Lecture 20 Nov 17

Model Building / Variable Selection

- PRIOR INFORMATION

- Data-driven techniques

$$\left\{\begin{array}{l}\text{Forward} \\ \text{Backward} \\ \text{FB Stepwise}\end{array}\right. \quad \text{Algorithms}$$

If there are K possible predictors: K steps

All possible models : 2^K

Criterion-driven

AIC Akaike information Criterion

BIC Bayesian information Criterion

Let M be the model :

$$Y_i = \mu(x_{1i}, x_{2i}, \dots x_{pi}) + \varepsilon_i$$

$$\mu(\underline{x}_i) = \beta_0 + \beta_1 x_{1i} + \dots + \beta_p x_{pi}$$

where $\varepsilon_i \overset{iid}{\sim} N(0, \sigma^2)$

The maximum likelihood estimate of σ^2 computed from the data $\{(x_i, Y_i), i = 1 \ldots, n\}$ is

$$\hat{\sigma}^2_{MLE} = \frac{\|\underline{R}\|^2}{\underset{\nwarrow\ n-(p+1)}{n}} \quad \text{where} \quad \underline{R} = \begin{pmatrix} R_1 \\ \vdots \\ R_n \end{pmatrix}, \quad R_i = Y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_{1i} + \ldots + \hat{\beta}_p x_{pi})$$

The AIC & BIC values for this model are, respectively:

$$AIC(x_1 \ldots x_p) = \log \hat{\sigma}^2_{MLE(p)} + \frac{n+(p+1)}{n-(p+1)-2},$$

$$BIC(x_1 \ldots x_p) = \underbrace{\log \hat{\sigma}^2_{MLE(p)}}_{\text{FIT}} + \underbrace{\frac{(p+1)\log n}{n}}_{\text{PENALTY}},$$

As $p \uparrow$ $\qquad \underline{\downarrow} \qquad \underline{\uparrow}$

Squared Prediction error criteria

$$Y_i = \mu(\underline{x}_i) + \varepsilon_i$$

$$\mu(\underline{x}_i) = \beta_0 + \beta_1 x_{1i} + \ldots + \beta_p x_{pi}$$

$$\varepsilon_i \overset{iid}{\sim} N(0, \sigma^2)$$

Cross-Validation

DATA

Training

$\{(\underline{x}_1, Y_1), (\underline{x}_3, Y_3), \dots\}$

$\Downarrow$ Fit the model

$$Y_i = \beta_0 + \beta_1 x_{1i} + \dots + \beta_p x_{pi} + \varepsilon_i$$

$i = 1, 3, 5, \dots$

$$\Rightarrow \hat{\underline{\beta}}_{\text{TRAINING}} = \begin{pmatrix} \hat{\beta}_{0,TR} \\ \vdots \\ \hat{\beta}_{p,TR} \end{pmatrix}$$

Testing

$\{(\underline{x}_2, Y_2), (\underline{x}_4, Y_4), \dots\}$

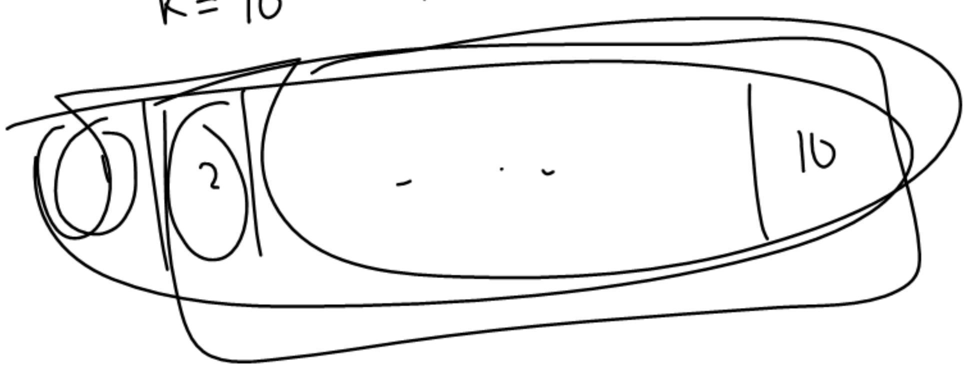
$$PE(i) = Y_i - (\hat{\beta}_{0,TR} + \hat{\beta}_{1,TR} x_{1i} + \dots + \hat{\beta}_{p,TR} x_{pi})$$

$i = 2, 4, 6, \dots$

$$SPE = \sum_{i=2,4,\dots} (PE(i))^2$$

"K-fold cross-validation"

$K = 10$



$$DATA = \bigcup_{g=1}^{10} Data_g$$

For $g=1:10$

$\quad Testing = Data_g$

$\quad Training = Data \setminus Data_g$

$\quad SPE_g$

END

$$SPE = \sum_{g=1}^{10} SPE_g$$

Multicollinearity $\longrightarrow$

$$Y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \varepsilon_i$$

$$\underline{Y} = X\underline{\beta} + \underline{\varepsilon}, \quad \underline{\varepsilon} \sim N(\underline{0}, I \otimes \sigma^2)$$

$$X = \begin{pmatrix} \mathbb{1} & x_{11} & x_{21} \\ & \vdots & \vdots \\ & x_{1n} & x_{2n} \end{pmatrix}_{n \times 3}$$

$$\hat{\underline{\beta}} = (X'X)^{-1} X' \underline{Y}$$

$$\operatorname{cov}(\hat{\underline{\beta}}) = (X'X)^{-1} \otimes \sigma^2 := V_{\hat{\beta}} \quad (3 \times 3)$$

$$\operatorname{Var}(\hat{\beta}_1) = V_{\hat{\beta}}(2,2)$$

$$= \left(\frac{1}{1 - r_{12}^2}\right) S_{x_1 x_2} \quad \sigma^2$$

where $$S_{x_1 x_2} = \sum_{i=1}^{n} (x_{1i} - \bar{x}_1)(x_{2i} - \bar{x}_2)$$

$$r_{12} = \frac{\sum (x_{1i} - \bar{x}_1)(x_{2i} - \bar{x}_2)}{\sqrt{\sum (x_{1i} - \bar{x}_1)^2 \cdot \sum (x_{2i} - \bar{x}_2)^2}}$$