Generalized Additive Models

Author: Trevor Hastie and Robert Tibshirani

Presenter: Taoran Sheng

Outline

1. The Linear Regression Model and its Smooth Extension
2. The Generalized Linear Models
3. Smooth Extensions of Generalized Linear Models and the Training Method for the Generalized Additive Models

Linear Regression

1. The Linear Regression Model and its Smooth Extension

1) Response Variable: Y
2) A set of predictor random variables X1, X2, ..., Xp
3) A set of n independent realizations
(y1, x11, ..., x1p), ..., (yn, xn1, ..., xnp)

Regression Procedure:

$$E(Y \mid X_1, X_2, \cdots, X_p) = \beta_0 + \beta_1 X_1 + \cdots + \beta_p X_p.$$

Given a sample, estimate coefficients

Smooth Extension(Additive Model)

The additive Model:

$$E(Y \mid X_1, X_2, \cdots, X_p) = s_0 + \sum_{j=1}^{p} s_j(X_j),$$

Sj(.) are smooth functions

Estimate coefficients $\Rightarrow$ Estimate smooth functions

Scatterplot Smoothers

- Our Model:

$$E(Y \mid X) = s(X)$$

- Any reasonable estimate of E(Y | X = x)
- Local average estimates (Ni: neighborhood of xi)

$$\hat{s}(x_i) = \mathrm{Ave}_{j \in N_i}\{y_j\}$$

$$N_i = \left\{\max\left(i - \frac{[wn]-1}{2}, 1\right), \ldots, i-1, i,\right.$$
$$\left. i+1, \ldots, \min\left(i + \frac{[wn]-1}{2}, n\right)\right\}$$

Scatterplot Smoothers

- Any reasonable estimate of E(Y | X = x)

$$\hat{s}(x_i) = \mathrm{Ave}_{j \in N_i}\{y_j\}$$

- Local average estimates (Ni: neighborhood of xi)

1) Running mean ---> arithmetic mean

2) Running lines smoother

$$\hat{s}(x_i) = \hat{\beta}_{0i} + \hat{\beta}_{1i} x_i \Rightarrow$$

$$\hat{\beta}_{0i} = \bar{y}_i - \hat{\beta}_{1i}\bar{x}_i$$

$$\hat{\beta}_{1i} = \frac{\sum_{j \in N_i} (x_j - \bar{x}_i) y_j}{\sum_{j \in N_i} (x_j - \bar{x}_i)^2}$$

where: $\bar{x}_i = (1/n) \sum_{j \in N_i} x_j$ $\bar{y}_i = (1/n) \sum_{j \in N_i} y_j$

Span Selection(Window size)

Consider the Extreme Choices

- W = 1/n Estimate of S(xi) is yi
  high Variance, not smooth
- W = 2 Estimate of S(xi) is global least squares regression
  too smooth, might be biased

- Window Size(Span) is chosen between 1/n and 2

  ----> Bias-Variance Tradeoff

- Minimize Cross-Validation Sum of Squares

$$\mathrm{CVSS}(w) = (1/n)\sum_{1}^{n}(y_i - \hat{s}_w^{-i}(x_i))^2.$$

Outline

1. The Linear Regression Model and its Smooth Extension
2. The Generalized Linear Models
3. Smooth Extensions of Generalized Linear Models and the Training Method for the Generalized Additive Models

Generalized Linear Models(GLM)

- General linear model
  Y = b0 + b1*X1 + ... + bm*Xm

- Generalized linear model
  Y = g(b0 + b1*X1 + ... + bm*Xm)

- Formally, the inverse function of g(...), gi(...) is called the link function
  gi(muY) = b0 + b1*X1 + ... + bm*Xm
  muY: expected value of Y

Generalized Linear Models(GLM)

gi(muY) = b0 + b1*X1 + ... + bm*Xm
muY: expected value of Y

Algorithm for estimate of b=(b0, b1, ..., bm):

- Adjusted dependent variable regression
  Given:
  Etah = b0 + b1*X1 + ... + bm*Xm (current estimate of linear predictor)
  muY: corresponding fitted value

