Splines Summary

Saturday, November 25, 2023 23:24

Basis Function – Just any transformation on original var x_i

e.g. polynomial → $\beta_1 x_i + \beta_2 x_i^2 + \beta_3 x_i^3 \dots + \varepsilon_i$ → Polynomial fcns bad at modeling ends

periodic → $\beta_1 x_i + \beta_2 I_{x_i > 0} + \dots + \varepsilon_i$

indicator → $\beta_1 \sin(x_i/k) + \beta_2 \cos(x_i/k) + \dots + \varepsilon_i$

Then the model becomes:

$$y_i = \underbrace{\sum_{m=1}^{M} \beta_m b_m(x_i)}_{\text{aka } g(x_i)} + \epsilon_i$$

($b_m(x_i)$ → basis function)

Piecewise regression – where the basis fcns of x_i interact w/ regional dummy vars. Discontinuous at knots ("jump discontinuities")

E.g. Piecewise Quadratic:

$$y_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \beta_3 D_{1i} + \beta_4 x_i D_{1i} + \beta_5 x_i^2 D_{1i} + \beta_6 D_{2i} + \beta_7 x_i D_{2i} + \beta_8 x_i^2 D_{2i}$$

$$D_{1i} = \begin{cases} 1 & 37 \le x_i < 42 \\ 0 & \text{otherwise} \end{cases} \qquad D_{2i} = \begin{cases} 1 & 42 \le x_i \\ 0 & \text{otherwise} \end{cases}$$

SPLINES

connected/continuous at knots (unlike piecewise)

LINEAR SPLINES

Discontinuous at knots IF you take 1st derivative

Example:

$$y_i = \beta_0 + \beta_1 x_i + \beta_2 (x_i - 36.5)_+ + \beta_3 (x_i - 41.5)_+$$

$\beta_1 x_i$ → global slope; $(x_i - 36.5)_+ = x_i - 36.5$ if $x \ge 36.5$, 0 otherwise; $(x_i - 41.5)_+ = x_i - 41.5$ if $x \ge 41.5$, 0 otherwise

CUBIC SPLINES

Continuous 2nd-derivatives

Example:

$$g(x_i) = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \beta_3 x_i^3 + \beta_4 (x_i - 36.5)_+^3 + \beta_5 (x_i - 41.5)_+^3$$

POLYNOMIAL SPLINES

The general form

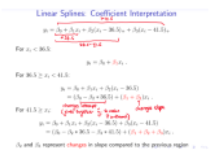
Truncated Power Formulation:

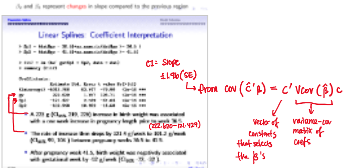
$$y_i = \beta_0 + \sum_{j=1}^{d} \beta_j x_i^j + \sum_{m=1}^{M} \beta_{d+m} (x_i - k_m)_+^d$$

$j = 1, \dots, d$ degree polynomial

$m = 1, \dots, M$ knots

(d=3 → cubic splines; d=1 → linear splines)





Do not interpret coefs individually (unlike linear)!

B-SPLINES

A different way of representing polynomial splines

Intuition: Turns $(2\beta_1)\left(\frac{x_1}{2}\right)$ into $\beta_1 x_1$

Scaled between 0 and 1

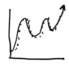
More numerical stability

Do not interpret coefs individually

Model Selection

CV can help us pick.

Choice of basis function doesn't have large impact on model fit, esp. if enough knots and $g(x)$ is smooth



Overfitting
Low bias
High variance
(when you use a test dataset)

Underfitting
Low variance
High bias

$$\text{LOOCV} = \frac{1}{n} \sum_{i=1}^{n} \left(y_i - \hat{g}_i^{(-i)} \right)^2$$

Leave one out cross-validation. Estimates pop. MSE aka $E[(Y - \hat{g}(x))^2]$ with a sample MSE.

($(-i)$: "not i"; $\hat{g}_i^{(-i)}$: estimate y_i on all data except y_i)

$$\text{LOOCV (OLS)} = \frac{1}{n} \sum_{i=1}^{n} \frac{(y_i - \hat{y}_i)^2}{(1 - H_{ii})^2}$$

ONLY IN LINEAR REG.

