北京航空航天大學
经济管理学院
BEIHANG UNIVERSITY
SCHOOL OF ECONOMICS AND MANAGEMENT

Generalized Linear Models

Lecture 11: Additive models

Outline

Additive models

- Multiple linear regression model

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_p x_{ip} + \varepsilon_i$$

- Replace each linear component $\beta_j x_{ij}$ with a (smooth) non-linear function $f_j(x_{ij})$.

$$y_i = \beta_0 + \sum_{j=1}^{p} f_j(x_{ij}) + \varepsilon_i,\ \varepsilon_i \sim N(0, \sigma^2).$$

Additive models

- By assuming additive surface, we can avoid curse of dimensionality.
- Restricts complexity but a much richer class of surfaces than parametric models (e.g., nonlinear relationships).
- Need to estimate p one-dimensional functions instead of one p-dimensional function.
- Usually set each f_j to have zero mean.
- Some f_j may be linear.

Additive models

- Up to p different bandwidths to select.
- Estimated functions, f_j, are analogues of coefficients in linear regression.
- Interpretation easy with additive structure.

Additive models

- In its basic form, the additive model will do poorly when strong interactions exist.
- Allow interaction between two continuous variables x_j and x_k by fitting a bivariate surface $f_{j,k}(x_j, x_k)$.
- Allow interaction betwen factor x_j and continuous x_k by fitting separate functions $f_{j,k}(x_k)$ for each level of x_j.

Additive models in R

- `gam` package: more smoothing approaches, uses a backfitting algorithm for estimation.
- `mgcv` package: simplest approach, with automated smoothing selection and wider functionality.
- `gss` package: smoothing splines only

Estimation in `gam`

Back-fitting-algorithm (Hastie and Tibshirani, 1990)

1. Set $\beta_0 = \bar{y}$.
2. Set $f_j(x) = \hat{\beta}_j x$ where $\hat{\beta}_j$ is OLS estimate.
3. For $j = 1, \dots, p, 1, \dots, p, 1, \dots, p, \dots$

$$f_j(x) = S(x_j, y - \beta_0 - \sum_{i \neq j} f_i(x_i))$$

where $S(x, u)$ means univariate smooth of u on x.

Iterate step 3 until convergence.

Estimation in `gam`

Back-fitting-algorithm (Hastie and Tibshirani, 1990)

1. Set $\beta_0 = \bar{y}$.
2. Set $f_j(x) = \hat{\beta}_j x$ where $\hat{\beta}_j$ is OLS estimate.
3. For $j = 1, \dots, p, 1, \dots, p, 1, \dots, p, \dots$

$$f_j(x) = S(x_j, y - \beta_0 - \sum_{i \neq j} f_i(x_i))$$

   where $S(x, u)$ means univariate smooth of u on x.

Iterate step 3 until convergence.

- S could be *any* univariate smoother.
- $y - \beta_0 - \sum_{i \neq j} f_i(x_i)$ is a “partial residual”

Outline

Generalized additive models (GAM)

Generalized Linear Model (GLM)

- Distribution of y
- Link function g
- $E(y \mid x_1, \ldots, x_p) = \mu$ where $g(\mu) = \beta_0 + \sum_{j=1}^{p} \beta_j x_j$.

Generalized additive models (GAM)

Generalized Linear Model (GLM)

- Distribution of y
- Link function g
- $\mathrm{E}(y \mid x_1, \dots, x_p) = \mu$ where $g(\mu) = \beta_0 + \sum_{j=1}^{p} \beta_j x_j$.

Generalised Additive Model (GAM)

- Distribution of y
- Link function g
- $\mathrm{E}(y \mid x_1, \dots, x_p) = \mu$ where $g(\mu) = \beta_0 + \sum_{j=1}^{p} f_j(x_j)$.

Examples:

- Y binary and $g(\mu) = \log[\mu(1-\mu)]$. This is a logistic additive model. (GAM for classification!)
- Y normal and $g(\mu) = \mu$. This is a standard additive model.

Estimation

Hastie and Tibshirani describe method for fitting GAMs using a method known as "local scoring" which is an extension of the Fisher scoring procedure.

Pros of GAM

- GAMs allow us to fit a non-linear f_j to each x_j, so that we can automatically model non-linear relationships that standard linear regression will miss. This means that we do not need to manually try out many different transformations on each variable individually.
- The non-linear fits can potentially make more accurate predictions for the response y.
- Because the model is additive, we can still examine the effect of each x_j on y individually while holding all of the other variables fixed. Hence if we are interested in inference, GAMs provide a useful representation.
- The smoothness of the function f_j for the variable x_j can be summarized via degrees of freedom.

Cons of GAM

- GAM is restricted to be additive. With many variables, important interactions can be missed.However, as with linear regression, we can manually add interaction terms to the GAM model by including additional predictors of the form $x_j \times x_k$. In addition we can add low-dimensional interaction functions of the form $f_{jk}(x_j, x_k)$ into the model; such terms can be fit using two-dimensional smoothers such as local regression, or two-dimensional splines (not covered here).

Outline

Generalised additive mixed models (GAMM)

- GAMM manages to combine the three major themes of this book.
  1. The response can be nonnormal from the exponential family of distributions.
  2. The error structure can allow for grouping and hierarchical arrangements in the data.
  3. Finally we can allow for smooth transformations of the response.