- Adjusted dependent variable:

$$z = \hat{\eta} + (y - \hat{\mu})\left(\frac{d\eta}{d\mu}\right).$$

Generalized Linear Models(GLM)

Adjusted dependent variable:

$$z = \hat{\eta} + (y - \hat{\mu})\left(\frac{d\eta}{d\mu}\right),$$

- Define weights W:

$$(W)^{-1} = \left(\frac{d\eta}{d\mu}\right)^2 V,$$

V: the variance of Y at mu = muh

Generalized Linear Models(GLM)

Adjusted dependent variable regression algorithm:

$$z = \hat{\eta} + (y - \hat{\mu})\left(\frac{d\eta}{d\mu}\right).$$

- 1) Algorithm regressiong z on x1, ..., xp with weights W,
- 2) Obtain an estimate coefficient b,
- 3) Using estimate b to computer the estimate of mu and eta,
- 4) A new z is computed
- 5) Repeat 1) to 4) until the change in deviance is small enough:

$$\mathrm{dev}(y, \hat{\mu}) = 2[l(y) - l(\hat{\mu})]$$

Outline

1. The Linear Regression Model and its Smooth Extension
2. The Generalized Linear Models
3. Smooth Extensions of Generalized Linear Models and the Training Method for the Generalized Additive Models

Smooth Extension of GLM

- S1(.), ..., Sp(.) are smooth functions:

$$\eta = \beta_0 + \beta_1 X_1 + \cdots + \beta_p X_p$$

$$\downarrow$$

$$\eta = s_0 + \sum_1^p s_j(X_j)$$

$$\downarrow$$

Generalized Additive Model: $g(\mu) = \eta$

Estimation – Backfitting Algorithm

Suppose $Y = s_0 + \sum_{j=1}^{p} s_j(X_j) + \varepsilon$ is correct

Define the partial residual: $R_j = Y - s_0 - \sum_{k \neq j} s_k(X_k)$

Then: $E(R_j \mid X_j) = s_j(X_j)$

Minimize: $$E\left(Y - s_0 - \sum_{j=1}^{p} s_j^m(X_j)\right)^2$$

Estimation – Backfitting Algorithm

Backfitting Algorithm

Initialization: $s_0 = E(Y), \quad s_1^1(\cdot) \equiv s_2^1(\cdot) \equiv \cdots \equiv s_p^1(\cdot) \equiv 0, \quad m = 0.$

Iterate: $m = m + 1$

for $j = 1$ to p do:

$$R_j = Y - s_0 - \sum_{k=1}^{j-1} s_k^m(X_k) - \sum_{k=j+1}^{p} s_k^{n-1}(X_k)$$

$$s_j^m(X_j) = E(R_j \mid X_j).$$

Until: $\mathrm{RSS} = E\left(Y - s_0 - \sum_{j=1}^{p} s_j^m(X_j)\right)^2$ fails to decrease.

General Local Scoring Algorithm

General Local Scoring Algorithm

Initialization: $s_0 = g(E(y)), \quad s_1^0(\cdot) \equiv s_2^0(\cdot) \equiv \cdots$

