CSCI 145 Problem 8: How Much Does Training Error Hide?

October 27, 2026

Submission Instructions

Please upload your work to Gradescope by **11:59pm on Wednesday, October 7**.

- Turn in a single PDF of *handwritten* solutions. Typed (including LaTeX) solutions will not be accepted.
- A complete, good-faith attempt earns submission credit. The grader may select any recent assigned problem for a brief conversation, so keep your work and be prepared to explain each solution and answer a related question. Scheduling details will be announced in class.

Problem 8: How Much Does Training Error Hide?

Several sensors measure the same constant quantity μ, but their readings contain noise. We can give each sensor its own prediction or pool their readings into a shared prediction.

There are $n \geq 2$ sensors. Sensor i contributes a reading Y_i to the training set and a reading Y_i' to the test set:

$$Y_i = \mu + \varepsilon_i, \qquad Y_i' = \mu + \varepsilon_i'.$$

All $2n$ errors are independent, with mean zero and variance $\sigma^2 > 0$. Before observing any readings, partition the sensors into k equal groups, where k divides n. Write $m = n/k$ and let $G(i)$ be sensor i's group. Predict its test reading using its group's initial average:

$$\widehat{Y}_i = \frac{1}{m} \sum_{j \in G(i)} Y_j.$$

Define the training-set and test-set mean squared errors by

$$\mathcal{L}_{\text{train}} = \frac{1}{n} \sum_{i=1}^{n} (Y_i - \widehat{Y}_i)^2, \qquad \mathcal{L}_{\text{test}} = \frac{1}{n} \sum_{i=1}^{n} (Y_i' - \widehat{Y}_i)^2.$$

1. What are the expectations of these two errors and their difference as functions of n, k, and σ^2?

2. Does choosing k by training error identify the most accurate predictions on the test set?

Takeaway: *More separate predictions lower training error but raise expected test error in this example. Choosing how many predictions to fit using training error alone rewards fitting the noise.*