Chapter 3
Fitting Generalized Additive Logistic Regression Model with GAM Procedure

“It is the mark of a truly intelligent person to be moved by statistics”
- George Bernard Shaw

3.1 Introduction

The GAM procedure is the most versatile of several new procedures for non-parametric regression. The methodology behind this procedure is more flexible than traditional parametric modeling tools such as linear or non-linear regression. It relaxes the usual parametric assumption and enables us to uncover structure in the relationship between independent variables and dependent variable in exponential family that we might otherwise miss. Hastie and Tibshirani (1990) proposed the fitting of Generalized Additive models using various methods. The general methods of fitting these models are Backfitting algorithm and local scoring algorithm. The backfitting algorithm fits an additive model using any regression type smoothers and it can be used with different smoothers such as smoothing splines and local regression smoothers.

PROC GAM fits the generalized additive model by using a modified term of adjusted dependent variable, with the additive predictor generally taking the role of linear predictor. This procedure uses the local scoring algorithm and the estimation of generalized additive models is accomplished by replacing the weighted linear regression

Most of the contents of this Chapter have been published in Journal of Mathematics and System Science (Ref. Sharma et al. (2013))
by a weighted backfitting algorithm. In fact, it fits a weighted additive model with an iterative algorithm. The GAM procedure implements the B-spline for univariate smoothing components and the thin plate smoothing spline for bivariate smoothing components.

The logistic regression model is extremely popular in the analysis of medical and epidemiological studies. This model measures the relationship between a categorical dependent variable and a set of independent variables which may be continuous or categorical in nature. The model predicts the probability of risk of an individual after specifying the values of predictors in the logistic function. Let y be a binary outcome variable (y=1 if a disease is present and y=0, if absent) and x is a vector of p explanatory variables (risk factors for disease). The aim is to model the posterior probability P(y=1|x), or simply P(x) and such a model describes the risk of presence of disease as a function of the risk factors. The linear logistic model has the form

\[ \text{Logit}[P(x)] = \log \left( \frac{P(x)}{1-P(x)} \right) = \alpha + \beta x. \]

Maximum likelihood method is used to estimate the parameters of linear logistic model. Hastie and Tibshirani proposed the additive non-parametric regression model of the form

\[ \text{Logit}[P(x)] = \alpha + \sum f_j(x_j), \]

where \( f_j \) are the general real-valued functions. Each of the \( f_j \) can be chosen to be either linear, general non-linear (estimated by a scatterplot smoothing) or step functions for discrete covariates. In this case, the score equations are non-linear and one has to find the solution iteratively. Generally, Newton Raphson iterative method is used to estimate the non-linear functions. Generalized additive models have the form

\[ \eta(x) = \alpha + \sum f_j(x_j), \]

where \( \eta \) might be the regression function in a multiple regression or the logistic transformation of the posterior probability \( P(y=1|x) \) in a logistic regression. These models generalize the whole family of generalized linear models \( \eta(x) = \beta'x \), where \( \eta(x) = g(\mu(x)) \) is some transformation of the regression function. To estimate the functions \( f_j(x_j) \) non-parametrically, we use the local scoring algorithm and scatterplot smoother as a building block.
In this Chapter, we discuss two methods of fitting generalized additive logistic regression model, one based on Newton Raphson method and other based on iterative weighted least square method using first and second order Taylor series expansion. The GAM procedure with a specified set of weights, using local scoring algorithm, has been applied to real life data sets. The cubic spline smoother is applied to the independent variables. Based on nonparametric regression and smoothing techniques, this procedure provides powerful tools for data analysis. To link up the problem, some basic models are discussed below. In Sections 3.2 and 3.3, we discuss the fitting of generalized additive logistic model using Newton-Raphson method and iterative weighted least square method respectively. Fitting of these models to real life datasets have been explored in Section 3.4. Local scoring algorithm is discussed in Section 3.5. The SAS programs are listed in Section 3.6.

Linear Models

The linear model is used for regression in a variety of contexts other than ordinary regression. Common examples include log-linear models, logistic regression, the proportional-hazards model for survival data, models for ordinal categorical responses and transformation models. The linear model is a convenient but crude first-order approximation to the predictor surface. However, GAM’s can be used in virtually any setting where linear models are used. Generalized Additive Models allows us to choose from a wide range of distributions for the dependent variable and link functions for the effects of the predictor variables on the dependent variables. For instance, we may specify following links for different distributions

<table>
<thead>
<tr>
<th>DISTRIBUTION</th>
<th>LINK</th>
</tr>
</thead>
<tbody>
<tr>
<td>Normal</td>
<td>Identity</td>
</tr>
<tr>
<td>Gamma</td>
<td>Inverse</td>
</tr>
<tr>
<td>Poisson</td>
<td>Log</td>
</tr>
<tr>
<td>Binomial</td>
<td>Logit</td>
</tr>
</tbody>
</table>
Fitting Generalized Additive Logistic Regression Model with GAM Procedure

Linear regression model is of the form

\[ E[Y|X_1, \ldots, X_k] = \beta_0 + \beta_1 X_1 + \ldots + \beta_k X_k = \beta_0 + \sum_{j=1}^{k} \beta_j X_j. \]

The Generalized Additive Model consists of a random component, an additive component and a link function, linking the two components. It is of the form

\[ g(E[Y|X_1, \ldots, X_k]) = S_0 + \sum_{j=1}^{k} f_j(X_j). \]

Logistic Model

Logistic regression is a mathematical modeling approach that can be used to describe the relationship of several independent variables to a dichotomous dependent variable.