$$\equiv s_p^0(\cdot) \equiv 0, \quad m = 0.$$

Iterate: $m = m + 1$

1. Form the adjusted dependent variable

$$Z = \eta^{m-1} + (Y - \mu^{m-1})(\partial\eta/\partial\mu^{m-1}),$$

where

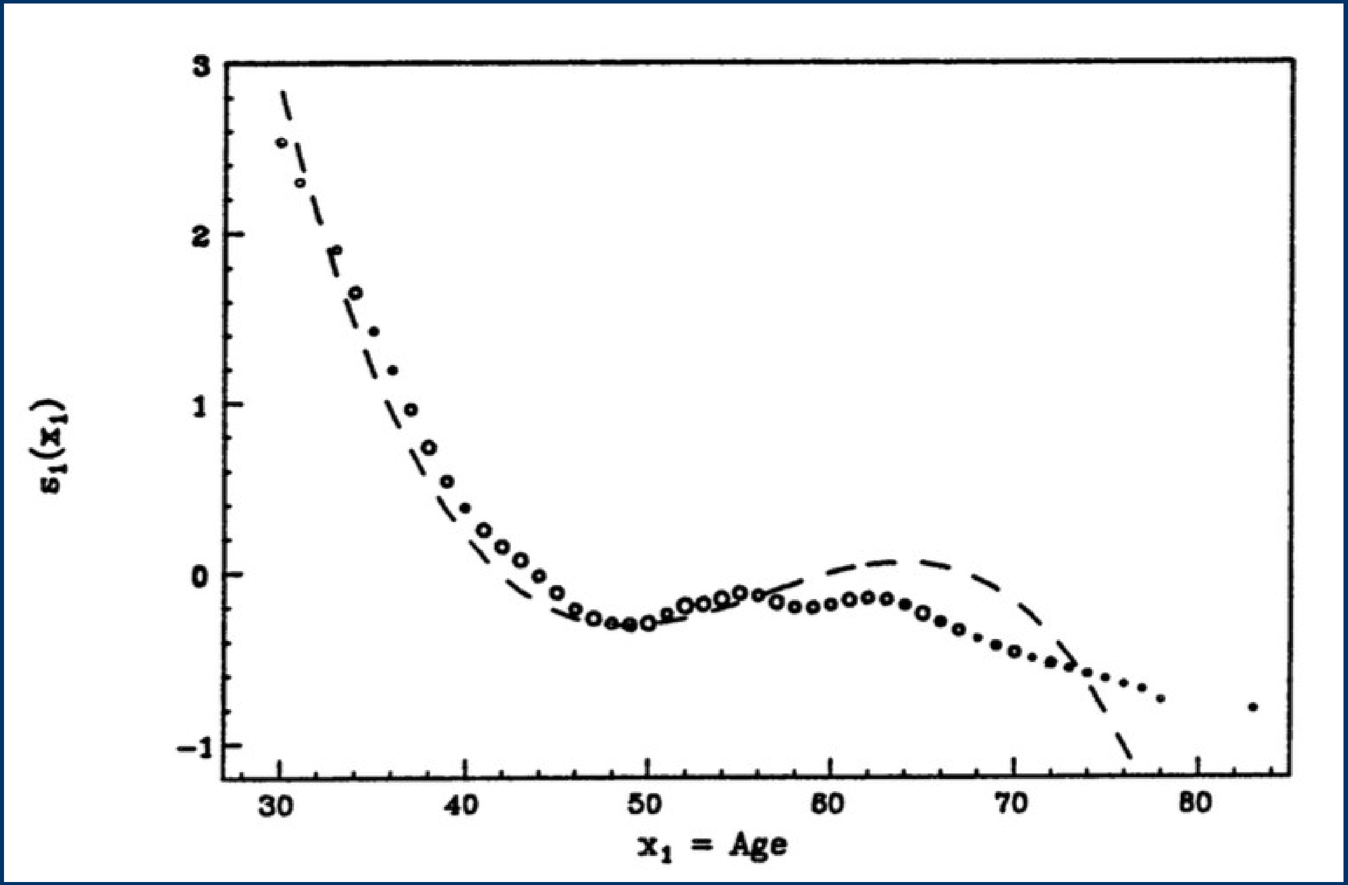
$$\eta^{m-1} = s_0 + \sum_{j=1}^{p} s_j^{m-1}(X_j) \quad \text{and}$$

$$\eta^{m-1} = g(\mu^{m-1}).$$

2. Form the weights $W = (\partial\mu/\partial\eta^{m-1})^2 V^{-1}$.
3. Fit an additive model to Z using the backfitting algorithm with weights W, we get estimated functions $s_j^m(\cdot)$ and model η^m.

