Homework 01

Imola Fodor SM3500474

Probabilistic Machine Learning 2022

Exercise 1.

The time in takes (in hours) to travel back home from the university is modeled by the density function

$$f(x) = \begin{cases} 1 - \frac{x}{2}, & 0 < x < 2 \\ 0, & otherwise \end{cases}$$

The cost for a trip of *x* hours is $y = 8 + 12x$ €. Compute the expected cost.

Result

If Y is a continuous random variable with the probability density function (pdf) $f_Y(y)$, then the expected value of Y is given by

$$\mathrm{E[Y]} = \int_{-\infty}^{\infty} \mathrm{y} \cdot \mathrm{f}_Y(\mathrm{y})\mathrm{dy}$$

In order to compute the expected value of Y, we need to know the pdf $f_Y(y)$.

There are two ways to compute the expected value. Since we do not have the knowledge of the pdf, we can use directly the LOTUS methodology for the solution, or we can work out first the pdf and subsequently the E[Y] formula reported above. Below the two approaches.

LOTUS – Law of The Unconscious Statistician

For a continuous random variable, defined through another random variable, the expected value can be written as:

$$E[g(X)] = \int_{-\infty}^{\infty} g(x) f_X(x) dx$$

, where g(X) stands for Y, and g(x) stands for y., in our case.

The solution reported below:

$$E[Y] = \int_{-\infty}^{\infty} (8+12x) fX(x)dx = \int_0^2 (8+12x)\left(1-\frac{x}{2}\right)dx = = \int_0^2 8+8x-6x^2 dx$$

$$= 16 + \frac{32}{2} - \frac{48}{3} = 32 - 16 = 16$$

Therefore, we can state that the estimated cost is 16€.

Long Way

In order to get the pdf of the random variable Y, we first need to find it's cumulative density function (cdf), differentiate it, and further use the initial formula of the expected value, $\mathrm{E[Y]} = \int_{-\infty}^{\infty} \mathrm{y} \cdot \mathrm{f(y)dy}$**.**

1. Find cdf of Y, $F_Y(y)$.

$$F_Y(y) = P(Y \leq y) = P(8+12X \leq y) = P\left(X \leq \frac{y-8}{12}\right)$$

By definition $P(X \leq k) = \int_{-\infty}^{k} f(x)dx$:

$$F_Y(y) = P\left(X \leq \frac{y-8}{12}\right) = \int_0^{\frac{y-8}{12}} 1 - \frac{x}{2} dx = -\frac{y^2}{576} + \frac{y}{9} - \frac{7}{9}$$

2. Differentiate $f_Y(y)$, to get the pdf of Y, $f_Y(y)$.

$$f_Y(y) = \frac{d}{dy} F_Y(y) = \begin{cases} \frac{1}{9} - \frac{y}{488}, & 8+12*0 < x < 8+12*2 \\ 0, & otherwise \end{cases}$$

$$E[Y] = \int_8^{32} y\left(\frac{1}{9} - \frac{y}{488}\right)dy = 16$$

Likewise, we can state that the estimated cost is 16€.

Note: For a concise result, intermediate computational steps are excluded.

Exercise 2.

Two random variables X and Y are said to be correlated if and only if their covariance is non-zero. Can two independent random variables X and Y be correlated?

Result

By their nature, when it comes to independent random variables, knowing the value of one of them, does not change the probabilities of the other. The joint probability of two independent random variables can therefore be written as:

$$P(X,Y) = P(X) \cdot P(Y)$$

Similarly, even though the expected value of the product of two random variables need not be equal to the product of their expectation, for independent variables the following still holds:

$$E[XY] \;=\; E[X] \cdot E[Y]$$

Furthermore, we define the identity for covariance, to later make use of it.

$$\begin{aligned} Cov(X,Y) &= \mathrm{E}[(\mathrm{X} - \mathrm{E}[\mathrm{X}]) \cdot (\mathrm{Y} - \mathrm{E}[\mathrm{Y}])] \\ &= E(XY) - E(X) \cdot E(Y) \end{aligned}$$

Two answer whether two independent random variables X and Y can be correlated, we can prove that their covariance is 0, and therefore conclude the non-correlation.

$$Cov(X,Y) \;=\; E(XY) - E(X) \cdot E(Y) \;=\; E(X) \cdot E(Y) \;-\; E(X) \cdot E(Y) \;=\; 0$$

Two independent random variables X and Y cannot be correlated, since their covariance is 0.