Logistic model is based on logistic function and \( P(x) = \frac{1}{1 + e^{-x}} \)

The model is designed to describe a probability, which is always some number between 0 and 1. In epidemiologic terms, such a probability gives the risk factor of an individual getting a disease. To obtain the logistic model from the logistic function we write \( y \) in the following linearized form

\[ y_i = \beta_0 + \beta_1 x_{i1} + \ldots + \beta_k x_{ik}, \]

where \( x_i = (x_{i1}, x_{i2}, \ldots, x_{ik}) \) denote the \( i \)-th setting of the \( k \) explanatory variables for \( i = 1, 2, \ldots, n \).

Therefore, \( P(x) = \frac{1}{1 + e^{-\left(\beta_0 + \sum \beta_j x_{ij}\right)}} \). Thus,

\[ LogitP(x) = \log \frac{P(x)}{1 - P(x)} = \beta_0 + \sum \beta_j x_{ij}. \]

This is the link known as log odds transformation, the logit, which converts this model into linear form. The coefficients \( \beta_j, j = 0,1,2,\ldots,k \) in this model represent unknown parameters and they can be estimated through maximum likelihood technique.
3.2 Fitting Generalized Additive Logistic Model using Newton-Raphson Method

The basic idea of Generalized Additive logistic model is to replace the linear component \( \sum_{j=1}^{k} \beta_j x_{ij} \) in the logistic model by an additive component \( \sum_{j=1}^{k} f_j(x_{ij}) \), in the functional form of the generalized additive model. Since the joint probability mass function of \( \{y_1, y_2, \ldots, y_k\} \) is the product of \( n \) binomial functions, hence

\[
\prod_{i=1}^{n} [P(x_i)]^{y_i} [1 - P(x_i)]^{n-y_i}
\]

\[
= \prod_{i=1}^{n} \left[ \frac{P(x_i)}{1 - P(x_i)} \right]^{y_i} [1 - P(x_i)]
\]

\[
= \left[ \prod_{i=1}^{n} [1 - P(x_i)] \right] \left[ \prod_{i=1}^{n} \exp \left[ \log \left( \frac{P(x_i)}{1 - P(x_i)} \right)^{y_i} \right] \right]
\]

\[
= \left[ \prod_{i=1}^{n} [1 - P(x_i)] \right] \exp \left[ \sum_{i=1}^{n} y_i \log \left( \frac{P(x_i)}{1 - P(x_i)} \right) \right]. \quad (3.2.1)
\]

Since,

\[
\log \frac{P(x_i)}{1 - P(x_i)} = f_0 + \sum_{j=1}^{k} f_j(x_{ij}), \quad (3.2.2)
\]

therefore, using (3.2.2) the second term of (3.2.1) becomes

\[
\exp \left[ \sum_{i=1}^{n} y_i \left( f_0 + \sum_{j=1}^{k} f_j(x_{ij}) \right) \right]
\]

\[
= \exp \left[ \sum_{i=1}^{n} y_i f_0 + \sum_{j=1}^{k} \sum_{i=1}^{n} y_i f_j(x_{ij}) \right]. \quad (3.2.3)
\]

We also know that

\[
[1 - P(x_i)] = \left[ 1 + \exp \left( f_0 + \sum_{j=1}^{k} f_j(x_{ij}) \right) \right]^{-1} \quad (3.2.4)
\]

Substituting (3.2.3) and (3.2.4) in (3.2.1), we get the likelihood as

\[
L = \prod_{i=1}^{n} \left[ 1 + \exp \left( f_0 + \sum_{j=1}^{k} f_j(x_{ij}) \right) \right] \left[ \exp \left( \sum_{i=1}^{n} y_i f_0 + \sum_{j=1}^{k} \sum_{i=1}^{n} y_i f_j(x_{ij}) \right) \right].
\]
The log likelihood becomes

\[ \log L = \log \left( \sum_{i=1}^{n} y_i \left( f(x_i) + \sum_{j=1}^{n} y_i \sum_{i=1}^{n} f_j(x_{ij}) \right) - \sum_{i=1}^{n} \log \left[ 1 + \exp \left( \sum_{j=1}^{n} f_j(x_{ij}) \right) \right] \). \]

We derive the score equations by differentiating \( L(f) \) w.r.t. elements of \( f \) and setting the result equal to zero. Thus,

\[ \frac{\partial L(f)}{\partial f_i} = \sum_{i=1}^{n} y_i f_i(x_{ia}) - \sum_{i=1}^{n} \log \left[ 1 + \exp \left( \sum_{j=1}^{n} f_j(x_{ij}) \right) \right] f_i(x_{ia}). \] (3.2.5)

