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

1. Mallow's C_p criterion:

For a given model with d (out of the p available) predictors

$$C_p = \frac{1}{n}\left(\text{RSS} + 2d\hat{\sigma}^2\right)$$

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

$\implies$ choose the model which has **minimum** C_p

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

2. Akaike Information Criterion (AIC)

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

$$AIC = \frac{1}{n\hat{\sigma}^2}\left(\text{RSS} + 2d\hat{\sigma}^2\right)$$

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

$\implies$ choose the model which has **minimum** AIC