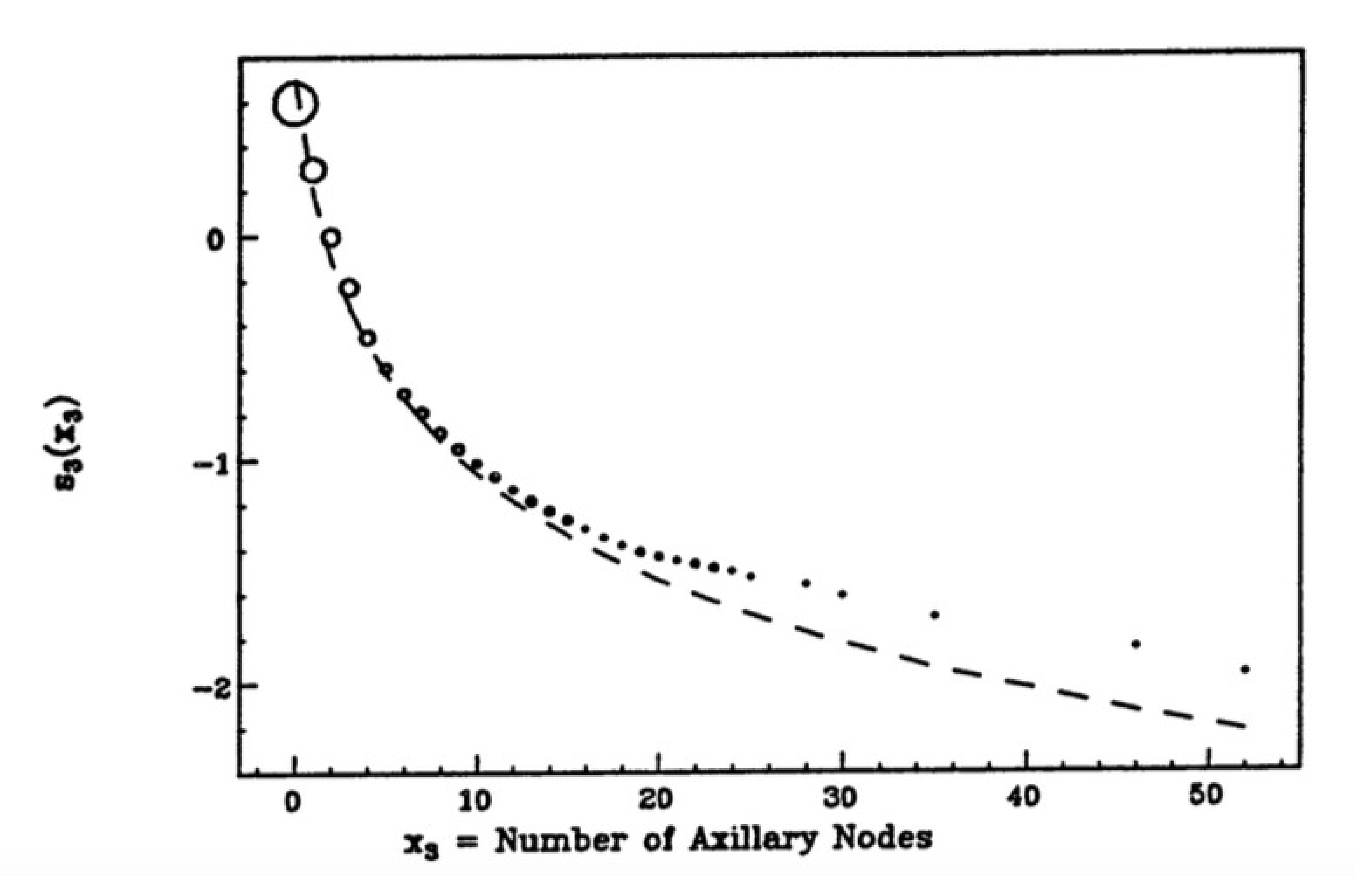
Until: $E\,\mathrm{dev}(Y, \mu^m)$ fails to decrease.

Experiment



- Breast Cancer Dataset

Experiment



- Breast Cancer Dataset

Thanks!