Suppose the initial estimator of \( \exp \left( \hat{f}_0 + \sum_{j=1}^{n} \hat{f}_j(x_{ij}) \right) \) is \( P_i \),

\[ \frac{\exp \left( \hat{f}_0 + \sum_{j=1}^{n} \hat{f}_j(x_{ij}) \right)}{1 + \exp \left( \hat{f}_0 + \sum_{j=1}^{n} \hat{f}_j(x_{ij}) \right)} = P_i. \]

Then equating (3.2.5) equal to zero, the score equation becomes

\[ \sum_{i=1}^{n} (y_i - P_i) f_i'(x_{ia}) = 0. \] (3.2.6)

This is similar to the normal equations which we obtain when we apply the method of least squares in linear regression model, that is,

\[ X' \hat{m} = X' \hat{P} \]

where \( \hat{m} = P_i \), and \( X' = f_i'(x_{ia}) \).

If there is prior information about the form of \( f(x) \), e.g. quadratic, cubic or any other polynomial of higher degree, then one can work out with the modified data sets, after taking the first derivative of the polynomial function \( f(x) \).

The method of estimating \( P_i \) or \( f_j \) can be carried out by using Newton Raphson iterative method for solving non-linear equations. It can solve likelihood equations that determine the location at which a function is maximized. This method requires an initial guess for the values that maximize the function. The function is approximated in a neighbourhood of that guess by a second-degree polynomial and the second guess is the location of that polynomial’s maximum value. The function is then approximated in the neighbourhood of the second guess by another second-degree polynomial and the third guess is the location of its maximum. In this way, this method generates a sequence of
guesses, and these guesses converge to the location of maximum when the function is suitable or initial guess is good.

The information matrix is the negative expected value of the matrix of second partial derivatives of the log likelihood, which is called Fisher’s information and is given by

$$I(Q) = -E \left[ \frac{\partial^2 L(f)}{\partial f \partial f^T} \right]$$

Under regularity conditions, the maximum likelihood estimators of parameters have a large-sample normal distribution with covariance matrix equal to the inverse of the transformation matrix, that is, \( \frac{1}{I(Q)} \).

For the logistic regression model, we have

$$\frac{\partial^2 L(f)}{\partial f \partial f^T} = \sum_{i=1}^{n} \left[ \frac{\exp \left\{ f_0 + \sum_{j=1}^{k} f_j(x_j) \right\}}{1 + \exp \left\{ f_0 + \sum_{j=1}^{k} f_j(x_j) \right\}} \right] \frac{f'_i(x_i)}{f'_i(x_i)} f'_i(x_i) f'_i(x_i)$$

$$= \sum_{i=1}^{n} f'_i(x_i) f'_i(x_i) P(x_i) [1 - P(x_i)]$$

(3.2.7)

Since (3.2.7) is not a function of \(\{y_i\} \), the observed and expected second derivative matrices are identical. We estimate the covariance matrix by substituting \(\hat{f} \) into the matrix having elements equal to negative of (3.2.7) and inverting. The estimated covariance matrix has the form

$$\text{cov} (\hat{f}) = [X' \text{diag} \{\hat{P}(1 - \hat{P})\} X]^{-1}$$

(3.2.8)

where \(\text{diag} \{\hat{P}(1 - \hat{P})\} \) denote the \(n \times n\) diagonal matrix having elements \(\{\hat{P}(1 - \hat{P})\} \) on the main diagonal.

**Newton-Raphson Method**

This method determines the value \(\hat{f} \) of \(f \) that maximizes \(g(f)\), where

$$g(f) = \frac{\exp \left\{ f_0 + \sum_{j=1}^{k} f_j(x_j) \right\}}{1 + \exp \left\{ f_0 + \sum_{j=1}^{k} f_j(x_j) \right\}}$$

65
Let \( q' = \left( \frac{\partial g}{\partial f_1}, \frac{\partial g}{\partial f_2}, \ldots \right) \) and \( H \) denote the matrix with entries \( h_{ab} = \frac{\partial^2 g(f)}{\partial f_a \partial f_b} \).

Let \( q^{(0)} \) and \( H^{(0)} \) be terms evaluated at \( f^{(0)} \), the \( t \)th guess for \( f \). At the \( t \)th step in the iterative process (\( t = 0, 1, 2, \ldots \)), \( g(f) \) is approximated near \( f^{(0)} \) by the terms up to second order in its Taylor series expansion written as

\[
Q^{(0)}(f) = g(f^{(0)}) + q^{(0)}(f - f^{(0)}) + \frac{1}{2}(f - f^{(0)})^T H^{(0)}(f - f^{(0)}).
\]

Differentiating w.r.t. \( f \), we get

\[
\frac{\partial Q^{(0)}(f)}{\partial f} = q^{(0)} + H^{(0)}(f - f^{(0)}).
\]

Equating to zero, yields the next guess for \( f \), given as

\[
f^{(t+1)} = f^{(t)} - (H^{(t)})^{-1} q^{(t)}
\]

where we assume that \( H^{(t)} \) is non-singular.