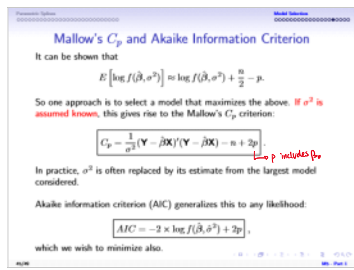
(H_{ii}: the i^{th} diagonal element of the hat matrix $X(X'X)^{-1}X'$)

$$\text{GCV} = \frac{1}{n} \frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{(1 - n^{-1} tr(H))^2}$$

GCV replaces each H_{ii} by the average of all H_{ii}. A weighted version of CV.

($tr(H)$: trace aka diagonals of H i.e. $\frac{1}{n} \sum_{i=1}^{n} H_{ii}$ aka the average)

Want to MINIMIZE GCV or CV.



Parametric Splines | Model Selection

Mallow's C_p and Akaike Information Criterion

It can be shown that

$$E\left[\log f(\hat{\beta}, \sigma^2)\right] \approx \log f(\hat{\beta}, \sigma^2) + \frac{n}{2} - p.$$

So one approach is to select a model that maximizes the above. If σ^2 is assumed known, this gives rise to the Mallow's C_p criterion:

$$C_p = \frac{1}{\sigma^2}(\mathbf{Y} - \hat{\beta}\mathbf{X})'(\mathbf{Y} - \hat{\beta}\mathbf{X}) - n + 2p.$$

(→ p includes β_0)

In practice, σ^2 is often replaced by its estimate from the largest model considered.

Akaike information criterion (AIC) generalizes this to any likelihood:

$$AIC = -2 \times \log f(\hat{\beta}, \hat{\sigma}^2) + 2p,$$

which we wish to minimize also.

Where $f(\beta, \sigma^2) = (2\pi\sigma^2)^{-n/2} \exp\left(-\frac{1}{2\sigma^2}(Y - X\beta)'(Y - X\beta)\right)$ aka Gaussian likelihood

Bias-Variance Decomposition

Know: $Y = g(x) + \epsilon$ (g: unknown fcn), $\epsilon \sim N(0, \sigma^2)$

$\hat{Y} = \hat{g}(x_i) \perp\!\!\!\perp \epsilon$

Find population MSE.

$$\text{Pop MSE} = E\left[(Y - \hat{g}(x))^2 \mid X = x_i\right] = E\left[Y - g(x_i) + g(x_i) - \hat{g}(x_i)\right]^2$$

(squared diff of Y and $\hat{g}(x)$; for a given x_i; $- g(x_i) + g(x_i)$: "add zero")

$$= E(Y - g(x_i))^2 + 2E\left[(Y - g(x_i))(g(x_i) - \hat{g}(x_i))\right] + E\left[g(x_i) - \hat{g}(x_i)\right]^2$$

($E(Y - g(x_i))^2 = \sigma^2$; $2E[(Y - g(x_i))(g(x_i) - \hat{g}(x_i))] = 2E[\varepsilon_i (g(x_i) - \hat{g}(x_i))]$)

$$= \sigma^2 + 2E\,\varepsilon_i\, E[g(x_i) - \hat{g}(x_i)] + E[g(x_i) - \hat{g}(x_i)]^2$$

($2E\,\varepsilon_i = 0$)

$$\text{MSE} = \sigma^2 + E[g(x_i) - \hat{g}(x_i)]^2$$

(σ^2: unknown; $g(x_i)$: true mean; $\hat{g}(x_i)$: estimate)

$$E[g(x_i) - \hat{g}(x_i)]^2 = E[g(x_i)^2] - 2E\,g(x_i)\hat{g}(x_i) + E[\hat{g}(x_i)^2]$$

($E[g(x_i)^2] = g(x_i)^2$; $2E\,g(x_i)\hat{g}(x_i) = 2g(x_i)E\,\hat{g}(x_i)$)

$$= g(x_i)^2 - 2g(x_i)E\,\hat{g}(x_i) + E[\hat{g}(x_i)^2]$$

$$\text{Bias}^2 = [g(x_i) - E\hat{g}(x_i)]^2 = g(x_i)^2 - 2g(x_i)E\hat{g}(x_i) + E[(\hat{g}(x_i))]^2$$

$$\text{Variance} = E[\hat{g}(x_i)^2] - E[\hat{g}(x_i)]^2$$

$$\text{Bias}^2 + \text{Variance} = g(x_i)^2 - 2g(x_i)E\hat{g}(x_i) + \cancel{E[\hat{g}(x_i)]^2} + E[\hat{g}(x_i)^2] - \cancel{E[\hat{g}(x_i)]^2}$$

$$= g(x_i)^2 - 2g(x_i)E\hat{g}(x_i) + E[\hat{g}(x_i)^2]$$

$$= \text{MSE}$$