MONASH University

MONASH BUSINESS SCHOOL

# ETC3580: Advanced Statistical Modelling

Week 11: Additive models

# Outline

# Additive models

Avoid curse of dimensionality by assuming additive surface:

$$y = \beta_0 + \sum_{j=1}^{p} f_j(x_j) + \varepsilon$$

where $\varepsilon \sim N(0, \sigma^2)$.

- Restricts complexity but a much richer class of surfaces than parametric models.
- 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.
- Generalization of multiple regression model

$$y = \beta_0 + \sum_{j=1}^{p} \beta_j x_j + \varepsilon$$

  which is also additive in its predictors.
- Estimated functions, $f_j$, are analogues of coefficients in linear regression.
- Interpretation easy with additive structure.

# Additive models

- Categorical predictors: fit constant for each level as for linear models.
- 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

**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 $j = 1, \ldots, p, 1, \ldots, p, 1, \ldots, p, \ldots$

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

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

Iterate step 3 until convergence.

# Estimation

**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 $j = 1, \ldots, p, 1, \ldots, p, 1, \ldots, p, \ldots$

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

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"

# Estimation

**Regression splines**

No need for iterative back-fitting as the model can be written as a linear model.

**Penalized regression splines**

No need for iterative back-fitting as the model can be written as a linear mixed-effects model.

# Inference for Additive Models

Each fitted function can be written as a linear smoother $\hat{f}_j = \boldsymbol{S}_j \boldsymbol{y}$ for some $n \times n$ matrix $\boldsymbol{S}_j$.

$\hat{\boldsymbol{f}}(\boldsymbol{x})$ is a linear smoother. Denote smoothing matrix as $\boldsymbol{S}$:

$$\hat{\boldsymbol{f}}(\boldsymbol{x}) = \boldsymbol{S}\boldsymbol{y} = \beta_0 \mathbf{1} + \sum_{j=1}^{p} \boldsymbol{S}_j \boldsymbol{y}$$

where $\mathbf{1} = [1, 1, \ldots, 1]^T$. Then $\boldsymbol{S} = \sum_{j=0}^{p} \boldsymbol{S}_j$ where $\boldsymbol{S}_0$ is such that $\boldsymbol{S}_0 \boldsymbol{y} = \beta_0 \mathbf{1}$.

Thus all inference results for linear smoothers may be applied to additive model.

# Outline

# Generalised additive models

## Generalized Linear Model (GLM)

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

# Generalised additive models

## 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 $g(\mu) = \beta_0 + \sum_{j=1}^{p} f_j(x_j)$.

# Generalised additive models

**Examples:**

- $Y$ binary and $g(\mu) = \log[\mu(1-\mu)]$. This is a logistic additive model.
- $Y$ 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.