Now suppose that \( g(f) \) is the log likelihood for logistic regression model. Using (3.2.6) and (3.2.7), let

\[
q^{(t)} = \frac{\partial L(f)}{\partial f} \bigg|_{f^{(t)}} = \sum_{i=1}^{n} (y_i - P_i) f(x_i) = X'(y - m^{(t)})
\]

and

\[
h^{(t)}_{ab} = \frac{\partial^2 L(f)}{\partial f_a \partial f_b} \bigg|_{f^{(t)}} = -\sum_{i=1}^{n} x_i(x_i) f(x_i) (P_i^{(t)} - P_i^{(t)}).
\]

Here \( P_i^{(t)} \), the \( t \)th approximation for \( P_i \), is obtained from \( f^{(t)} \) through

\[
P_i^{(t)} = \frac{\exp \left[ f_0 + \sum_{j=1}^{k} f_j^{(t)}(x_j) \right]}{1 + \exp \left[ f_0 + \sum_{j=1}^{k} f_j^{(t)}(x_j) \right]}.
\]

We use \( q^{(t)} \) and \( H^{(t)} \) from the previous formulas to obtain the next value \( f^{(t+1)} \), which is given by
Fitting Generalized Additive Logistic Regression Model with GAM Procedure

\[ f^{(i+1)} = f^{(i)} + \left[ X' \text{diag} \{ P^{(i)} (1 - P^{(i)}) \} X \right]^{-1} X' (y - m^{(i)}) , \]

where \( m^{(i)} = P^{(i)} \).

This is used to obtain \( P^{(i+1)} \) and so on. After making an initial guess \( f^{(0)} \), we use equation (3.2.9) to obtain \( P^{(0)} \) and for \( t > 0 \) the iterations proceed as described earlier.

In the limit, \( P^{(t)} \) and \( f^{(t)} \) converge to the maximum likelihood estimators of \( \hat{P} \) & \( \hat{f} \).

### 3.3 Fitting Generalized Additive Logistic Model using Iterative Weighted Least Square Method

The logit is defined as

\[
\text{Logit}(P(x)) = \log \frac{P(x)}{1 - P(x)} = f_0 + \sum_{j=1}^{k} f_j(x_{ij}).
\]

Using definition of GAM from section 1.3 (of Chapter1), we know that \( E(y|X) = \mu \) and therefore, \( \eta(x) = g(\mu) = \log \frac{P(x)}{1 - P(x)} \).

where \( \eta \) is a function of \( p \) variables.

When \( X = x \) is given and \( Y \sim \text{Bernoulli}(P(x)) \), this implies that

\[
E(y|X = x) = P(x) = \mu ;
\]

\[
V(y|X = x) = P(x)(1 - P(x)) = \mu(1 - \mu).
\]

In usual weighted least squares, the weights are such that they are proportional to the inverse of variance at each level. So, if there are \( n \) number of \( y \) observations over \( x_1, ..., x_k \), then

\[
V(y_i|X_i = x_i) = P(x_i)(1 - P(x_i)) = \mu_i(1 - \mu_i)
\]

\[
\Rightarrow w_i \propto \frac{1}{P(x_i)(1 - P(x_i))} = \frac{1}{\mu_i(1 - \mu_i)} \quad (3.3.1)
\]

where \( w_i \) is the weight for the \( i^{th} \) observation.

#### 3.3.1 Method to Estimate Weights

Let us assume \( y = \eta(x) + \varepsilon \), then given some initial estimator of \( \eta(x) \), we construct a dependent variable, using first order Taylor series expansion as

\[
z_i = \eta_i + (y_i - \mu_i) \left( \frac{\partial \eta}{\partial \mu} \right)_{\mu_i} \]

which implies

\[
z = \eta + (y - P(x_i))g'(P(x_i))
\]
and since logit is one-to-one transformation, hence instead of y, we regress the logit transformed variable z. We have

\[ E(z|X = x) = \eta \]

\[ V(z|X = x) = V(\eta + (y - P(x)))g'(P(x))|X = x) \]

\[ = g'(P(x))^2 V(y|X = x) \quad \text{as } \eta \text{ is a constant under } X = x \]

\[ = g'(P(x))^2 V(y|X = x). \]

Now \( g'(P(x)) = \frac{d}{dP(x)} g(P(x)) = \frac{\partial \eta}{\partial \mu}, \)

where \( g(P(x)) = \eta \) and \( P(x) = \mu. \)

So \( V(z|X = x) = \left( \frac{\partial \eta}{\partial \mu} \right)^2 V(y). \)

Hence, weight for the ith observation

\[ w_i \propto \left( \frac{\partial \eta_i}{\partial \mu_i} \right)^2 V(y_i). \]

The constant of proportionality can be taken as 1 and therefore,

\[ w_i = \left( \frac{\partial \eta_i}{\partial \mu_i} \right)^2 V(y_i)^{-1} \]

\[ = \left( \frac{\partial \mu_i}{\partial \eta_i} \right)^2 V^{-1}(y_i) \quad (3.3.2) \]

Using backfitting algorithm, one may calculate the \((m-1)^{th}\) estimator of \( \eta_i \) as

\[ \eta^{m-1} = f_0 + \sum_{j=1}^{p} f_j^{m-1}(x_j), \]

where \( f_j^{m-1} \) are \((m-1)^{th}\) stage estimators of the non-parametric functions obtained through smoothing.

