Exercise: Bias and Variance

Bookmark this page

Exercise: Bias and Variance

5/5 points (graded)

Throughout this exercise, we continue off the example of estimating the probability of heads θ for a coin.

Note that there are different ways in which one can compute an estimate $\hat{\theta}$. In 6.008.1x, we mainly use maximum likelihood estimation, and as we'll see later, we also use MAP estimation.

- For the coin case, the maximum likelihood (ML) estimate $\hat{\theta} = \frac{n_{\text{heads}}}{n}$. Note that $\hat{\theta}$ is a function of the training data $X^{(1)}, \dots, X^{(n)}$. Before conditioning on a specific observed value of the training data, the ML estimate $\hat{\theta}$ is a random variable since $n_{\text{heads}} \sim \text{Binomial}(n, \theta)$; some times to make this explicit, we write $\hat{\theta}(X^{(1)}, \dots, X^{(n)})$, making the dependence on the training data clear. In particular,

$$\hat{\theta}(X^{(1)}, \dots, X^{(n)}) = \frac{n_{\text{heads}}}{n} = \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}\{X^{(i)} = \text{heads}\}.$$

What is $\mathbb{E}[\hat{\theta}(X^{(1)}, \dots, X^{(n)})]$ as a function of θ, the true unknown parameter?

In this part, please provide your answer as a mathematical formula (and not as Python code). Use ^ for exponentiation, e.g., x^2 denotes x^2. Explicitly include multiplication using *, e.g. x*y is xy.

You can use the variable theta.

theta

✔ **Answer:** theta

θ

Solution:

We have

$$\hat{\theta}(X^{(1)}, \dots, X^{(n)}) = \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}\{X^{(i)} = \text{heads}\},$$

so by linearity of expectation,

$$\mathbb{E}[\hat{\theta}(X^{(1)}, \dots, X^{(n)})] = \frac{1}{n}\sum_{i=1}^{n} \mathbb{E}[\mathbf{1}\{X^{(i)} = \text{heads}\}] = \frac{1}{n} \cdot n \cdot \theta = \boxed{\theta}.$$

How do we tell how good an estimate $\hat{\theta}(X^{(1)}, \dots, X^{(n)})$ for θ is? For example, if we had an estimate $\hat{\theta}(X^{(1)}, \dots, X^{(n)})$ that was just always 0 regardless of the training data we collect, then intuitively such an estimate for θ would probably be quite awful.

The *bias* of an estimate $\hat{\theta}$ for a parameter θ is

$$\mathbb{E}[\hat{\theta}(X^{(1)}, \ldots, X^{(n)})] - \theta,$$

where we note that $\hat{\theta}(X^{(1)}, \ldots, X^{(n)})$ is a random variable.

- Is the ML estimate for θ unbiased, i.e., it has bias equal to 0?

  - (selected) Yes ✔
  - No

  Solution:

  $\boxed{\text{Yes}}$. The answer to the previous part is equal to θ, so $\mathbb{E}[\hat{k}_{\text{ML}}] - \theta = \theta - \theta = 0$.

- Consider a terrible estimator $\hat{\theta}(X^{(1)}, \ldots, X^{(n)}) = 0$ regardless of what the training data are. What is the bias of this terrible estimator?

  In this part, please provide your answer as a mathematical formula (and not as Python code). Use ^ for exponentiation, e.g., `x^2` denotes x^2. Explicitly include multiplication using `*`, e.g. `x*y` is xy.

  You can use the variable `theta`.

-theta

Answer: -theta

$-\theta$

Solution:

Clearly the expectation of $\hat{\theta}$ in this case is 0, so the bias is going to be

$$0 - \theta = \boxed{\theta}.$$

The variance of an estimator θ is

$$\mathbf{var}(\hat{\theta}) = \mathbb{E}[(\hat{\theta} - \mathbb{E}[\hat{\theta}])^2].$$

Recall that for any random variable Z and any constant $a \in \mathbb{R}$,

$$\mathbf{var}(aZ) = a^2 \mathbf{var}(Z),$$

and for random variables $Z_1, \ldots, Z_n$ that are i.i.d. each with the same distribution as Z,

$$\mathbf{var}\Big(\sum_{i=1}^{n} Z_i\Big) = n\mathbf{var}(Z).$$

- For the ML estimate $\hat{\theta}(X^{(1)}, \ldots, X^{(n)}) = \frac{n_{\text{heads}}}{n}$ for the probability of heads of the coin, what is the variance of this estimator? Express your answer as a function of the true probability of heads θ and the number of training data points n (unless one or both of these don't show up in the answer).

  In this part, please provide your answer as a mathematical formula (and not as Python code). Use `^` for exponentiation, e.g., `x^2` denotes x^2. Explicitly include multiplication using `*`, e.g. `x*y` is xy.

  You can use the variable `theta`.

  theta*(1-theta)/n

  Answer: theta*(1-theta)/n

  $\frac{\theta \cdot (1-\theta)}{n}$

  Solution:

  We have

$$\begin{aligned}
\text{var}(\hat{\theta}_{\text{ML}}) &= \text{var}\Big(\frac{1}{n}\sum_{i=1}^{n} \mathbf{1}\{X^{(i)} = \text{heads}\}\Big) \\
&= \frac{1}{n^2}\text{var}\Big(\sum_{i=1}^{n} \mathbf{1}\{X^{(i)} = \text{heads}\}\Big) \\
&= \frac{1}{n^2} \cdot n \cdot \text{var}(\mathbf{1}\{X^{(1)} = \text{heads}\})
\end{aligned}$$

$$= \frac{1}{n^2} \cdot n \cdot \text{var}(\text{Ber}(\theta)\})$$
$$= \frac{1}{n^2} \cdot n \cdot \theta(1-\theta)$$
$$= \boxed{\frac{\theta(1-\theta)}{n}}.$$

- For the terrible estimator $\hat{\theta}(X^{(1)}, \dots, X^{(n)}) = 0$ regardless of the training data, what is the variance of this estimator? Express your answer as a function of the true probability of heads θ and the number of training data points n (unless one or both of these don't show up in the answer).

  In this part, please provide your answer as a mathematical formula (and not as Python code). Use `^` for exponentiation, e.g., `x^2` denotes x^2. Explicitly include multiplication using `*`, e.g. `x*y` is xy.

  You can use the variable `theta`.

  0

  ✔ **Answer:** 0

  0

  Solution:

$$\text{var}(\hat{\theta}(X^{(1)}, \dots, X^{(n)})) = \text{var}(0) = \boxed{0}.$$

Submit You have used 1 of 5 attempts

✔ Correct (5/5 points)

Discussion

Topic: Introduction to Parameter Learning - Maximum Likelihood and MAP Estimation / Exercise: Bias and Variance

Show Discussion