STAT 151A Lecture 28

Henry Liev

31 October 2025

Remark 0.1 (Abbreviated Derivation of Mallow's C_p)

$$\frac{\mathbb{E}[\hat{y}_m - \mathbf{X}\beta]^2}{\sigma^2} = \frac{\mathbb{E}\left[\hat{y}_m - H_m\mathbf{X}\beta\right]^2 + (\mathbb{H}_m\mathbf{X}\beta - \mathbf{X}\beta)^2}{\sigma^2}$$

$$\frac{\mathbb{E}\left[(\mathbb{H}_m y - \mathbf{X}\beta)^\intercal(\mathbb{H}_m y - \mathbf{X}\beta)\right]}{\sigma^2}$$

$$= \frac{\mathbb{E}\left[(\mathbb{H}_m y - \mathbb{H}_m\mathbf{X}\beta)^\intercal(\mathbb{H}_m y - \mathbb{H}_m\mathbf{X}\beta)\right] + (\mathbb{H}_m\mathbf{X}\beta - \mathbf{X}\beta)^\intercal(\mathbb{H}_m\mathbf{X}\beta - \mathbf{X}\beta)}{\sigma^2}$$

Use linearity and trace and Mallow's C_p is derived

Remark 0.2 (More on Cross-Validation)

k-fold cross-validation
Instead of leaving out a single point, you leave out a batch of points
2 advantages:

(1) Less computation

(2) Look at the broader set of perturbations in your data

$$CV_{LOO} = \frac{1}{n}\sum\left(\frac{e_i}{1-h_{ii}}\right)^2$$

LOO = Leave one out
Generalized CV mostly for computational reasons (will be ignored)

Remark 0.3 (Akaike Information Criterion (AIC))

AIC: Comes from Kullback-Leiber divergence K-L Divergence between probability distributions
Let f (full) and f_m (reduced) be densities for y
The K-L divergence: $\mathcal{J}(f, f_m) = \int \log\left(\frac{f(y)}{f_m(y)}\right) f(y)dy \geq 0$
Let's estimate $\mathcal{J}(f, f_m)$: density of $(y_1, \ldots, y_n)$
$f(y_1, \ldots, y_m) \overset{NLMiid}{=} \prod f(y_i)$
$\log f(y_1, \ldots, y_m) = \sum \log f(y_i)$
This strategy requires a probability model, but we can use the resulting score as a heuristic
Assuming NLM
$f_m(y_i) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(y_i - x_{i_m}^\intercal \beta)^2}{2\sigma^2}\right)$
$f(y_i) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(y_i - x_i^\intercal \beta)^2}{2\sigma^2}\right)$
Where we're going:
$AIC(m) = \hat{\mathcal{J}}(f, f_m) = -2\log \mathcal{L}(\hat{\beta}_m, \hat{\sigma}^2) + 2(p_m + 1) = n \log \frac{ErrSS_m}{n} + 2(p_m + 1) + C$
$\mathcal{L}(\beta_m) = \prod f_m(y_i|\beta_m) \overset{NLM}{\to} \prod \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(y - x_{i_m}^\intercal \beta_m)^2}{\sigma^2}\right)$
$\log \mathcal{L}(\beta_m) = -\frac{n}{2}\log 2\pi - \frac{n}{2}\log\sigma^2 - \frac{1}{2\sigma^2}\sum(y_i - x_{i_m}^\intercal \beta_m)^2$