To estimate the weights, we plug these estimators into formula of \( w_i \) and get

\[ \tilde{w}_i = \left( \frac{\partial \mu_i^{m-1}}{\partial \eta_i^{m-1}} \right)^2 V_i^{-1} \quad \text{where } V_i = Var(y_i). \quad (3.3.3) \]

The process is to be continued until the convergence criterion of backfitting algorithm is satisfied (see Section 3.5).

For some given initial estimator of \( \eta(x) \), and using second order Taylor series, we create a dependent variable of the form

\[ z = \eta + (y - P(x))g'(P(x)) + \frac{(y - P(x))^2}{2!} g''(P(x)) \]

68
Fitting Generalized Additive Logistic Regression Model with GAM Procedure

\[ \eta + (y - P(x)) \frac{\partial \eta}{\partial \mu} + \frac{(y - P(x))^2}{2!} \frac{\partial^2 \eta}{\partial \mu^2}. \]

Hence, \( V(z|X = x) = \left( \frac{\partial \eta}{\partial \mu} \right)^2 V(y - P(x)) + \left( \frac{\partial^2 \eta}{\partial \mu^2} \right)^2 \frac{1}{4} V((y - P(x))^2) + 2 \left( \frac{\partial \eta}{\partial \mu} \right) \left( \frac{\partial^2 \eta}{\partial \mu^2} \right) \text{Cov}\left((y - P(x)), (y - P(x))^2\right) \) (3.3.4)

Consider, \( \text{Cov}\left((y - P(x)), (y - P(x))^2\right) \)

\[ = E\left[(y - P(x))(y - P(x))^2\right] - E(y - P(x))E((y - P(x))^2) \]

\[ = E\left[(y - P(x))^3\right] \quad \text{(as E}(y) = P(x)) \]

\[ = E\left[y^3 - (P(x))^3 - 3yP(x)(y - P(x))\right] \]

\[ = E\left[y^3 - (P(x))^3 - 3y^2P(x) + 3y(P(x))^2\right] \]

\[ = E\left[y^3 - (P(x))^3 - 3yP(x) + 3y(P(x))^2\right] \]

\[ = E(y) - (P(x))^3 - 3P^2(x) + 3(P(x))^3 \]

(since for Bernoulli \( y^3 = y \))

\[ = P(x) + 2(P(x))^3 - 3(P(x))^2 \]

\[ = P(x)[1 + 2(P(x))^2 - 3P(x)]. \] (3.3.5)

Also,

\[ \left((y - P(x))\right)^2 = V[y^2 + (P(x))^2 - 2yP(x)], \; \text{and} \; \text{under } X = x \]

\[ = V(y^2) + 4V(y)P(x)^2 - 4\text{Cov}(y^2, yP(x)) \]

(since for Bernoulli \( y^2 = y \))

\[ = V(y) + 4(P(x))^2V(y) - 4P(x) \text{ Cov}(y, y) \]

\[ = V(y) + 4(P(x))^2V(y) - 4P(x) V(y). \]

Thus \( V(y - P(x))^2 = V(y)(1 - 2P(x))^2 \). (3.3.6)

Moreover, \( V(y - P(x)) = V(y), \; \text{under } X = x \). (3.3.7)

Substituting (3.3.5), (3.3.6) and (3.3.7) in (3.3.4)

\[ V(z|X = x) = \left( \frac{\partial \eta}{\partial \mu} \right)^2 V(y) + \left( \frac{\partial^2 \eta}{\partial \mu^2} \right)^2 \frac{1}{4} (1 - 2P(x))^2 V(y) + 2 \left( \frac{\partial \eta}{\partial \mu} \right) \left( \frac{\partial^2 \eta}{\partial \mu^2} \right) P(x)[2(P(x))^2 - 3P(x) + 1]. \]
For $i^{th}$ observation

$$w_i = \frac{1}{V(z|X = x)}$$

$$= \left[ \left( \frac{\partial \eta_i}{\partial \mu_i} \right)^2 V(y_i) + \left( \frac{\partial^2 \eta_i}{\partial \mu_i^2} \right)^2 \left( \frac{1 - 2P(x)}{4} \right)^2 V(y_i) \right]^{-1} \left[ \left( \frac{\partial^2 \eta_i}{\partial \mu_i^2} \right)^2 \left( \frac{1 - 2P(x)}{4} \right)^2 V(y_i) + 2 \frac{\partial \eta_i}{\partial \mu_i} \left( \frac{\partial^2 \eta_i}{\partial \mu_i^2} \right)^2 P(x)(2P(x)^2 - 3P(x) + 1) \right]^{-1}$$

