1 December 2020

PRESS

$$\begin{aligned} \text{PRESS} &= \sum_{i=1}^{n}(y_i - \hat{y}_{(i)})^2 \\ &= \sum_{i=1}^{n}\frac{\hat{e}_i^2}{(1-h_{ii})^2} \end{aligned}$$

where h_{ii} is the $(i,i)^{th}$ of $\underset{\sim}{H} = \underset{\sim}{X}(\underset{\sim}{X}^T\underset{\sim}{X})^{-1}\underset{\sim}{X}^T$.

Properties of h_{ii}

1. $\sum_{i=1}^{n} h_{ii} = p'$
2. $\frac{1}{n} \leq h_{ii} \leq 1$

Summary on Best Subset Selection Methods

1. Choose the model with the smallest Cp statistic.
2. Choose the model with the smallest PRESS.
3. Choose the model with the largest R^2.
4. Choose the model with the largest R^2_{adj}, where $R^2_{adj} = 1 - \frac{MSE|_p}{TotalS.S./(n-1)}$.

   $\Rightarrow$ Choose the model with the smallest $\hat{\sigma}^2$.
5. Choose the model with the smallest AIC, where $\text{AIC} = n\log(\frac{\text{Res.S.S.}}{n}) + 2p'$.
6. Choose the model with the smallest BIC, where $\text{BIC} = n\log(\frac{\text{Res.S.S.}}{n}) + p'\log n$.

Residual Analysis

The mean and variance of $\hat{\boldsymbol{e}}$ are

$$\begin{aligned} \text{E}(\hat{\boldsymbol{e}}) &= \mathbf{0} \\ \text{Var}(\hat{\boldsymbol{e}}) &= \sigma^2(\boldsymbol{I} - \boldsymbol{H}) \end{aligned}$$

where $\boldsymbol{H} = \boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T$.

$\Rightarrow \text{E}(\hat{e}_i) = 0$, $\text{var}(\hat{e}_i) = \sigma^2(1-h_{ii})$

Studentized residual

1. The first uses $\hat{\sigma}^2$ to estimate σ^2, giving the formula

   $$r_i = \frac{\hat{e}_i}{\hat{\sigma}\sqrt{1-h_{ii}}}$$

   The r_i are called *internally Studentized residuals* because the estimate of σ^2 uses all of the data including the ith case.

2. The second scaling uses an estimate of σ^2 obtained when the ith case is excluded from the regression, *externally Studentized residual*, i.e.

   $$t_i = \frac{\hat{e}_i}{\hat{\sigma}_{-i}\sqrt{1-h_{ii}}}\,.$$

Model Diagnostic

1. Use Residual Plot(s) ($\hat{e}_i$ = RESIDUAL; r_i = STUDENT; t_i = RSTUDENT vs $\hat{y}_i$, x_{i1}, ..., x_{ik}) to check zero mean, homoscedasticity and linearity.

   no pattern $\Rightarrow$ assumptions on zero mean, homoscedasticity and linearity are valid.

2. Use QQ Plot to check normality.

   straight line $\Rightarrow$ assumption on normality is valid.

Transformation

Box-Cox Transformation - transformation on y

$$z_i^{\lambda} = \begin{cases} \dfrac{y_i^{\lambda} - 1}{\lambda[GM(y)]^{\lambda-1}} & \lambda \neq 0 \\ GM(y)\log(y_i) & \lambda = 0 \end{cases}$$

Find λ such that $L(\lambda)$ is maximized or $RSS_{\lambda}(\boldsymbol{Z})$ is minimized.