EML Assignment 3 Problem 1

Muhammed Saeed and Ali Salaheldin Ali Ahmed

December 2022

Question 1

In k-fold cross-validation, the original sample is randomly partitioned into k equal sized subsamples. Of the k subsamples, a single subsample is retained as the validation data for testing the model, and the remaining k-1 subsamples are used as training data. The cross-validation process is then repeated k times, with each of the k subsamples used exactly once as the validation data. The k results from the folds can then be averaged to produce a single estimation.

The value of k in k-fold cross-validation determines the number of subsamples that the original sample is divided into, and therefore has a direct impact on the number of times the model is trained and evaluated. A larger value of k provides a more accurate estimate of the model's performance, but also requires more computational resources.

K-fold cross-validation is a compromise between the validation set approach and leave-one-out cross-validation (LOOVC). In the validation set approach, a single fixed subset of the original sample is used as the validation data, which may not provide an accurate estimate of the model's performance. In LOOVC, the model is trained and evaluated k times, with each of the k subsamples used as the validation data exactly once. This provides a very accurate estimate of the model's performance, but is computationally expensive.

K-fold cross-validation provides a more efficient estimate of the model's performance compared to LOOVC, while still providing a better estimate than the validation set approach. Additionally, k-fold cross-validation ensures that all observations are used for both training and validation, which can help to reduce bias in the model.

Question 2

The hat matrix H is the matrix which maps the original observations' responses Y to their predicted or fitted values, $\hat{Y}$.

$$\hat{Y} = HY$$

The diagonal elements h_i, are known as the leverage values for each sample, i. h_i represents the contribution of sample x_i (based on y_i) in the its own fitted value $\hat{y}_i$.

Samples with high leverage values have a greater influence on the model's predicted values compared to samples with low leverage values. Removing a sample with high leverage from the dataset can have a significant effect on the model estimation, as the model's predicted values will be based on a different set of influential samples. This can lead to a change in the model's coefficients and overall performance.

Therefore, it is important to identify samples with high leverage and assess their potential impact on the model estimation. In some cases, it may be necessary to remove such samples from the dataset in order to improve the model's performance.

Question 3

$$\beta^{-t} = (u^{t1} - ht)(X'X)^{-1}X't, \qquad (A)$$

where β is the estimate using all data and $\beta^{(t)}$ the estimate when leaving out $X(t)$, observation t. Let Xt be defined as a row vector such that $y^t = Xt\beta$. u^t are the residuals.

The proof uses the following matrix algebraic result.

$$\begin{aligned} Let A &= a\text{nonsingular matrix,} \\ b &= \text{a vector,} \\ \lambda &= \text{a scalar.} \\ If \lambda &\neq -1b'A^{-1}b \\ Then (A + \lambda bb')^{-1} &= A^{-1} - (\lambda 1 + \lambda b'A^{-1}b)A^{-1}bb'A^{-1} \qquad (B) \end{aligned}$$

We first show that $\beta^{-t} = (u^{t1} - ht)(X'X)^{-1}X't$, (A) where β is the estimate using all data and $\beta^{(t)}$ the estimate when leaving out $X(t)$, observation t. Let Xt be defined as a row vector such that $y^t = Xt\beta$. u^t are the residuals.

The proof uses the following matrix algebraic result.

Let A be a nonsingular matrix, b a vector and λ a scalar. If $\lambda \neq -1b'A^{-1}b$ Then $(A + \lambda bb')^{-1} = A^{-1} - (\lambda 1 + \lambda b'A^{-1}b)A^{-1}bb'A^{-1}$ (B) The proof of (B) follows immediately from verifying

$$\{A^{-1} - (\lambda 1 + \lambda b'A^{-1}b)A^{-1}bb'A^{-1}\}(A + \lambda bb') = I.$$

The following result is helpful to prove (A)

$$(X'(t)X(t))^{-1}X't = (1 - ht)(X'X)^{-1}X't. \qquad (C)$$

Proof of (C): By (B) we have, using $\sum_{t=1}^{T} X'tXt = X'X$,

$$\begin{aligned}(X'(t)X(t))^{-1} &= (X'X - X'tXt)^{-1} \\ &= (X'X)^{-1} + (X'X)^{-1}X'tXt(X'X)^{-1} - Xt(X'X)^{-1}X't.\end{aligned}$$

So we find

$$\begin{aligned}(X'(t)X(t))^{-1}X't &= (X'X)^{-1}X't + (X'X)^{-1}X't(Xt(X'X)^{-1}X't - Xt(X'X)^{-1}X't) \\ &= (1 - ht)(X'X)^{-1}X't.\end{aligned}$$

The proof of (A) now follows from (C): As $X'X\beta = X'y$, we have

$$\begin{aligned}(X'(t)X(t) + X'tXt)\beta &= X'(t)y(t) + X'tyt, \\ \{I_k + (X'(t)X(t))^{-1}X'tXt\}\beta &= \beta^{(t)} + (X'(t)X(t))^{-1}X't(Xt\beta + u^t).\end{aligned}$$

So,

$$\begin{aligned}\beta &= \beta^{(t)} + (X'(t)X(t))^{-1}X'tu^t \\ &= \beta^{(t)} + (X'X)^{-1}X't.\end{aligned}$$

where the last equality follows from (C). (1)

Now, note $ht = Xt(X'X)^{-1}X't$. (2)

Multiply through in (A) by $Xt,$ add yt on both sides and rearrange to get, (3)

with $u^{(t)}$ the residuals resulting from using $\beta^{(t)}(yt - Xt\beta^{(t)})$, (4)

$u^{(t)} = u^t + (u^t 1 - ht)ht$ (5)

or (6)

$u^{(t)} = u^t(1 - ht) + u^t ht 1 - ht = u^t 1 - ht$ (7)

Note the solution is reference is link