Advanced Data Analysis
Homework Week 2

Aswin Vijay

April 24, 2023

Homework 2

The Mean Squared Error for Leave one out Cross Validation is given by,

$$MSE = \frac{1}{n}\sum_{i=1}^{n}\left(\hat{f}_i(x_i) - y_i\right)^2$$

where $\hat{f}_i(x)$ is learned from data other than (x_i, y_i). The given linear model is,

$$f_\theta(x) = \sum_{j=1}^{b}\theta_j\phi(x)_j = \phi^T\theta$$

Calculating optimal $\hat{\theta}_i$ learned by removing (x_i, y_i) we get the optimization objective as,

$$\hat{\theta}_i = \arg\min_\theta\left[\frac{1}{2}\left(\sum_{k=1}^{n}(f_\theta(x_k) - y_k)^2 - (f_\theta(x_i) - y_i)^2\right) + \frac{\lambda}{2}||\theta||^2\right]$$

Using $\Phi, \mathbf{y}, \theta, \phi_i, y_i$ we get the regularized objective as,

$$\hat{\theta}_i = \arg\min_\theta\left[\frac{1}{2}||\Phi\theta - \mathbf{y}||^2 - \frac{1}{2}||\phi_i\theta - y_i||^2 + \frac{\lambda}{2}||\theta||^2\right]$$

Taking the derivative and setting to zero we get,

$$\nabla_\theta\left[\frac{1}{2}||\Phi\theta - \mathbf{y}||^2 - \frac{1}{2}||\phi_i\theta - y_i||^2 + \frac{\lambda}{2}||\theta||^2\right] = 0$$

$$\Phi^T\Phi\theta - \Phi^T\mathbf{y} - \phi_i^T\phi_i\theta + \phi_i y_i + \lambda\theta = 0$$

$$\hat{\theta}_i = \frac{\Phi^T\mathbf{y} - \phi_i y_i}{\Phi^T\Phi \ - \phi_i^T\phi_i + \lambda I}$$

Where, Setting $U = \Phi^T\Phi \ + \lambda I$ we get,

$$\hat{\theta}_i = \frac{\Phi^T \mathbf{y} - \phi_i y_i}{U - \phi_i^T \phi_i}$$

Calculating the MSE using prediction for y_i as $\phi_i^T \hat{\theta}_i$ and summing the squared error we get,

$$\begin{aligned}
MSE &= \frac{1}{n}\sum_{i=1}^{n}\left(\phi_i^T \hat{\theta}_i - y_i\right)^2 \\
&= \frac{1}{n}\sum_{i=1}^{n}\left(\phi_i^T \left[\frac{\Phi^T \mathbf{y} - \phi_i y_i}{U - \phi_i^T \phi_i}\right] - y_i\right)^2 \\
&= \frac{1}{n}\sum_{i=1}^{n}\left[(U - \phi_i^T \phi_i)^{-1}(\phi_i^T \Phi^T \mathbf{y} - \phi_i^T \phi_i y_i) - y_i\right]^2
\end{aligned}$$

Using the special case of Sherman-Morrison-Woodbury below we have,

$$\begin{aligned}
MSE &= \frac{1}{n}\sum_{i=1}^{n}\left[(U - \phi_i^T \phi_i)^{-1}(\phi_i^T \Phi^T \mathbf{y} - \phi_i^T \phi_i y_i) - y_i\right]^2 \\
&= \frac{1}{n}\sum_{i=1}^{n}\left[(U - \phi_i^T \phi_i)^{-1}(\phi_i^T \Phi^T \mathbf{y} - U y_i)\right]^2 \\
&= \frac{1}{n}||(U - \Phi^T \Phi)^{-1}(\Phi^T \Phi^T \mathbf{y} - U\mathbf{y})||^2 \\
&= \frac{1}{n}||\left([U^{-1} + \frac{U^{-1}\Phi^T \Phi U^{-1}}{I - \Phi U^{-1}\Phi^T}][\Phi^T \Phi^T \mathbf{y} - U\mathbf{y}]\right)||^2 \quad Sherman - Morrison - Woodbury \\
&= \frac{1}{n}||\left(U^{-1}\Phi^T \Phi^T \mathbf{y} - \mathbf{y} + \frac{U^{-1}\Phi^T \Phi U^{-1}\Phi^T \Phi^T \mathbf{y} - U^{-1}\Phi^T \Phi \mathbf{y}}{I - \Phi U^{-1}\Phi^T}\right)||^2
\end{aligned}$$

Next we expand out the denominators and using $H = I - \Phi U^{-1}\Phi^T$ we get the result,

$$MSE = \frac{1}{n}\left\|\frac{H\mathbf{y}}{H}\right\|$$