Name:

Short Quiz 2A

22 September 2025

Question 1: What is a validation dataset? From validation data $(x_1, y_1), \ldots, (x_V, y_V)$, how do we estimate the performance of a predictor f for a cost c?

Answer: A validation dataset is used to select the best learning procedure among several ones. The performance of a predictor f for a cost c is estimated with

$$\frac{1}{V} \sum_{v=1}^{V} c(y_v, f(x_v)).$$

Question 2: What is overfitting for a predictor?

Answer: A predictor overfits if it has a small train error and a large test error: it has learnt "by heart" the training data, but generalizes poorly.

Question 3: True or false? Circle the right answer.

T   F   The estimation error increases with the size of the model $\mathcal{F}$.

T   F   Complex models tend to overfit more.

T   F   Cross-validation is useful when the quantity of data is scarce.

Answer: True — True — True.