Statistical Methods - Nonparametric Regression

Lecture 7

Mattias Villani

Sveriges Riksbank and Stockholm University

May 6, 2010

Overview Lecture 7

- Additive models
- Surface fitting
- Nonparametric Generalized linear models
- Generalized additive models (GAMs)
- Flexible error distribution - continuous reponse variable
- Flexible error distribution - exponential family reponse

Additive models

- How do we extend splines to situations with more than one covariate?
- Additive models. Let $x = (x_1, x_2, ..., x_q)'$

$$y = \alpha + f_1(x_1) + ...f_q(x_q) + \varepsilon,$$

  where $f_1, ..., f_q$ are smooth nonparametric functions, e.g. splines.
- Identify by assuming: $\sum_{i=1}^{n} f_j(x_{ij}) = 0 \forall j$.
- Additive models (also for GLM-type responses) can be fitted with the back-fitting algorithm (HTF, Algorithm 9.1, GAM package in R):
  - Step 0: Initialize $\hat{\alpha} = \bar{y}, \hat{f}_j = 0, \forall i, j$.
  - Step 1: Cycle $j = 1, 2, ..., q, 1, 2, ..., q, ...$

$$\hat{f}_j \leftarrow \mathcal{S}_j \left[\left\{ y_i - \hat{\alpha} - \sum_{k \neq j} \hat{f}_k(x_{ik}) \right\}_{i=1}^{n} \right]$$

$$\hat{f}_j \leftarrow \left[\hat{f}_j - \frac{1}{n} \sum_{i=1}^{n} \hat{f}_j(x_{ij}) \right]$$

  until the functions $\hat{f}_j$ change less than s prespecified threshold.

Surface fitting

- Additive models assume that there are no interactions between covariates.
- Splines with radial basis functions replaces $(x - \kappa)_+^p$ with

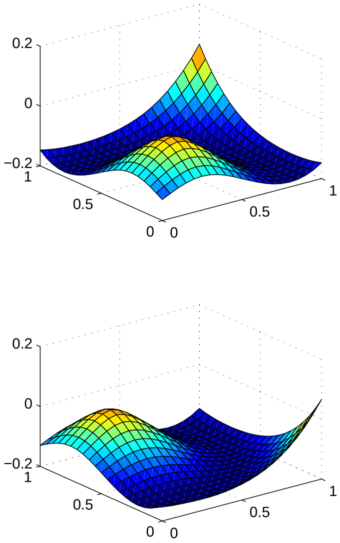
$$r\left(|x - \kappa|\right)$$

  where $r(u)$ is some smooth function from $\mathbb{R} \rightarrow \mathbb{R}$.
- Examples: $r(u) = u \ln u$ (thin plate), $r(u) = (a + u^2)^{1/2}$ (multiquadric) and $r(u) = u^3$ (cubic).
- Radial basis functions can be easily generalized to more than one covariate

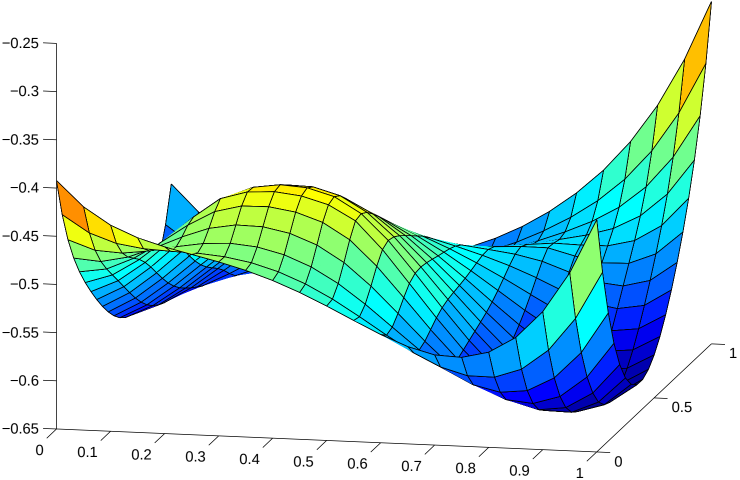
$$r\left(\|\mathbf{x} - \kappa\|\right),$$

  where $x = (x_1, ..., x_q)'$ is a q-dimensional covariate vector and $\kappa = (\kappa_1, ...\kappa_q)'$ is a q-dimensional knot vector.
- Same principle as before: each knot corresponds to a covariate. Knots can be chosen by a clustering algorithm in covariate space.

Thin plate splines



Thin plate spline fit



- GLMs: $y_1, ...y_n$ are independent conditional on the covariates and

$$y_i | x_i \sim ExpFamily(\theta_i)$$

$$g(\theta_i) = x_i' \beta,$$

where $ExpFamily(\theta)$ is some distribution in the exponential family (e.g. normal, Poisson, gamma, ...) with parameter θ.

- $g()$ is the link function that link the parameters in the distribution (θ) to the linear predictor, $x_i'\beta$. Examples:
  - Identity link: $g(\theta) = \theta$
  - Log link: $g(\theta) = ln(\theta)$
  - Logit link: $g(\theta) = \ln \frac{\theta}{1-\theta}$
- GLMs be estimated by a unified Newton-Raphson algorithm (with Fisher Scoring) which goes under the name **iteratively re-weighted least squares**.

