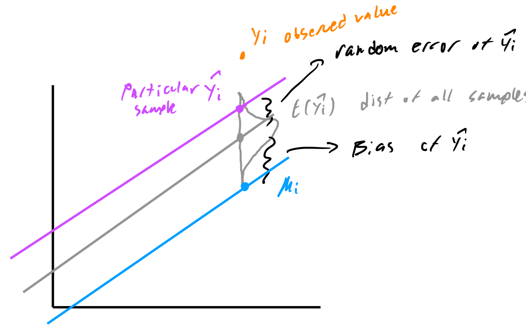
Mallows Cp



→ MSE of fitted values

Bias², random error, Variance

$$\to MSE_{\hat{y}_i} = (\hat{y}_i - \mu_i)^2 = [\,\underbrace{(E(\hat{y}_i) - \mu_i)}_{\text{Bias}} + \underbrace{(\hat{y}_i - E(\hat{y}_i))}_{\text{random error}}]^2 = \underbrace{(E(\hat{y}_i) - \mu_i)^2}_{\text{Bias}^2} + \underbrace{\sigma^2_{\hat{y}_i}}_{\text{variance}}$$

$\hat{y}_i$: fitted response

μ_i: True response function

$E(\hat{y}_i)$: Expectation of fitted value for given regression model

$$\Rightarrow \text{Total MSE} = \sum_{i=1}^{n} [(E(\hat{y}_i) - \mu_i)^2 + \sigma^2_{\hat{y}_i}]$$

$$= \sum_{i=1}^{n} (E(\hat{y}_i) - \mu_i)^2 + \sum_{i=1}^{n} \sigma^2_{\hat{y}_i}$$

$$\to \Gamma_p = \frac{\text{Total MSE}}{\text{True error variance}} = \frac{1}{\sigma^2}\left(\sum_{i=1}^{n} E(\hat{y}_i - \mu_i)^2 + \sum_{i=1}^{n} \sigma^2_{\hat{y}_i}\right)$$

can be shown:

$\to E(SSE_p) = \sum (E(\hat{y}_i) - \mu_i)^2 + (n-p)\sigma^2$

$\to \sum_{i=1}^{n} \sigma^2_{\hat{y}_i} = p\sigma^2$

rearrange

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

$= -n\sigma^2 + p\sigma^2 + p\sigma^2$

$= -n\sigma^2 + 2p\sigma^2$

$$= \frac{E(SSE_p)}{\sigma^2} + (2p - n)$$

$$\to \hat{\Gamma}_p = \frac{E(\widehat{SSE_p})}{\hat{\sigma}^2} + (2p-n) = \boxed{\frac{SSE_p}{MSE(X_1, \dots, X_{p-1})} + (2p-n) = C_p}$$

$\hat{\sigma}^2$: MSE estimates this accurately with the full model (unbiased)

SSE_p: SSE candidate model

$MSE(X_1, \dots, X_{p-1})$: MSE full model

$(2p-n)$: penalty term b/c want to minimize C_p

C_p ✓

→ Optimizing C_p

$$\to C_p = \frac{SSE_p}{MSE_{full}} + (2p - n)$$

$$= \frac{(n-p)\,MSE_p}{MSE_{full}} + (2p-n)$$

→ if smaller model doesn't inflate estimate of σ, then $MSE_p \approx MSE_{full}$

$$\approx \frac{(n-p)\,\cancel{MSE_p}}{\cancel{MSE_{full}}} + (2p-n)$$

$$\approx p$$

☆ Another way to think about it

$$\to C_p = \frac{1}{\sigma^2}\left(\sum_{i=1}^{n} \underbrace{(E(\hat{y}_i) - \mu_i)^2}_{} + \underbrace{\sum_{i=1}^{n} \sigma^2_{\hat{y}_i}}_{= p\sigma^2}\right)$$

★ when bias is small ⇒ ≈ 0

$$= \frac{1}{\sigma^2}\left(\sum (\approx 0)^2 + p\sigma^2\right)$$

$$\approx \frac{1}{\cancel{\sigma^2}}(p\,\cancel{\sigma^2})$$

$$\approx p$$

→ By definition → if $p = P - 1$ (all variables in model)

$$\to C_p = \frac{SSE_{full = p}}{MSE_{full = p}} + (2p - n)$$

$$= \frac{\cancel{SSE_{full}}}{\cancel{SSE_{full}}/(n-p)} + (2p-n)$$

$$= n - p + 2p - n$$

$$= p \text{ exactly}$$