Nguyen Vu Anh Ngoc
Machine Learning 1
August 22, 2023

Homework Week 2: Basic Probability

Question 1. Compute

a) The marginal distributions $p(x)$ and $p(y)$.

$p(x_1) = 0.16$
$p(x_2) = 0.17$
$p(x_3) = 0.11$
$p(x_4) = 0.22$
$p(x_5) = 0.34$

$p(y_1) = 0.26$
$p(y_2) = 0.47$
$p(y_3) = 0.27$

b) The conditional distributions $p(x|Y = y_1)$ and $p(y|X = x_3)$.

$$p(x_1|y_1) = \frac{0.01}{0.26} = 0.038$$
$$p(x_2|y_1) = \frac{0.02}{0.26} = 0.077$$
$$p(x_3|y_1) = \frac{0.03}{0.26} = 0.114$$
$$p(x_4|y_1) = \frac{0.01}{0.26} = 0.038$$
$$p(x_5|y_1) = \frac{0.01}{0.26} = 0.038$$

$$p(y_1|x_3) = \frac{0.03}{0.11} = 0.273$$
$$p(y_2|y_3) = \frac{0.05}{0.11} = 0.454$$
$$p(y_3|y_3) = \frac{0.03}{0.11} = 0.273$$

Question 2. Prove $E_X[x] = E_Y[E_x[x \mid y]]$.

$$\begin{aligned}
&(1) & E_Y[E_x[x \mid y]] &= E_Y[\int xp(x|y)\,dx] \\
&(2) & &= \int p(y) \int xp(x|y)\,dxdy \\
&(3) & &= \iint p(y)x\frac{p(x,y)}{p(y)}\,dxdy \\
&(4) & &= \iint xp(x,y)\,dxdy \\
&(5) & &= \int x \int p(x,y)\,dy\,dx \\
&(6) & &= \int xp(x)\,dx \\
&(7) & &= E_X[x]
\end{aligned}$$

Question 3. Prove $V_X[x] = E_X[x^2] - (E_x[x])^2$.

$$\begin{aligned}
&(1) & V_X[x] &= \sum_{i=1}^{n}(x_i - \mu)^2 p(x_i) \\
&(2) & &= E[(x-\mu)^2] \\
&(3) & &= E[x^2 - 2\mu x + \mu^2] \\
&(4) & &= E[x^2] - 2\mu E[x] + \mu^2 \\
&(5) & &= E[x^2] - 2E[x]^2 + E[x]^2 \\
&(6) & &= E_X[x^2] - (E_x[x])^2
\end{aligned}$$

Question 4. In a study, physicians were asked what the odds of breast cancer would be in a woman who was initially thought to have a risk of cancer but who ended up with a positive mammogram result (a mammogram accurately classifies about 80% of cancerous tumors and 90% of benign tumors). 95 out of a hundred physicians estimated the probability of cancer to be about 75%. Do you agree?

$$\begin{aligned} P(P|C) = 0.8 &\rightarrow P(\neg P|C) = 0.2 \\ P(\neg P|\neg C) = 0.9 &\rightarrow P(P|\neg C) = 0.1 \\ P(C) = 0.01 &\rightarrow P(\neg C) = 0.99 \end{aligned}$$

C: event that one has breast cancer.
P: event that one has positive mammogram result.

$$P(C|P) = \frac{P(P|C)P(C)}{P(P)} \tag{1}$$
$$= \frac{P(P|C)P(C)}{P(P,C) + P(P,\neg C)} \tag{2}$$
$$= \frac{P(P|C)P(C)}{P(P|C)P(C) + P(P|\neg C)P(\neg C)} \tag{3}$$
$$= \frac{0.8 \times 0.01}{0.8 \times 0.01 + 0.1 \times 0.99} \tag{4}$$
$$\approx 0.075 \tag{5}$$

The probability is 7.5% far more lower than 75% expected by physicians.

Question 5. You find yourself on a game show, and the host presents you with four doors. Behind three doors are an assortment of gummy bears, and the remaining door has a pure gold car behind it!You pick a door, and then the host reveals a door behind which he knows is only a couple of red gummy bears. Then you are given the choice to stick with your first choice or switch to one of the other two unopened doors.
If the probability that you will win the car if you switch is $\frac{a}{b}$, where a and b are coprime positive integers, what is $a + b$?

X: event of winning.
O: event of choosing the right door from the beginning.

$P(O) = \frac{1}{4}, P(\neg O) = \frac{3}{4}$

$P(X) = P(X, O) + P(X, \neg O) = P(X|O)P(O) + P(X|\neg O)P(\neg O)$

	$P(X\|O)$	$P(X\|\neg O)$	$P(X)$
Switch	0	1/2	$0 \times 1/4 + 1/2 \times 3/4 = 3/8$
No Switch	1	0	$1 \times 1/4 + 0 \times 3/4 = 1/4$

Since $\frac{3}{8} > \frac{1}{4}$, we should switch the door. $a + b = 11$.