Elements of Machine Learning

Assigment 2 - Problem 5

Sangeet Sagar(7009050), Philipp Schuhmacher(7010127)
{sasa00001,phsc00003}@stud.uni-saarland.de

December 2, 2021

Problem 5 (T, 8 Points)

Prove that for linear and polynomial least squares regression, the LOOCV estimate for the test MSE can be calculated using the following formula

$$\mathrm{CV}_{(n)} = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{y_i - \hat{y}_i}{1 - h_i}\right)^2$$

where h_i is the leverage. Note, that the leverage h_i for point i is given by the diagonal element pertaining to data point i of the hat matrix H.

Hint: First, properly understand and define all variables for the case with sample i left-out.

As we know: $\hat{y} = X\hat{\beta} = X(X^TX)^{-1}X^Ty$

i.e. $\hat{\beta} = \cancel{x}(X^TX)^{-1}X^Ty$

Now i^{th} observation is left out:-

$$\hat{\beta}_{-i} = (X_{-i}^T X_{-i})^{-1} X_{-i}^T Y_{-i}$$

$$CV_{(n)} = \frac{1}{n}\sum_{i=1}^{n} MSE_i = \frac{1}{n}\sum_{i=1}^{n}(y_i - x_i^T\hat{\beta}_{-i})^2$$

where $\hat{\beta}_{-i}$ = estimated Coefficent after i^{th} observation left out.

we need to Compute $\hat{\beta}_{-i}$ using $\hat{\beta}_{-i} = (X_{-i}^T X_{-i})^{-1} X_{-i}^T Y_{-i}$

using Sherman Morrison woodbury formula:-

$$(X^TX)^{-1} = (X^TX - x_i x_i^T)^{-1} = (X^TX)^{-1} + \frac{(X^TX)^{-1}x_i x_i^T (X^TX)^{-1}}{1 - x_i^T(X^TX)^{-1}x_i}$$

where:- $(X^TX)_{-i} :\Rightarrow [X^TX - x_i x_i^T]$

i.e. x_i^{th} row left out.

Also:- $X_{-i}^T Y_{-i} = X^Ty - x_i y_i$

$$\Rightarrow (X_{-i}^T X_{-i})^{-1} X_{-i}^T Y_{-i} = \underbrace{(X_{-i}^T X_{-i})^{-1}}_{\hat{\beta}_{-i}} (X^Ty - x_i y_i)$$

$$\Rightarrow \hat{\beta}_{-i} = \left[(X^TX)^{-1} + \frac{(X^TX)^{-1}x_i x_i^T(X^TX)^{-1}}{1 - x_i^T(X^TX)^{-1}x_i}\right](X^Ty - x_i y_i)$$

$$= (X^TX)^{-1}(X^Ty - x_i y_i) + \frac{(X^TX)^{-1} x_i x_i^T (X^TX)^{-1}}{1 - x_i^T (X^TX)^{-1} x_i} (X^Ty - x_i y_i)$$

$$= \underbrace{(X^TX)^{-1}X^Ty}_{\hat{\beta}} - (X^TX)^{-1}x_i y_i + \frac{(X^TX)^{-1}x_i x_i^T (X^TX)^{-1}X^Ty - x_i y_i (X^TX) x_i x_i^T (X^TX)^{-1}}{1 - x_i^T (X^TX)^{-1} x_i}$$

$$= \hat{\beta} - \frac{\left[(X^TX)^{-1}x_i y_i - (X^TX)^{-1}x_i y_i \cancel{x_i^T (X^TX)^{-1} x_i} - (X^TX)^{-1}x_i x_i^T (X^TX^{-1})X^Ty + x_i y_i (X^TX)^{-1} \cancel{x_i} x_i^T (X^TX)^{-1}\right]}{1 - x_i^T (X^TX)^{-1} x_i}$$

$$= \hat{\beta} - \left[\frac{(X^TX)^{-1}x_i y_i - (X^TX)^{-1}x_i x_i^T \underbrace{(X^TX)^{-1}X^Ty}_{\beta}}{1 - x_i^T (X^TX)^{-1} x_i}\right]$$

$$= \hat{\beta} - \left[\frac{(X^TX)^{-1}x_i y_i - (X^TX)^{-1}x_i x_i^T \hat{\beta}}{1 - x_i^T (X^TX)^{-1} x_i}\right]$$

$$\boxed{\hat{\beta}_{-i} = \hat{\beta} - \left[\frac{(X^TX)^{-1}x_i (y_i - x_i^T \hat{\beta})}{1 - h_{ii}}\right]}$$

$h_{ii} \rightarrow \text{diag } X(X^TX)^{-1}X$

~~MSE~~

$$CV_{(n)} = \frac{1}{n} \sum_{i=1}^{n} \left(y_i - x_i^T \hat{\beta}_{-i}\right)^2$$

$$= \frac{1}{n} \sum_{i=1}^{n} \left(y_i - x_i^T \left[\hat{\beta} - \frac{(X^T X)^{-1} x_i \, (y_i - x_i^T \hat{\beta})}{1 - h_{ii}} \right] \right)^2$$

$$= \frac{1}{n} \sum_{i=1}^{n} \left(y_i - x_i^T \hat{\beta} + \frac{x_i^T (X^T X)^{-1} x_i \, (y_i - x_i^T \hat{\beta})}{1 - h_{ii}} \right)^2$$

$$= \frac{1}{n} \sum_{i=1}^{n} \left(y_i - x_i^T \hat{\beta} + \frac{h_{ii} \, (y_i - x_i^T \hat{\beta})}{1 - h_{ii}} \right)^2$$

$$= \frac{1}{n} \sum_{i=1}^{n} \left(\frac{y_i - \cancel{y_i h_{ii}} - x_i^T \hat{\beta} + \cancel{x_i^T \hat{\beta} h_{ii}} + \cancel{h_{ii} y_i} - \cancel{h_{ii} x_i^T \hat{\beta}}}{1 - h_{ii}} \right)^2$$

$$= \frac{1}{n} \sum_{i=1}^{n} \left(\frac{y_i - x_i^T \hat{\beta}}{1 - h_{ii}} \right)^2$$

$$\boxed{CV_{(n)} = \frac{1}{n} \sum_{i=1}^{n} \left(\frac{y_i - \hat{y}_i}{1 - h_{ii}} \right)^2}$$

These resources were throughly studied to understand the logic behind the error estimates of LOOCV [Meijer] and [Hyndman]

References

Rob J Hyndman. Fast computation of cross-validation in linear models. `https://robjhyndman.com/hyndsight/loocv-linear-models/`. Accessed: 2021-12-02.

Rosa Meijer. Efficient approximate leave-one-out cross-validation for ridge and lasso. `https://repository.tudelft.nl/islandora/object/uuid:d9b5456d-722a-401d-9f1a-c530c46d6491/datastream/OBJ/download`. Accessed: 2021-12-02.