	-1.2558	0.0447	787.6218	<.0001
	01: 'success'		Ref	Ref	Ref	Ref
marital_n	01: 'divorced','single'	1	0.1951	0.0263	55.0716	<.0001
	02: Married & others		Ref	Ref	Ref	Ref
education_n	01: 'primary', 'secondary'	1	-0.0954	0.0257	13.7597	0.0002
	02: 'tertiary', 'others'		Ref	Ref	Ref	Ref
contact_n	01:'cellula'	1	0.4875	0.0452	116.2995	<.0001
	02:'telepho'	1	0.281	0.0716	15.3937	<.0001
	02:'Others'		Ref	Ref	Ref	Ref
previous		1	0.0795	0.0104	57.8922	<.0001

Generalized Additive Logistic Regression					
Regression Model Analysis					
Parameter	Categories	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept		-2.8128	0.19068	-14.75	<.0001
Linear(balance)		-0.00051936	1.95E-05	-26.64	<.0001
duration		0.00447	9.67E-05	46.22	<.0001
Linear(age)		0.00415	0.0024	1.73	0.0834
housing	01: Yes	0.73103	0.0571	12.8	<.0001
	02: No	Ref	Ref	Ref	Ref
Loan	01: Yes	0.57761	0.08671	6.66	<.0001
	02: No	Ref	Ref	Ref	Ref
poutcome_n	01 : 'unknown','other','failure'	-2.47527	0.09721	-25.46	<.0001
	01: 'success'	Ref	Ref	Ref	Ref
marital_n	01: 'divorced','single'	0.27462	0.05666	4.85	<.0001
	02: Married & others	Ref	Ref	Ref	Ref
education_n	01: 'primary', 'secondary'	-0.20967	0.05507	-3.81	0.0001
	02: 'tertiary', 'others'	Ref	Ref	Ref	Ref
contact_n	01:'cellula'	1.28846	0.08539	15.09	<.0001
	02:'telepho'	0.83111	0.138	6.02	<.0001
	02:'Others'	Ref	Ref	Ref	Ref
previous		0.09234	0.01117	8.27	<.0001

From Table 4, we can observe that the estimate for age is highly insignificant, followed by balance in logistic regression, but in GAM the linear component of balance and age are statistically significant at 5% level of significance.

Table 5 provides a summary of significance of the smoothing parameters used in the model.

Table 5: Summary of Significance of the Smoothing Parameters

Smoothing Model Analysis			
Approximate Analysis of Deviance			
Source	DF	Chi-Square	Pr > ChiSq
Spline(balance)	352.1504	18801.4343	<.0001
Spline(age)	12.43418	259.5807	<.0001

For each smoothing effect in the model, Table 5 gives a χ^2 test comparing the deviance between the full model and the model without the nonparametric component of this variable. The analysis of deviance results indicates that the non-parametric effect of age and balance are highly significant.

Performance Analysis

Table 6: Performance Results of Generalised Additive Models - I

Interval	Minmum Score	Maximum Score	Total Accounts	Number of goods	Number of bads	Good Rate	% goods	%bads	Cummulative goods : G(i)	Cummulative bads B(i)	KS
1	0.60130	0.73110	2,260	1,232	1,028	54.51%	45.88%	5.16%	45.88%	5.16%	40.72%
2	0.53990	0.60130	2,261	706	1,555	31.23%	26.29%	7.81%	72.18%	12.97%	59.21%
3	0.52250	0.53990	2,260	308	1,952	13.63%	11.47%	9.80%	83.65%	22.77%	60.88%
4	0.51490	0.52250	2,261	166	2,095	7.34%	6.18%	10.52%	89.83%	33.28%	56.55%
5	0.51040	0.51490	2,260	75	2,185	3.32%	2.79%	10.97%	92.63%	44.25%	48.37%
6	0.50730	0.51040	2,261	48	2,213	2.12%	1.79%	11.11%	94.41%	55.36%	39.05%
7	0.50490	0.50730	2,261	28	2,233	1.24%	1.04%	11.21%	95.46%	66.57%	28.88%
8	0.50310	0.50490	2,260	16	2,244	0.71%	0.60%	11.27%	96.05%	77.84%	18.22%
9	0.50160	0.50310	2,261	5	2,256	0.22%	0.19%	11.33%	96.24%	89.16%	7.08%
10	0.50000	0.50160	2,260	101	2,159	4.47%	3.76%	10.84%	100.00%	100.00%	0.00%
			22605	2685	19920						60.88%