Exercise 3.

You buy a laptop whose lifetime is the number of years it works properly before breaking down. You know that $P(T \geq \; t) = e^{-\frac{t}{5}}$ for all $t \geq 0$, corresponding to the probability that it lasts at least years.

If the laptop still works properly after two years, what is the probability that it breaks in the third year?

Result

In order to get the answer to the probability that the laptop breaks in the third year, given it is still working after two years, we pose the following formula of conditional probability:

$$P(A|B) = \frac{P(A)}{P(B)} = \frac{P(2 \leq T \leq \; 3)}{P(T \geq 2)}$$

To calculate the probability that T, lies between the two values 2 and 3 (years) we use the following result:

$$P(2 \leq T \leq 3) = \int_{2}^{3} f_T(x)dx$$

For $P(x \geq 2)$ instead:

$$P(T \geq 2) = \int_{2}^{\infty} f_T(x)dx$$

Getting $f_T(x)$, the pdf of our random variable T, is possible through cdf of $F_T(t) = P(T \geq t) = e^{-t/5}$, as follows:

$$P(T \geq t) = \int_t^{\infty} f_T(x)dx$$
$$\frac{d}{dt} e^{-\frac{t}{5}} = f_T(x)$$
$$-\frac{1}{5} e^{-\frac{t}{5}} \Big|_t^0 = f_T(x)$$
$$\frac{1}{5} e^{-\frac{t}{5}} = f_T(x)$$

Plugging in the gotten formula for the pdf:

$$P(2 \leq T \leq 3) = \int_2^3 \frac{1}{5} e^{-\frac{t}{5}} = 0.12$$

$$P(T \geq 2) = \int_2^{\infty} \frac{1}{5} e^{-\frac{t}{5}} = 0.67$$

Finally:

$$\frac{P(2 \leq T \leq 3)}{P(T \geq 2)} = \frac{0.12}{0.67} = 0.17$$

The probability of the laptop breaking in it’s third year of “life” is 0.17.

Exercise 4.

Find the VC dimension of the set of functions

$$H = \{1_{x \in C} \ : \ C \ is \ a \ convex \ polytope \ in \ \mathbb{R}^2 \},$$

where $1_{x \in C}$ is the indicator function.

Result

The VC dimension of a classifier is defined by Vapnik and Chervonenkis to be the cardinality (size) of the largest set of points that the classification algorithm can shatter, namely perfectly partition the plane so that the positive points are separated from the negative points, no matter the labeling of the points.

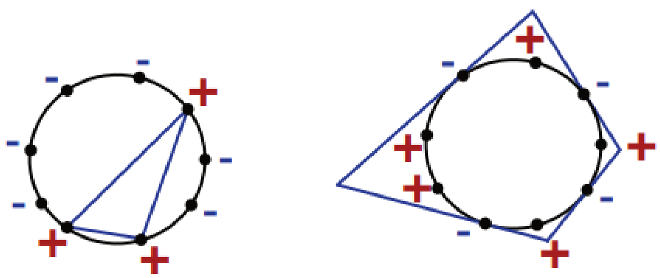
When it comes to the set of functions defined by H, they represent the different convex polygons.

This class of convex d-gons has no more than d vertices in R^2, and lies on a plane.

To get a lower bound, we show that any set of 2d+1 points can be shattered. To do this, we select 2d+1 points that lie on a circle, and for a particular labeling, if there are more negative than positive labels, then the points with the positive labels are used as the polygon's vertices, as in figure (a). Otherwise, the tangents of the negative points serve as the edges of the polygon, as shown in (b).

To derive an upper bound, it can be shown that choosing points on the circle maximizes the number of possible dichotomies, and thus VCdim(convex d-gons) = 2d + 1.

Note also that VCdim(convex polygons) = $+\infty$.



|positive points| < |negative points| (a)

|positive points| > |negative points| (b)

Bibliography

Mohri et al - Foundations of Machine Learning