Nonparametric Generalized Linear Models

- It is now obvious how to extend GLMs to the spline setting: extend the covariate vector x with additional bases, just like in the usual regression case.
- Additive models and surface models can be handled similarly. The back-fitting algorithm can be extended by:
  - estimating each smoother using the usual iteratively re-weighted least squares algorithm
  - redefining the fitting criterion in terms of deviance

Flexible modeling of the error distribution

- So far we have (implicitly) assumed the errors to be Gaussian (when the response is continuous) or belonging to the exponential family (when the responses are counts, binary or proportions).
- What if the error distribution is mis-specified? Does it matter? It can matter for two reasons:
  - Wrong error distribution may ruin $E(y|x)$, and/or the uncertainty about $E(y|x)$.
  - We often care about higher order moments ($Var(y|x)$ is the focus in financial analysis, $Pr(y > c|x)$ is important in meteorology) or even the whole distribution $p(y|x)$.

Flexible error distribution - continuous reponse

- Student t-distribution

$$y_i = x_i'\beta + \varepsilon_i,\ \ \varepsilon_i \overset{iid}{\sim} t_\nu(0, \sigma^2).$$

- Exponential power distribution with parameters μ, α and β:

$$p(x) = \frac{\beta}{2\alpha\Gamma(1/\beta)} e^{-(|x-\mu|/\alpha)^\beta}$$

- Mixture of normals

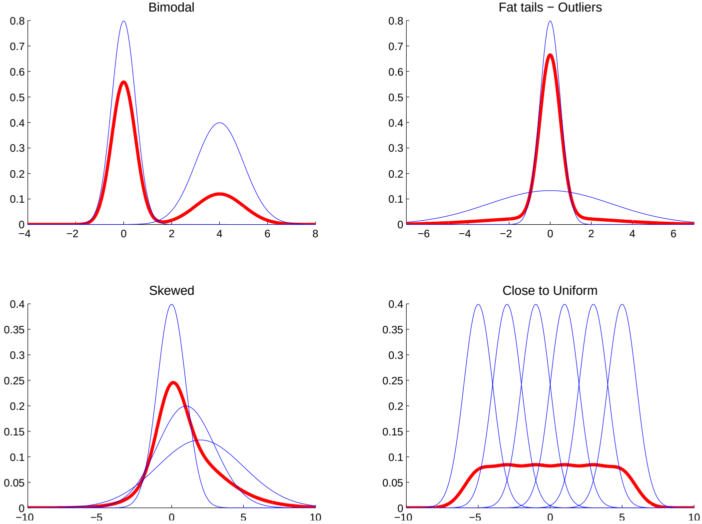
$$y_i = x_i'\beta + \varepsilon_i,\ \ \varepsilon_i \overset{iid}{\sim} MoN(0, \sigma^2, \pi),$$

  where $\pi = (\pi_1, ..., \pi_q)$ and $\sigma^2 = (\sigma_1^2, ..., \sigma_q^2)$.
- The mixture of normals model is defined by: $z \sim MoN(\mu, \sigma^2, \pi)$ if

$$p(z) = \sum_{j=1}^{q} \pi_j \cdot N(z|\mu_j, \sigma_j^2),$$

  where $\sum_{j=1}^{q} \pi_j = 1$.
- (Box-Cox)

Mixture of Normals



Flexible error distribution - exponential family reponse

- Many members of the exponential family are very restrictive since the mean and variance are deterministically related:

$$E(y|x) = \mu(x) = b'(\theta)$$
$$Var(y|x) = a(\phi)b''(\theta) = a(\phi)V(\mu)$$

- Example 1: $y|x \sim Poisson(\mu)$, then $E(y|x) = Var(y|x) = \mu(x)$. Large mean = large variance.
- Example 2: $E(y|x) = np(x)$, $Var(y|x) = np(x)[1 - p(x)]$. Large mean = small variance.
- We can obtain more flexibility using an over-dispersed model:
  - Poisson can be generalized to Negative Binomial
  - Binomial can be generalized to Beta-Binomial
  - The Exponential family can be extend to allow for over-dispersion using Efron's double exponential family.

Flexible error distribution - exponential family reponse

- We can play other tricks to get more flexibility as well. Example: Zero-inflated Poisson model

$$y|x \sim \begin{cases} \mathit{Poisson}[\mu(x)] & \text{with probability } \pi \\ 0 & \text{with probability } 1-\pi \end{cases}$$

- Mixtures of GLMs.
- Bayesian non-parametrics. Put a prior on the class of all distributions. Buzz words: Dirichlet process priors and Polya trees.

Multinomial regression models

- Sometimes the response vector is categorical: $y_i \in \{1, 2, ..., C\}$. Examples: Choice of brands in marketing research {'CocaCola','Fanta','Sprite','UbuntuCola,'Other'}. Modes of transportation {'Train','Bus','Car','Bike','Walk'}.
- We want to explain peoples choice of transportation using a bunch of covariates.
- Multinomial logit:

$$\Pr(y = c|x) = \frac{\exp(x'\beta_c)}{\sum_{j=1}^{C} exp(x'\beta_j)}$$

  where $\beta_1 = 0$ for identification. Interpretation

$$\ln \frac{\Pr(y = c|x)}{\Pr(y = 1|x)} = x'\beta_c$$

  since $\beta_1 = 0$.
- Multinomial probit can also be used.