Development GINI	71.80%
Development KS	60.88%

Table 7: Performance Results of Generalised Additive Models - II

Interval	Minmum Score	Maximum Score	Total Accounts	Number of goods	Number of bads	Good Rate	% goods	%bads	Cummulative goods : G(i)	Cummulative bads B(i)	KS
0				0	0		0%	0%	0.00%	0.00%	0.00%
1	0.6013	0.7311	2,260	1,177	1,083	47.90%	5.88%	41.59%	5.88%	41.59%	35.71%
2	0.5399	0.6013	2,261	1,592	669	29.59%	7.96%	25.69%	13.84%	67.28%	53.44%
3	0.5225	0.5399	2,261	1,936	325	14.37%	9.68%	12.48%	23.52%	79.76%	56.24%
4	0.5149	0.5225	2,260	2,093	167	7.39%	10.46%	6.41%	33.99%	86.18%	52.19%
5	0.5104	0.5149	2,261	2,159	102	4.51%	10.79%	3.92%	44.78%	90.09%	45.31%
6	0.5073	0.5104	2,261	2,196	65	2.87%	10.98%	2.50%	55.76%	92.59%	36.83%
7	0.5049	0.5073	2,260	2,218	42	1.86%	11.09%	1.61%	66.85%	94.20%	27.35%
8	0.5031	0.5049	2,261	2,240	21	0.93%	11.20%	0.81%	78.05%	95.01%	16.96%
9	0.5016	0.5031	2,261	2,248	13	0.57%	11.24%	0.50%	89.29%	95.51%	6.22%
10	0.5000	0.5016	2,260	2,143	117	5.17%	10.71%	4.49%	100.00%	100.00%	0.00%
			22606	20002	2604						56.24%

IN-time GINI	66.05%
In-time KS	56.24%

Tables 6 and 7 provide the performance results of generalised additive models, the model's rank order the risk monotonically across deciles. The model achieves a Max KS of 60.8/56.24 in development and in-time validation, the model achieves GINI of 71.80/66.05 in development and in-time validation.

Table 8: Performance Results of Logistic Regression - I

Interval	Minmum Score	Maximum Score	Total Accounts	Number of goods	Number of bads	Good Rate	% goods	%bads	Cummulative goods : G(i)	Cummulative bads B(i)	KS
1	0.60130	0.73110	2,260	1,277	983	56.50%	47.56%	4.93%	47.56%	4.93%	42.63%
2	0.53990	0.60130	2,261	658	1,603	29.10%	24.51%	8.05%	72.07%	12.98%	59.09%
3	0.52250	0.53990	2,260	349	1,911	15.44%	13.00%	9.59%	85.07%	22.58%	62.49%
4	0.51490	0.52250	2,261	163	2,098	7.21%	6.07%	10.53%	91.14%	33.11%	58.03%
5	0.51040	0.51490	2,260	95	2,165	4.20%	3.54%	10.87%	94.67%	43.98%	50.70%
6	0.50730	0.51040	2,261	72	2,189	3.18%	2.68%	10.99%	97.36%	54.96%	42.39%
7	0.50490	0.50730	2,261	37	2,224	1.64%	1.38%	11.16%	98.73%	66.13%	32.60%
8	0.50310	0.50490	2,260	17	2,243	0.75%	0.63%	11.26%	99.37%	77.39%	21.98%
9	0.50160	0.50310	2,261	11	2,250	0.49%	0.41%	11.30%	99.78%	88.68%	11.09%
10	0.50000	0.50160	2,260	6	2,254	0.27%	0.22%	11.32%	100.00%	100.00%	0.00%
			22605	2685	19920						62.49%

