Quiz 22 Solutions

written by Alvin Wan . alvinwan.com/cs70

Thursday, April 14, 2016

This quiz does not count towards your grade. It exists to simply gauge your understanding. Treat this as though it were a portion of your midterm or final exam. Today, we will walk through conditional expectation with problems of increasing difficulty.

1 Linear Regression

1. We are pulling marbles at random from a bag and the placing the marble back in before sampling again. Let X be the number times we sample before we get a red marble. Let Y be the number of times we sample before we get a blue marble. Assume there are a red, b blue, and c randomly-colored marbles in the bag. In this problem, we will walk through computing $L[Y|X]$.

   Note: This problem is largely to walk you through computing these values. Linear regression is useful when you have only *data* and need to find a model to fit it.

   (a) Compute $E[X]$.

   Solution: Note $X \sim Geom(\frac{a}{a+b+c})$. Thus $E[X] = \frac{a+b+c}{a}$.

   (b) Compute $E[Y]$.

   Solution: Note $Y \sim Geom(\frac{b}{a+b+c})$. Thus $E[Y] = \frac{a+b+c}{b}$.

   (c) Compute $var(X)$

   Solution: Since this is geometric, the variance is $\frac{1-p}{p^2}$, where $p = \frac{a}{a+b+c}$

   (d) Compute $cov(X, Y)$. (Note that this gives us a fairly boring result.)

   Solution: X and Y are independent, so $cov(X, Y) = 0$.

(e) Finally, compute $L[Y|X]$. Remember that the formula for linear regression is the following:

$$L[Y|X] = E[Y] + \frac{cov(X,Y)}{var(X)}(X - E[X])$$

Solution: Since, $cov(X,Y) = 0$, we have that $L[Y|X] = E[Y]$. This is always true for two independent random variables X and Y. Unfortunately, it's not very interesting.

2. You are given a set of sample points, taken uniformly at random from $\Omega = \{1,2,3,4,5,6\}$.

Ω	1	2	3
$X(\Omega)$	2	4	6
$Y(\Omega)$	2	1	3

(a) Compute $E[X]$.

Solution: By inspection, $E[X] = 4$. Simply add take the average of the three points, assuming a uniform distribution (as stated in the question).

$$E[X] = \sum_x xPr[X = x] = \frac{1}{3}(2 + 4 + 6) = 4$$

(b) Compute $E[Y]$.

Solution: By inspection, $E[Y] = 2$.

(c) Compute $var(X)$.

Solution: $\frac{8}{3}$

We know the following.

$$var(X) = E[X^2] - E[X]^2$$

We have that $E[X^2] = 4$ from part a. Now, compute $E[X^2]$.

$$\begin{aligned} E[X^2] &= \sum_x x^2 P[X = x] \\ &= \frac{1}{3}(2^2 + 4^2 + 6^2) \\ &= \frac{1}{3}(4 + 16 + 36) \end{aligned}$$

$$= \frac{56}{3}$$

Given that $E[X] = 4$, $E[X]^2 = 16$. Thus,

$$var(X) = \frac{56}{3} - \frac{48}{3} = \frac{8}{3}$$

(d) Compute $cov(X, Y)$.

Solution: $\frac{2}{3}$

We know that to compute $cov(X, Y)$, we need

$$cov(X, Y) = E[XY] - E[X]E[Y]$$

Now, we compute $E[XY]$.

$$\begin{aligned} E[XY] &= \sum_{x,y} xyPr[X = x, Y = y] \\ &= \frac{1}{3}(4 + 4 + 18) \\ &= \frac{26}{3} \end{aligned}$$

We know that $E[X] = 4$ and $E[Y] = 2$, so $E[X]E[Y] = 8$, making

$$cov(X, Y) = \frac{26}{3} - \frac{24}{3} = \frac{2}{3}$$

Thus, $cov(X, Y) = \frac{2}{3}$.

(e) Compute $L[Y|X]$.

Solution: We plug into the formula.

$$\begin{aligned} L[Y|X] &= E[Y] + \frac{cov(X, Y)}{var(X)}(X - E[X]) \\ &= 2 + \frac{2/3}{8/3}(X - 4) \\ &= 2 + 4(X - 4) \\ &= 4X - 14 \end{aligned}$$

(f) Estimate Y, when $X = 4$. (It's not 1!)

Solution: 2

$$L[Y|X = 4] = 4X - 14 = 16 - 14 = 2$$