Can We Leave One Out Without Fitting Again?

Due: Wednesday, September 30 at 11:59pm

Instructions: Please submit your *handwritten* solutions as a single PDF on Gradescope.

To evaluate a prediction rule, we would like to test it on observations whose labels it did not use for training. In *leave-one-out evaluation*, we remove each observation in turn, fit least squares to the remaining observations, and predict the omitted label. This appears to require fitting n separate models.

Let $\mathbf{X} \in \mathbb{R}^{n \times d}$ be the design matrix and $\mathbf{y} \in \mathbb{R}^n$ the labels. Assume the columns of $\mathbf{X}$ remain linearly independent after deleting any row. The full-data fit is

$$\mathbf{y}^\star = \mathbf{H}\mathbf{y}, \qquad \mathbf{H} = \mathbf{X}(\mathbf{X}^\top \mathbf{X})^{-1}\mathbf{X}^\top.$$

Fix an observation i. Let $\mathbf{w}_{-i}$ be the least-squares fit with observation i omitted, and let $t = [\mathbf{X}\mathbf{w}_{-i}]_i$. Define $\mathbf{z} = \mathbf{y} + (t - y_i)\mathbf{e}_i$, where $\mathbf{e}_i$ has a one in coordinate i and zeros elsewhere.

1. Replacing the omitted label by t adds a perfectly predicted observation. Prove that $\mathbf{w}_{-i}$ is also the least-squares fit to all rows of $\mathbf{X}$ with label vector $\mathbf{z}$.
2. The full-data prediction map is linear in the labels. Prove that $t = y_i^\star + H_{ii}(t - y_i)$.
3. The error on the omitted observation is $y_i - t$. Express this residual using only y_i, $y_i^\star$, and H_{ii}.

Hint: For any label vector $\mathbf{z} \in \mathbb{R}^n$, the fitted predictions on the same design matrix are $\mathbf{H}\mathbf{z}$. Under the stated independence assumption, $0 \le H_{ii} < 1$ for every i.

Takeaway: *For least squares, we can compute all the leave-one-out errors from one fit. This gives the same evaluation as fitting a separate model for each omitted point, with much less repeated work.*