STAT 151A Lecture 39

Henry Liev

1 December 2025

Remark 0.1

So far: Models have the form $Y = \mathbf{X}\beta + \varepsilon$
Requires that $\mathbb{E}[Y|\mathbf{X}]$ is a hyperplane
What if $\mathbb{E}[Y|\mathbf{X}]$ is actually nonlinear or "curved"
Seen this in special cases
Transformations - linear after a specific transformation
Logistic Regression - linear after a specific transformation
This week: Be more general about nonlinearity

Remark 0.2 (Roadmap)

Univariate x (Regression Splines), Generalized Additive Models (Multivariate x):
Polynomial Regression
Piecewise Constant
Regression Trees
Big picture: Move away from thinking about individual coefficients.
Basis expansion: Model in ?

Remark 0.3 (Polynomial Regression)

$y_i = \hat{\alpha} + \hat{\beta}x_i \to y_i \approx \hat{\alpha} + \hat{\beta}_1 x_i + \hat{\beta}_2 x_i^2 + \hat{\beta}_3 x_i^3 + \cdots$
Fits very nicely into our traditional regression framework: Define a design matrix

$$\mathbf{X} = \begin{pmatrix} 1 & x_1 & x_1^2 & x_1^3 & \cdots \\ 1 & x_2 & x_2^2 & x_2^3 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \\ 1 & x_n & x_n^2 & x_n^3 & \cdots \end{pmatrix}$$

Proceed as usual: $\hat{\beta} = \left(\mathbf{X}^\intercal\mathbf{X}\right)^{-1}\mathbf{X}^\intercal y$ and so on, $\hat{y}, \vec{e}, R^2$, etc.
Inference: Assume NLM
$Y = \mathbf{X}\beta + \varepsilon, \varepsilon \sim \mathcal{N}(\vec{0}, \sigma^2\mathbb{I}_n)$
All of the usual diagnostics and inference results still hold
Biggest difference: We don't really care about individual $\hat{\beta}_j$'s anymore, the role or impact of change of difference in x_i is now explained by many different $\hat{\beta}_j$'s
Still might want to get uncertainty estimates for $\hat{y}_i$ or $\mathbb{E}[\hat{y}_i|x_i]$
$\hat{y}_i = \hat{f}(x_i) = \hat{\beta}_0 + \hat{\beta}_1 x_i + \hat{\beta}_2 x_i^2 + \cdots + \hat{\beta}_d x_i^d$
Assume NLM: and let $\vec{z}_i = \begin{pmatrix} 1 & x_i & x_i^2 & \cdots & x_i^d \end{pmatrix}$
$\text{Var}(\hat{y}_i) = z_i^\intercal \left(\sigma^2(\mathbf{Z}^\intercal\mathbf{Z})^{-1}\right) z_i$
Also $\hat{y}_i \sim N, \hat{\sigma}^2 \sim \chi^2_{n-p-1}$
Create pointwise CI for $\hat{f}(x_i) : \hat{f}(x_i) \pm t_{1-\alpha/2,n-p-1} \cdot \hat{\sigma}\sqrt{z_i^\intercal(\mathbf{Z}^\intercal\mathbf{Z})^{-1}z_i}$

Remark 0.4 (Downsides)

- Can overfit a lot, especially for high powers of x
- Fitted models do very weird things outside the range of observed data (global implications for every point)

Remark 0.5 (Piecewise Constant Regresion)

Goal: keep local information local
$\vec{x} \to$ choose quantiles $c_1, \ldots, c_k$
Define $C_0(x_i) = 1\{x_i < c_1\}$
$C_1(x_i) = 1\{c_1 \leq x_i < c_2\}$
$\vdots$
$C_k(x_i) = 1\{c_k \leq x_i\}$
Create a design matrix

$$\mathbf{X} = \begin{pmatrix} \uparrow & \uparrow & & \uparrow \\ C_0(\vec{x}) & C_1(\vec{x}) & \cdots & C_k(\vec{x}) \\ \downarrow & \downarrow & & \downarrow \end{pmatrix}$$

Fit a linear model
$\hat{y}_i = \sum_{j=1}^{k} \hat{\beta}_j C_j(x_i)$
Piecewise constant, mean of each quantile
Good: Very simple and does not extrapolate
Bad: Overfits with small bins, choosing bins is a headache, discontinuities are ugly and make it hard to model trends

Remark 0.6 (Big Idea)

Combine piecewise constant and polynomial regression $\to$ Both do basis expansion, take a single $\vec{x} \to$ turn into many columns
First attempt: Piecewise polynomials $\to$ not good $\to$ Take a piecewise polynomial model:
- Add restrictions

(1) Require continuity of cut points (knots), functions agree

(2) Require differentiability at cut points, functions derivatives agree (regression spline)