As done for first order Taylor series, we plug in the estimators of $\eta_i$ in the above equation to estimate weights

$$\tilde{w}_i = \left[ \left( \frac{\partial \eta_i}{\partial \mu_i} \right)^2 V(y_i) + \left( \frac{\partial^2 \eta_i}{\partial \mu_i^2} \right)^2 \left( \frac{1 - 2P(x)}{4} \right)^2 V(y_i) \right]^{-1} \left[ \left( \frac{\partial^2 \eta_i}{\partial \mu_i^2} \right)^2 \left( \frac{1 - 2P(x)}{4} \right)^2 V(y_i) + 2 \frac{\partial \eta_i}{\partial \mu_i} \left( \frac{\partial^2 \eta_i}{\partial \mu_i^2} \right)^2 P(x)(2P(x)^2 - 3P(x) + 1) \right]^{-1}$$

(3.3.8)

For first order Taylor series, the estimated weights $\tilde{w}_i$ (given by (3.3.3)) and for second order Taylor series, $\tilde{w}_i$ (given by (3.3.8)) can be used in GAM procedure.

### 3.4 Fitting Logistic GAM to Real Life Data

During pregnancy certain biomarkers like AFP (Alpha Fetoprotein), uE3 (Unconjugated Estradiol), bhCG (human chorionic gonadotropin) and PAPP-A (pregnancy associated plasma protein A) are produced by the foetus and travel into mother’s blood circulation. These biomarkers can be evaluated twice during pregnancy, that is, during 10-13 weeks of the pregnancy and 15-20 weeks of pregnancy. By estimation of these biomarkers, we can categorize pregnant ladies into low risk or high risk. Low risk means that chances of having chromosomal abnormality or neural tube defects are low whereas high risk does not mean that foetus is affected but it commands further confirmatory test through ultrasonography for evaluation of foetus or karyotyping of amniotic fluid to diagnose whether foetus is affected or not. Along with these biomarkers there are three other predictors viz. age, gestation and weight.

A sample of size 1083 from Govt. Medical College and Hospital (GMCH), Chandigarh is taken. There were 424 cases with complications and 659 with non-complications. The predicted risk of developing a complication based on biomarkers and...
other predictors have been worked out using logistic GAM. The data was analysed using local scoring algorithm method and the results are presented below:

Fig. 3.4.1 shows a scatterplot matrix for all the predictors in the Pregnancy data set. Blue circles and red plus symbols distinguish patients with complications and no complications, respectively. One can observe that the blue circles (complications) indicate that there is no major correlation between variables and therefore the impact of multicollinearity on parameter estimates does not have any effect.

![Scatterplot Matrix](image)

**Figure 3.4.1**

The GAM procedure has been used to fit a model with 4, 6 and 8 degrees of freedom (df). The results for GCV score for smoothing components have been shown in Table 3.4.1.
Table 3.4.1: GCV Score for Smoothing Components

<table>
<thead>
<tr>
<th>Component</th>
<th>GCV Score</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>DF=4</td>
</tr>
<tr>
<td>Spline(age)</td>
<td>4.192</td>
</tr>
<tr>
<td>Spline(gestation)</td>
<td>6.278</td>
</tr>
<tr>
<td>Spline(afp)</td>
<td>3.394</td>
</tr>
<tr>
<td>Spline(hcgb)</td>
<td>3.596</td>
</tr>
<tr>
<td>Spline(uE3)</td>
<td>2.643</td>
</tr>
<tr>
<td>Spline(weight)</td>
<td>5.293</td>
</tr>
</tbody>
</table>

It is evident from Table 3.4.1 that GCV (generalized cross validation) score with 4 df for all predictors was minimum and hence we have presented further results only with 4 df.

In the PROC GAM procedure, 1083 \([424 \text{ (with complications)} + 659 \text{ (without complications)}]\) observations were used with binomial distribution and logit link function. Iterations for local scoring algorithm and backfitting algorithm were carried out till the convergence criterion was met. The procedure stopped after 5 iterations for local scoring and after 1 iteration for backfitting algorithm.

The output given below provides analytical information about the fitted model, including estimates of parameters for the linear portion of the model, fit summary for smoothing components, and the “Analysis of Deviance” table. The linear parameter estimates for predictors age and gestation were significant \((p < 0.01)\) and all other linear predictors were non-significant (Table 3.4.2). The linear predictors may not be important but smoother spline play important role in GAM, since they are estimated in a non-parametric fashion. The chi-square values in the “Analysis of Deviance” (Table 3.4.4) indicate that the three smoothers, that is, age, HCGB, and UE3 are not statistically significant but AFP, gestation and weight are highly significant \((p < 0.01)\). The deviance of the final estimate is found to be 1308.21.

