Some extra problems for COMS 4771 Fall 2025

Daniel Hsu

September 24, 2025

1 Problems

In this part of the assignment, you'll work out why $f_{\text{avg}} := \frac{1}{M}\sum_{t=1}^{M} f_t$ satisfies

$$\mathbb{E}[(f_{\text{avg}}(\vec{X}) - Y)^2] = \frac{1}{M}\sum_{t=1}^{M}\mathbb{E}[(f_t(\vec{X}) - Y)^2] - \frac{1}{2M^2}\sum_{s=1}^{M}\sum_{t=1}^{M}\mathbb{E}[(f_s(\vec{X}) - f_t(\vec{X}))^2] \tag{1}$$

for any random example $(\vec{X}, Y)$ and any real-valued functions $f_1, \dots, f_M$.

> **Problem 1.1 (1 point).** Suppose A and B are independent and identically distributed random variables, each with variance σ^2. Determine the relationship between $\mathbb{E}[(A - B)^2]$ and σ^2. Briefly (but precisely) explain your answer. *Hint: The bias-variance decomposition can be useful here. Or, just expand the square.*

It turns out the original claim in Equation (1) is true even if we remove the expectations and replace $(\vec{X}, Y)$ with an arbitrary (non-random) example $(\vec{x}, y)$:

$$(f_{\text{avg}}(\vec{x}) - y)^2 = \frac{1}{M}\sum_{t=1}^{M}(f_t(\vec{x}) - y)^2 - \frac{1}{2M^2}\sum_{s=1}^{M}\sum_{t=1}^{M}(f_s(\vec{x}) - f_t(\vec{x}))^2. \tag{2}$$

The original claim in Equation (1) follows from Equation (2) simply by replacing $(\vec{x}, y)$ with $(\vec{X}, Y)$ and taking expectations. So let's just focus on understanding why Equation (2) is true.

> **Problem 1.2 (2 points).** Rewrite Equation (2) using the relationship identified in Problem 1.1 (in particular to change the final term on the right-hand side), so that the rewritten equation can be interpreted as an instance of the bias-variance decomposition. Briefly (but precisely) explain the interpretation. *Hint: Define a random variable T whose distribution is uniform over $\{1, 2, \dots, M\}$, and consider the random variable $f_T(\vec{x})$.*

2 Solutions

Problem 1.1 Let us define μ to be the mean of A (which is also the mean of B). We first write

$$\mathbb{E}[(A - B)^2] = \mathbb{E}\Big[\mathbb{E}[(A - B)^2 \mid A]\Big] \tag{3}$$

using the tower property of conditional expectations. Now we apply the bias-variance decomposition to re-write the "inner" (conditional) expectation:

$$\mathbb{E}[(A-B)^2 \mid A] = \underbrace{(A-\mathbb{E}[B \mid A])^2}_{\text{squared bias}} + \underbrace{\mathbb{E}[(B-\mathbb{E}[B \mid A])^2 \mid A]}_{\text{variance}}$$

(Here, "squared bias" and "variance" are understood to be conditional on A.) But since A and B are independent, the right-hand side can be simplified as follows:

$$\begin{aligned}(A-\mathbb{E}[B \mid A])^2 + \mathbb{E}[(B-\mathbb{E}[B \mid A])^2 \mid A] &= (A-\mathbb{E}[B])^2 + \mathbb{E}[(B-\mathbb{E}[B])^2] \\ &= (A-\mu)^2 + \mathbb{E}[(B-\mu)^2] \\ &= (A-\mu)^2 + \sigma^2.\end{aligned}$$

Plugging this back into (3) and computing the "outer" expectation,

$$\begin{aligned}\mathbb{E}[(A-B)^2] &= \mathbb{E}[(A-\mu)^2 + \sigma^2] \\ &= \mathbb{E}[(A-\mu)^2] + \sigma^2 \quad \text{(by linearity of expectation)} \\ &= \sigma^2 + \sigma^2 = 2\sigma^2.\end{aligned}$$

We see that $\mathbb{E}[(A-B)^2]$ for iid random variables A and B is twice the variance of each random variable.

Problem 1.2 Our task is to explain how

$$(f_{\text{avg}}(\vec{x}) - y)^2 = \frac{1}{M}\sum_{t=1}^{M}(f_t(\vec{x}) - y)^2 - \frac{1}{2M^2}\sum_{s=1}^{M}\sum_{t=1}^{M}(f_s(\vec{x}) - f_t(\vec{x}))^2 \tag{4}$$

is a consequence of the bias-variance decomposition.

Let T be a random variable uniformly distributed in $\{1, 2, \ldots, M\}$, and let $A := f_T(\vec{x})$. Observe that $\mathbb{E}[A] = f_{\text{avg}}(\vec{x})$. The bias-variance decomposition implies

$$\mathbb{E}[(A-y)^2] = (\mathbb{E}[A] - y)^2 + \text{var}(A),$$

which we can re-arrange to

$$(\mathbb{E}[A] - y)^2 = \mathbb{E}[(A-y)^2] - \text{var}(A). \tag{5}$$

If we write out each expectation explicitly, (5) becomes

$$\underbrace{(f_{\text{avg}}(\vec{x}) - y)^2}_{(\mathbb{E}[A]-y)^2} = \underbrace{\frac{1}{M}\sum_{t=1}^{M}(f_t(\vec{x}) - y)^2}_{\mathbb{E}[(A-y)^2]} - \underbrace{\frac{1}{M}\sum_{t=1}^{M}(f_t(\vec{x}) - f_{\text{avg}}(\vec{x}))^2}_{\text{var}(A)}. \tag{6}$$

We claim that (6) is equivalent to (4). To see this, recall that by Problem 6,

$$\text{var}(A) = \frac{1}{2}\,\mathbb{E}[(A-B)^2], \tag{7}$$

where B is an "independent copy" of A (i.e., B has the same distribution as A and is independent of A). If we write out the expectation in (7), we obtain

$$\begin{aligned} \mathrm{var}(A) = \frac{1}{2}\,\mathbb{E}[(A-B)^2] &= \frac{1}{2}\sum_{s=1}^{M}\sum_{t=1}^{M}\Pr[(A,B)=(s,t)] \times (f_s(\vec{x}) - f_t(\vec{x}))^2 \\ &= \frac{1}{2M^2}\sum_{s=1}^{M}\sum_{t=1}^{M}(f_s(\vec{x}) - f_t(\vec{x}))^2. \end{aligned}$$

Plugging this back into (6) yields the equation (4) that we wanted to establish.