CSCI 145 Problem 6 Solution: Leave One Out Without Fitting Again

Solution

1. The Leave-One-Out Residual

Fix an observation i. Let $\mathbf{w}_{-i}$ be the least-squares weights fitted without that observation, and let t be their prediction on it. We will put observation i back with label t instead of its original label y_i. Adding a perfectly predicted observation cannot change the optimal weights.

To see this, use sums of squared errors, which have the same minimizers as their averages. For the modified dataset, any weights $\mathbf{w}$ have loss

$$\sum_{j \neq i} \big(y_j - [\mathbf{X}\mathbf{w}]_j\big)^2 + \big(t - [\mathbf{X}\mathbf{w}]_i\big)^2.$$

The first term is minimized by $\mathbf{w}_{-i}$, and the second is nonnegative and equals zero at $\mathbf{w}_{-i}$. Thus $\mathbf{w}_{-i}$ minimizes their sum. The independent feature columns make this minimizer unique, so fitting the modified dataset still predicts t at observation i.

The modified label vector is $\mathbf{y} + (t - y_i)\mathbf{e}_i$, where $\mathbf{e}_i$ has a 1 in entry i and zeros elsewhere. Apply the supplied formula to this label vector:

$$\begin{aligned} \text{modified fitted predictions} &= \mathbf{H}\big(\mathbf{y} + (t - y_i)\mathbf{e}_i\big) \\ &= \mathbf{y}^\star + (t - y_i)\mathbf{H}\mathbf{e}_i. \end{aligned}$$

Its ith entry must be t, and the ith entry of $\mathbf{H}\mathbf{e}_i$ is H_{ii}. Therefore

$$t = y_i^\star + H_{ii}(t - y_i).$$

Subtract this identity from y_i and collect the leave-one-out residual:

$$\begin{aligned} y_i - t &= y_i - y_i^\star + H_{ii}(y_i - t), \\ (1 - H_{ii})(y_i - t) &= y_i - y_i^\star. \end{aligned}$$

Since $H_{ii} < 1$, we can divide to obtain

$$y_i - t = \frac{y_i - y_i^\star}{1 - H_{ii}}.$$

The same argument applies to every observation. All leave-one-out residuals therefore follow from the full-data residuals and the diagonal of $\mathbf{H}$, without fitting another model. The factor $1/(1 - H_{ii})$ shows how a small training residual can grow when that observation is omitted.