Table 3.4.2: Regression Model Analysis, Parameter Estimates

| Parameter  | Parameter Estimate | Standard Error | t-value | Pr > |t| |
|------------|--------------------|----------------|---------|------|--|--|---|---|
| Intercept  | 1.9353             | 1.2020         | 1.6100  | 0.1077|
| Linear(age)| 0.0543             | 0.0164         | 3.3100  | 0.001*|
| Linear(gestation) | -0.2280    | 0.0572         | -3.9900 | <.0001*|
| Linear(afp) | 0.2808             | 0.1866         | 1.5000  | 0.1327|
| Linear(hcgb)| -0.0569            | 0.1072         | -0.5300 | 0.5954|
| Linear(uE3) | -0.1726            | 0.1768         | -0.9800 | 0.3292|
| Linear(weight)| 0.0019            | 0.0061         | 0.3100  | 0.7530|

*: Significant
Fitting Generalized Additive Logistic Regression Model with GAM Procedure

Table 3.4.3: Smoothing Model Analysis,
Fit Summary for Smoothing Components

<table>
<thead>
<tr>
<th>Component</th>
<th>Smoothing Parameter</th>
<th>DF</th>
<th>GCV</th>
<th>Number of unique observation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Spline(age)</td>
<td>1.00000</td>
<td>3</td>
<td>4.19299</td>
<td>755</td>
</tr>
<tr>
<td>Spline(gestation)</td>
<td>0.99942</td>
<td>3</td>
<td>6.27845</td>
<td>40</td>
</tr>
<tr>
<td>Spline(afp)</td>
<td>0.99999</td>
<td>3</td>
<td>3.39453</td>
<td>166</td>
</tr>
<tr>
<td>Spline(hcgb)</td>
<td>1.00000</td>
<td>3</td>
<td>3.59684</td>
<td>241</td>
</tr>
<tr>
<td>Spline(uE3)</td>
<td>0.99999</td>
<td>3</td>
<td>2.64353</td>
<td>179</td>
</tr>
<tr>
<td>Spline(weight)</td>
<td>0.99998</td>
<td>3</td>
<td>5.29316</td>
<td>102</td>
</tr>
</tbody>
</table>

Table 3.4.4: Smoothing Model Analysis, Analysis of Deviance

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>Chi-Square</th>
<th>Pr &gt; Chi-Square</th>
</tr>
</thead>
<tbody>
<tr>
<td>Spline(age)</td>
<td>3</td>
<td>2.7943</td>
<td>0.4244</td>
</tr>
<tr>
<td>Spline(gestation)</td>
<td>3</td>
<td>76.1399</td>
<td>&lt;.0001*</td>
</tr>
<tr>
<td>Spline(afp)</td>
<td>3</td>
<td>11.4987</td>
<td>0.0093*</td>
</tr>
<tr>
<td>Spline(hcgb)</td>
<td>3</td>
<td>2.6935</td>
<td>0.4413</td>
</tr>
<tr>
<td>Spline(uE3)</td>
<td>3</td>
<td>1.2183</td>
<td>0.7486</td>
</tr>
<tr>
<td>Spline(weight)</td>
<td>3</td>
<td>15.5997</td>
<td>0.0014*</td>
</tr>
</tbody>
</table>

Figure 3.4.2
Fig. 3.4.2 shows the smoothing components for complications along with 95% confidence limits. Partial predictions corresponding to afp, gestation and weight were found to be significant with quadratic pattern for afp and cubic for gestation and weight. After removing the non-significant predictors viz. age, hcg and uc3, Proc GAM procedure was carried out again and the following output was obtained which can be used to evaluate the predicted risk of complications.

### Table 3.4.5: Smoothing Model Analysis

<table>
<thead>
<tr>
<th>Smoothing Component</th>
<th>Smoothing Parameter</th>
<th>DF</th>
<th>GCV</th>
<th>Unique Obs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Spline(gestation)</td>
<td>0.999422</td>
<td>3</td>
<td>6.354205</td>
<td>40</td>
</tr>
<tr>
<td>Spline(afp)</td>
<td>0.999987</td>
<td>3</td>
<td>3.200451</td>
<td>166</td>
</tr>
<tr>
<td>Spline(weight)</td>
<td>0.999984</td>
<td>3</td>
<td>5.451533</td>
<td>102</td>
</tr>
</tbody>
</table>

### Table 3.4.6: Smoothing Model Analysis, Analysis of Deviance

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>Chi-Square</th>
<th>Pr &gt; Chi-Square</th>
</tr>
</thead>
<tbody>
<tr>
<td>Spline(gestation)</td>
<td>3</td>
<td>75.4875</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Spline(afp)</td>
<td>3</td>
<td>11.1468</td>
<td>0.011</td>
</tr>
<tr>
<td>Spline(weight)</td>
<td>3</td>
<td>18.6369</td>
<td>0.0003</td>
</tr>
</tbody>
</table>

![Figure 3.4.3](image-url)
Fig. 3.4.3 shows the partial predictions corresponding to three significant predictors viz. gestation, afp and weight along with 95% confidence limits. It has been seen that gestation and weight have a cubic pattern whereas afp has a quadratic pattern.

