Model Selection Criteria

Four ways to estimate test performance using an approximation

Full model has p predictors

RSS is the residual sum of squares for model with d predictors

$\hat{\sigma}^2 = \text{RSS}_p/(n-p-1)$ is an estimate of the error variance for full model

3. Bayesian Information Criterion (BIC)

For linear models: equivalent to Mallow's C_p (proportional to)

$$BIC = \frac{1}{n\hat{\sigma}^2}\left(\text{RSS} + \underbrace{\log(n)}_{\text{heavier penalty}}d\hat{\sigma}^2\right)$$

we are penalizing models of higher dimensionality (larger d, greater penalty)

$\Longrightarrow$ choose the model which has **minimum** BIC

Model Selection Criteria

Four ways to estimate test performance using an approximation

Full model has p predictors

RSS is the residual sum of squares for model with d predictors

$\hat{\sigma}^2 = \text{RSS}_p/(n-p-1)$ is an estimate of the error variance for full model

4. Adjusted R-squared value

Adjust the regular R^2 by taking into account number of predictors

$$\text{Adjusted-}R^2 = 1 - \frac{\text{RSS}/(n-d-1)}{\text{TSS}/(n-1)}$$

$\Longrightarrow$ choose the model which has **maximum** Adjusted-R^2