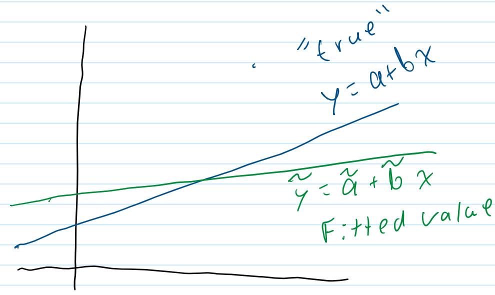
# 26. Variable Selection

November 7, 2025 9:29 AM

Last time: MSE and MSPE $\leftarrow$ $Var(\,) + Bias(\,)^2$

MSE $\uparrow$ $\tilde\beta$ parameters

MSPE $\uparrow$ $\tilde y$ Fitted values




$$MSPE = \frac{1}{\sigma^2}\sum_{i=1}^{n}\left(\tilde y_i - E y_i\right)^2$$

$$= \frac{1}{\sigma^2}\sum_{i=1}^{n}\left(\tilde y_i - E\tilde y_i + E\tilde y_i - E y_i\right)^2$$

$$= \frac{1}{\sigma^2}\sum_{i=1}^{n}\left(Var(\tilde y_i) + \boxed{Bias(\tilde y_i)^2}\right) \qquad \leftarrow Var(\tilde y) = \sigma^2 P_p$$

$$\sum_{i=1}^{n} Var(\tilde y_i) = \sigma^2\, tr(P_p) = \sigma^2\, ColRank(X_p) = \sigma^2(p+1)$$

sum diagonal of Matrix $P_p$ — inputs — intercept

obs — Fit

$$E\left(SS_{res}(p_1)\right) = E\left[\sum_{i=1}^{n}(y_i - \tilde y_i)^2\right]$$

$$= E\left[\sum_{i=1}^{n}\left(y_i - E\tilde y_i + E\tilde y_i - E y_i + E y_i - \tilde y_i\right)^2\right]$$

Noisy obs — expected Fitted value — expected true value — Fitted value

$$= E\left[\sum_{i=1}^{n}\left(\underbrace{(E\tilde y_i - E y_i)}_{Bias(\tilde y_i)} + \underbrace{(y_i - \tilde y_i)}_{\tilde r_i} - \underbrace{E(y_i - \tilde y_i)}_{E\tilde r_i}\right)^2\right]$$

$E\tilde r_i$ $\leftarrow$ $=0$ in Chpt 1 as unbiased. Now, it may not $=0$

$$= \sum_{i=1}^{n}\left[\boxed{Bias(\tilde y_i)^2} + \underbrace{E(\tilde r_i - E\tilde r_i)^2}_{Var(\tilde r_i)} + 0\right]$$

$$\sum_{i=1}^{n} Var(\tilde r_i) = \sigma^2\, tr(I - P_p) = \sigma^2(n - p_1 - 1)$$

$$\sum_{i=1}^{n} Bias(\tilde y_i)^2 = E\left[SS_{res}(p_1)\right] - \sigma^2(n - p_1 - 1)$$

$$MSPE = \frac{1}{\sigma^2}\sum_{i=1}^{n}\left(Bias(\tilde y_i)^2 + Var(\tilde y_i)\right)$$

$$= \frac{E\left[SS_{res}(p_1)\right]}{\sigma^2} - (n - p_1 - 1) + (p_1 + 1)$$

$$\frac{\quad}{\sigma^2}$$

$$= \frac{1}{\sigma^2}\mathbb{E}\left[SS_{res}(p_1)\right] - n + 2(p_1+1)$$

unknown, but they aren't $\beta$'s

Estimate $\sigma^2$ by $\frac{SS_{res}}{n-p-1}$

Mallows' $C_p$ Statistic:

reduced Model

Fixed

$$(n-p-1)\times\frac{SS_{res}(p_1)}{SS_{res}(p)} - n + 2(p_1+1) = C_p$$

Fixed

Full Model

$2\times(\#\text{ param!})$

The only terms that change / depend on the sub Model are $SS_{res}(p_1)$ and $p_1$

$\therefore$ We want to pick the submodel that Minimizes $C_p$ or Minimizes the MSPE

---

AIC = AKaike Information Criteria

- Mallows' $C_p$ above is specifically for linear Regression
- AIC is for any Model comparison problem

Coincide for linear regression under Normal errors $\varepsilon$

cost per parameter

$$AIC := -2\log(\text{likelihood}) + 2(\#\text{ parameters})$$

cost / penalty to using complex Models

- Max likelihood $\Rightarrow$ min $-\log$ Like
- Min AIC $\Rightarrow$ pick a model that Maximizes the likelihood But doesn't use too many parameters

General AIC

In Stat 378 with Normal errors

$$\boxed{AIC = n\log\left(\frac{SS_{res}}{n}\right) + 2(p+1)}$$

Minimize $SS_{res}$ — Not too many parameters

Step(…)

Stepwise Variable Selection (Greedy Search)

Stepwise Variable Selection (Greedy Search)

$p$ inputs $\Rightarrow$ $2^p$ Models to check!

- Forward Selection: Start with $y = \beta_0 + \varepsilon$ add add terms into it

- Backwards Selection: Start with $y = \beta_0 + \beta_1 x_1 + \dots + \beta_p x_p + \varepsilon$ and delete terms.

Point: Add or remove terms that best improve the AIC