Table 3.4.7: Partial Output of Logistic GAM

<table>
<thead>
<tr>
<th>Obs.</th>
<th>gestation</th>
<th>afp</th>
<th>weight</th>
<th>complication</th>
<th>p_complication</th>
<th>Pred</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>15.14</td>
<td>0.4</td>
<td>50</td>
<td>1</td>
<td>0.18429</td>
<td>0.546</td>
</tr>
<tr>
<td>2</td>
<td>17</td>
<td>1.7</td>
<td>57.39</td>
<td>0</td>
<td>0.61855</td>
<td>0.650</td>
</tr>
<tr>
<td>3</td>
<td>16.43</td>
<td>0.9</td>
<td>60</td>
<td>1</td>
<td>0.62432</td>
<td>0.651</td>
</tr>
<tr>
<td>4</td>
<td>20.57</td>
<td>1</td>
<td>50</td>
<td>1</td>
<td>0.67666</td>
<td>0.663</td>
</tr>
<tr>
<td>5</td>
<td>17</td>
<td>0.5</td>
<td>55</td>
<td>1</td>
<td>0.52830</td>
<td>0.629</td>
</tr>
<tr>
<td>6</td>
<td>16.43</td>
<td>0.6</td>
<td>70</td>
<td>0</td>
<td>0.52943</td>
<td>0.629</td>
</tr>
<tr>
<td>7</td>
<td>18.43</td>
<td>0.7</td>
<td>40</td>
<td>0</td>
<td>0.12129</td>
<td>0.530</td>
</tr>
<tr>
<td>8</td>
<td>18.43</td>
<td>0.4</td>
<td>45</td>
<td>0</td>
<td>0.12279</td>
<td>0.531</td>
</tr>
<tr>
<td>9</td>
<td>18.29</td>
<td>0.5</td>
<td>43</td>
<td>0</td>
<td>0.12366</td>
<td>0.531</td>
</tr>
<tr>
<td>10</td>
<td>18.43</td>
<td>0.4</td>
<td>73</td>
<td>1</td>
<td>0.18389</td>
<td>0.546</td>
</tr>
<tr>
<td>11</td>
<td>18.29</td>
<td>0.3</td>
<td>57</td>
<td>1</td>
<td>0.18402</td>
<td>0.546</td>
</tr>
</tbody>
</table>

P_complication: is the predicted logit, Pred: represents the predicted probability.

Table 3.4.7 shows the output generated by the logistic GAM, including the predicted logit and predicted probability. The predicted logit is not of direct interest and the inverse link function is used to convert logit to probability. For example, for observation 5, when gestation week is 17, weight 55, afp 0.5, then according to logistic GAM estimated from the data, the probability of having a complication is 0.629.
3.5 Local Scoring Algorithm

1. Initialize:
   \[ f_0 = g \left( \sum_{i=1}^{n} \frac{y_i}{n} \right), \quad f_1^0 = \ldots = f_p^0 = 0 \]

2. Iterate:
   \[ m = m + 1 \]
   from the adjusted dependent variable
   \[ z_i = \eta_i + (y_i - \mu_i) \cdot \left( \frac{\partial \eta_i}{\partial \mu_i} \right) \]
   where \( \eta^{m-1} = f_0 + \sum_{j=1}^{p} f_j^{m-1}(x_i) \)
   and \( f_j^{m-1} \) are \((m-1)\)th stage estimators of the non-parametric functions obtained through smoothing.

3. The process is to be continued until the change deviance \( D \) is sufficiently small, where
   \[ D = 2 \{ \log(L(y, P)) - \log(L(y, P^{LM})) \} \]
   \[ = 2 \{ \log(L(y, \mu)) - \log(L(y, \mu^{LM})) \} \]
   Or compute \( \Delta(\eta^m, \eta^{m-1}) = \frac{\sum_{j=1}^{p} \| f_j^m - f_j^{m-1} \|}{\sum_{j=1}^{p} \| f_j^{m-1} \|} \).

The GAM procedure uses the following condition as the convergence criterion for the back-fitting algorithm:

\[ \frac{\sum_{j=1}^{p} \sum_{i=1}^{n} w_i \cdot (f_j^{m-1}(x_{ij}) - f_j^m(x_{ij}))^2}{\sum_{j=1}^{p} \sum_{i=1}^{n} w_i \cdot f_j^{m-1}(x_{ij})^2} \]

where \( \epsilon = 10^{-8} \) by default. The weights are defined in equations (3.3.3) and (3.3.8) for first and second order Taylor series expansion.
3.6 Programs in SAS for Section 3.4

SAS code for scatter plot

```
proc sgscatter data=out.data1;
matrix age gestation afp hcgb ue3 weight/
  group=complication markerattrs=(size=4);
run;
```

SAS code for fitting GAM model on all 6 parameters

```
proc gam data=out.data1;
model complication(event='1')=spline(age) spline(gestation)
  spline(afp) spline(hcgb)
  spline(ue3) spline(weight)/
  dist=binomial;
output out=out.data1_DF4 p uclm lclm ;
run;
```

SAS code for fitting GAM on 3 significant parameters

```
proc gam data=out.data1 plots(unpack)=components(clm);
model complication(event='1')=spline(gestation) spline(afp) spline(weight)/
  dist=binomial;
output out=out.data1_DF4_3vars p uclm lclm ;
run;
```

SAS code for table of predicted probability

```
data bphat_1(keep = gestation afp weight complication p complication phat);
set out.data1_DF4_3vars;
phat = exp(p_complication)/(1+exp(p_complication));
run;
```