Development GINI	**76.20%**
Development KS	**62.49%**

Table 9: Performance Results of Logistic Regression - II

Interval	Minmum Score	Maximum Score	Total Accounts	Number of goods	Number of bads	Bad Rate	% goods	%bads	Cummulative goods : G(i)	Cummulative bads B(i)	KS
1	0.60130	0.73111	2260	994	1266	56.02%	4.97%	48.62%	4.97%	48.62%	43.65%
2	0.53990	0.60130	2261	1672	589	26.06%	8.36%	22.62%	13.33%	71.24%	57.91%
3	0.52250	0.53990	2261	1935	326	14.42%	9.67%	12.52%	23.00%	83.76%	60.75%
4	0.51490	0.52250	2260	2071	189	8.36%	10.35%	7.26%	33.36%	91.01%	57.66%
5	0.51040	0.51490	2261	2170	91	4.03%	10.85%	3.49%	44.21%	94.51%	50.30%
6	0.50730	0.51040	2261	2194	67	2.96%	10.97%	2.57%	55.17%	97.08%	41.91%
7	0.50490	0.50730	2260	2222	38	1.68%	11.11%	1.46%	66.28%	98.54%	32.26%
8	0.50810	0.50490	2261	2243	18	0.80%	11.21%	0.69%	77.50%	99.23%	21.73%
9	0.50150	0.50810	2261	2251	10	0.44%	11.25%	0.38%	88.75%	99.62%	10.86%
10	0.50000	0.50150	2260	2250	10	0.44%	11.25%	0.38%	100.00%	100.00%	0.00%
			22606	20002	2604						60.75%

Development GINI	**75.41%**
Development KS	**60.75%**

Tables 8 and 9 provide the performance results of logistic regression, the model's rank order the risk monotonically across deciles. The model achieves a Max KS of 62.4/60.75 in development and in-time validation, the model achieves GINI of 76.80/75.41 in development and in-time validation.

Conclusion

In this paper, we have outlined generalised additive logistic models with their application for marketing response models. It is shown that generalised additive models perform equivalently well with logistic regression. As generalised additive model relaxes the assumption of linearity between the predictors and the response and avoids the problem of model misspecification, which is often prone to happen in generalised linear model, it is easily address by generalised additive models. Secondly, by incorporating non-linear effects, generalised additive model helps discover the hidden pattern of predictors and therefore improves performance on models when they are applied on larger datasets for scoring and also ensures that the models are stable over a period of time.

References

Craven, P., & Wahba, G. (1979). Smoothing noisy data with spline functions: Estimating the correct degree of smoothing by the method of generalized cross-validation Number. *Math.,* 31, 377-403.

Hastie, T. J., & Tibshirani, R. J. (1990). *Generalized additive models.* New York: Chapman & Hall.

Liu, W., & Cela, J. (2007). Improving credit scoring by generalized additive model. SAS global forum 2007 (paper 078-2007).

McCullagh, P., & Nelder, J. (1989). *Generalized linear models*. Chapman and Hall.

Moro, S., Laureano, R., & Cortez, P. (2011). *Using Data Mining for Bank Direct Marketing: An Application of the CRISP-DM Methodology*. In P. Novais et al. (Eds.), Proceedings of the European Simulation and Modelling Conference - ESM'2011, pp. 117-121, Guimarães, Portugal, October, 2011. EUROSIS.

Muller, M. (2000). *Semi-parametric extensions to generalized linear models*. Habilitationsschrift.

Wood, S. N. (2006). Generalized additive model: An introduction with R. Chapman and Hall/CRC.

Stone, C. J. (1985). Additive regression and other non-parametric models. *Annals of Statistics,* 13, 689-705.

Web Reference

http://support.sas.com/kb/32/927.html