ML dseb62 w1

Nguyen Tuan Duy - 11204971

August 2022

1 Problem 1

	x1	x2	x3	x4	x5	p(y)
y1	0.01	0.02	0.03	0.1	0.1	0.26
y2	0.05	0.1	0.05	0.07	0.2	0.47
y3	0.1	0.05	0.03	0.05	0.04	0.27
p(x)	0.16	0.17	0.11	0.22	0.34	1

a) Based on the table above:

$$\begin{aligned} p(x) &= p(x1) + p(x2) + p(x3) + p(x4) \\ &= 0.16 + 0.17 + 0.11 + 0.22 + 0.34 \\ &= 1 \end{aligned} \tag{1}$$

$$\begin{aligned} p(y) &= p(y1) + p(y2) + p(y3) \\ &= 0.26 + 0.47 + 0.27 \\ &= 1 \end{aligned} \tag{2}$$

b) we have the following formula for conditional probability:

$$p(x_i|y = y_1) = \frac{p(x_i, y_1)}{p(y_1)} \tag{3}$$

by applying this formula we can calculate

$$\begin{aligned} p(x_1|Y = y_1) &= \frac{p(x_1, y_1)}{p(y_1)} \\ &= \frac{0.01}{0.26} \\ &= \frac{1}{26} \end{aligned} \tag{4}$$

by performing the same operation on all values of X, we have the following table:

i	1	2	3	4	5
$p(x_i\|Y=y_1)$	1/26	1/13	3/26	5/13	5/13

similarly, we can calculate $p(x|Y=y_3)$ and obtain the following result:

i	1	2	3	4	5
p($x_i\|Y=3$)	10/27	5/27	1/9	5/27	4/27

2 Problem 2

For discrete variables:

$$\begin{aligned} E_y[E_x[x|y]] &= E_y[\sum_x x * p(x|y)] \\ &= \sum_y [\sum_x xp(x|y)]p(y) \\ &= \sum_y \sum_x xp(x,y) \\ &= \sum_x x \sum_y p(x,y) \\ &= \sum_x xp(x) \\ &= E_x[x] \end{aligned} \tag{5}$$

For continuous variables:

E$_y[E_x[x|y]] = \int E_x[x|y] f_y(y)dy$
$= \int \int x f_{X|Y}(x|y) f_y(y) dxdy$
$= \int \int x f(x,y)dxdy$
$= \int xdx \int f(X,y)dy$
$= \int x f(x)dx$
$= E_x[x]$
(6)

3 Problem 3

Let x be the event that people use product X
Let y be the event that people use product Y
According to the problem, we have:
$p(x) = 0.207$
$p(y) = 0.5$
$p(x|y) = 0.365$
a)

$$\begin{aligned} p(x, y) &= p(x|y) * p(y) \\ &= 0.365 * 0.5 \\ &= 0.1825 \end{aligned} \tag{7}$$

b) $p(\bar{x}) = 1 - p(x) = 0.793$
$p(\bar{x}|y) = 1 - p(x|y) = 1 - 0.365 = 0.635$

$$\begin{aligned} p(y|\bar{x}) &= \frac{p(y, \bar{x})}{p(\bar{x}} \\ &= \frac{p(\bar{x}|y)p(y)}{p(\bar{x})} \\ &= \frac{0.635 * 0.5}{0.793} \\ &= 0.4 \end{aligned} \tag{8}$$

4 Problem 4

raw-score expression for the variance:

$$\sigma^2 = \frac{\sum_N (x-\mu)^2}{N}$$
$$= \frac{\sum_N (x^2 - 2x\mu + \mu^2)}{N}$$
$$= \frac{\sum_N x^2}{N} - \frac{\sum_N 2x\mu}{N} + \frac{\sum_N \mu^2}{N}$$
$$= \frac{\sum_N x^2}{N} - 2\mu^2 + \frac{N * \mu^2}{N}$$
$$= \frac{\sum_N x^2}{N} - 2\mu^2 + \mu^2$$
$$= \frac{\sum_N x^2}{N} - \mu^2$$
$$= E[x^2] - E[x]^2$$
(9)

5 Problem 5

Let x_i be the event that the player choose door i (i $\in$ {1, 2, 3})
Let c_i be the event that the car is in door i (i $\in$ {1, 2, 3})

Let m_i be the event that Monty chooses door (i $\in \{1, 2, 3\}$) Suppose that Monty chooses door 3 and player chooses door 1, we have $p(x1|c1) = \frac{1}{3}$
$p(c1) = 1/3$
$p(x1) = 1$
$p(m3|c2, x1) = 1$
$p(m3|c1, x1) = \frac{1}{2}$
$p(m3|x1) = \frac{1}{2}$
The event that we win a car by staying is that the car is in door 1 given that we chooses door 1 and monty chooses door 3:

$$\begin{aligned} p(c1|x1, m3) &= \frac{p(c1, x1, m3)}{p(x1, m3)} \\ &= \frac{p(m3|x1, c1)}{p(m3|x1)p(x1)} \\ &= \frac{\frac{1}{2}\frac{1}{3}}{\frac{1}{2}} \\ = \frac{1}{3} \end{aligned} \tag{10}$$

The event that we win a car by switching is that the car is in door 2 given that we chooses door 2 and monty chooses door 3:

$$\begin{aligned} p(c2|x1, m3) &= \frac{p(c2, x1, m3)}{p(x1, m3)} \\ &= \frac{p(m3|x1, c2)}{p(m3|x1)p(x1)} \\ &= \frac{1\frac{1}{3}}{\frac{1}{2}} \\ &= \frac{2}{3} \end{aligned} \tag{11}$$

In conclusion, the probability that we win by staying is only $\frac{1}{3}$ whereas if we switch, the probability that we win is $\frac{2}{3}$, thus switching is better in this case