CS 70 Discussion 13A

November 27, 2024

Estimation

Problem: Let's say I have a random variable Y, and I want to predict the value of Y. What should my prediction for Y be?
Solution: Let's call our estimator $\hat{Y}$. First, how do we determine if a prediction is "good"? We will use the **Mean-Squared Error**:

$$\mathbb{E}\left[\left(\hat{Y} - Y\right)^2\right]$$

We can observe that $\hat{Y} = \mathbb{E}[Y]$ minimizes this error. This means we should just guess the expectation as the value of Y. We call this our **Minimum Mean-Squared Estimate (MMSE)**.

Estimation (Cont.)

Problem: What if I want to predict the value of random variable Y, but I am also given the value of another random variable X? What is the MMSE this time?
Solution: Since we are given more information to make a prediction, we will want to minimize the following error function this time:

$$\mathbb{E}\left[\left(\hat{Y} - Y\right)^2 | X\right]$$

What we find is that $\hat{Y} = \mathbb{E}[Y|X]$ minimizes our error!

Linear Estimation

Problem: What if I don't know $\mathbb{E}[Y|X]$? How do I create an estimator?
Solution: One option is to restrict ourselves to the **linear least-squares estimator (LLSE)** (meaning that our estimator $\hat{Y}$ is a linear function of X). We use the same error function as with the previous slide:

$$\mathbb{E}\left[\left(\hat{Y} - Y\right)^2 |X\right]$$

We find that $\hat{Y} = \mathbb{E}[Y] + \frac{\text{Cov}(X,Y)}{\text{Var}(X)}(X - \mathbb{E}[X])$ is the best linear estimator (i.e. it produces the lowest error among all linear estimators).
Note: We only need unconditioned expectations for the linear estimator formula.