Probability HW2

109350008 張詠哲

Problem1

Problem 1.

(a). We need to proof $P(A \cap B) = P(A) + P(B)$

by the problem statement we can know that:

$P(A) = 1-p$, $P(B) = p \cdot \varepsilon_0 + (1-p) \cdot \varepsilon_1$

$P(A \cap B) = (1-p) \cdot \varepsilon_1$

∴ We need to substitude into independent condion.

$$(1-p) \cdot \varepsilon_1 = (1-p) \cdot \left(p \cdot \varepsilon_0 + (1-p) \cdot \varepsilon_1\right)$$

If $1-p \neq 0 \Rightarrow p \neq 1$.

$\Rightarrow \varepsilon_1 = p \cdot \varepsilon_0 + (1-p) \cdot \varepsilon_1$

$\Rightarrow \varepsilon_1 = p \cdot \varepsilon_0 + \varepsilon_1 - p \cdot \varepsilon_1$

$\Rightarrow p(\varepsilon_0 - \varepsilon_1) = 0$

∴ If $\begin{cases} p = 0 \\ p = 1 \\ \varepsilon_1 = \varepsilon_0 \end{cases}$ conditions hold, We can say that A & B are independent event.

(b). We need to proof

$$P(A \cap B \mid C) = P(A \mid C) \cdot P(B \mid C)$$

We can know $P(C) = p \cdot P(C \mid \text{bit } 0) + (1-p) \cdot P(C \mid \text{bit } 1)$

$$= p(\alpha_0 + \varepsilon_0) + (1-p)(\alpha_1 + \varepsilon_1)$$

$$P(A \mid C) = \frac{P(A \cap C)}{P(C)} = \frac{(1-p) \cdot (\alpha_1 + \varepsilon_1)}{P(C)}$$

$$P(B \mid C) = \frac{P(B \cap C)}{P(C)} = \frac{p \cdot \varepsilon_0 + (1-p) \cdot \varepsilon_1}{P(C)}$$

$$P(A \cap B \mid C) = \frac{P(A \cap B \cap C)}{P(C)} = \frac{(1-p) \cdot \varepsilon_1}{P(C)}$$

$$\therefore \frac{(1-p) \cdot \varepsilon_1}{P(C)} = \frac{(1-p) \cdot (\alpha_1 + \varepsilon_1)}{P(C)} \cdot \frac{p \cdot \varepsilon_0 + (1-p) \cdot \varepsilon_1}{P(C)}$$

$$\Rightarrow \frac{(1-p) \cdot \varepsilon_1}{P(C)} = \left(\frac{\alpha_1 (1-p)}{P(C)} + \frac{\varepsilon_1 (1-p)}{P(C)} \right) \cdot \left(\frac{p \varepsilon_0}{P(C)} + \frac{\varepsilon_1 (1-p)}{P(C)} \right)$$

$\therefore$ by the above equation, we can know that A & B are not conditionally independent given C, unless some specific condition, such as $\varepsilon_1 = \varepsilon_0 = (1-\alpha_1)$

(1) Let $X \in \{0,1\}$ be the transmitted bit.

by Total Probability theorem

We can know

$$P(\text{correct Delivery}) = p \cdot P(\text{correct} \mid X=0) + (1-p)\, P(\text{correct} \mid X=1)$$

$$\Rightarrow P(\text{correct Delivery}) = \underline{p(1-\varepsilon_0-\alpha_0) + (1-p)\cdot(1-\varepsilon_1-\alpha_1)}$$

Problem2

Problem 2.

(a). Let $\begin{cases} X \text{ be the number of 1s transmitted} \\ Y \text{ be the number of 0s transmitted} \end{cases}$

Since each bit is either 1 or 0.

$\therefore Z = X+Y$, $Z \sim \text{Posson}(\lambda)$ and $X \mid Z=z \sim \text{Binomial}(z,p)$

$$P(X=x) = \sum_{z=0}^{\infty} P(X=x \mid Z=z)\, P(Z=z)$$

$$\Rightarrow P(X=x) = \sum_{z=x}^{\infty} \frac{z!}{(z-x)!} (1-p)^{z-x} \cdot \frac{\lambda^z e^{-\lambda}}{z!}$$

$$= \frac{p^x \lambda^x e^{-\lambda}}{x!} \sum_{z=x}^{\infty} \frac{[\lambda(1-p)]^{z-x}}{(z-x)!}$$

$$= \frac{p^x \lambda^x e^{-\lambda}}{x!} e^{\lambda(1-p)}$$

$$= \frac{(p\lambda)^x e^{-x\lambda}}{x!}$$

$\therefore$ We can know that $X \sim \text{Posson}(p\lambda)$

(b) 1. From the question, we can know $P(X_i=k)=(1-p)^{k-1}p$, $k=1,2,\dots$

The CDF of X_i is $F_{X_i}(k)=P(X_i\le k)=1-(1-p)^k$.

$X=\max(X_1,X_2,X_3\dots X_n)$

$F_X(k)=P(X\le k)=P(X_1\le k, X_2\le k\dots X_n\le k)=\underbrace{\prod_{k=1}^{n}P(X_i\le k)}_{\text{Cause independent}}$

$=[F_X(k)]^n=(1-(1-p))^{k}$

$P(X=k)=F_X(k)-F_X(k-1)=(1-(1-p))^k-(1-(1-p))^{k-1}$

2. We want find $F_Y(k)-F_Y(k-1)$, for $Y=\min(X_1,X_2\dots X_n)$

note that $F_Y(k)=P(Y\le k)$ is the probability of at least one $X_i\le k$

$\therefore\quad F_Y(k)=P(Y\le k)=1-P(Y>k)=1-\prod_{i=1}^{n}P(X_i>k)$

$=1-[(1-p)^k]^n$

$=1-(1-p)^{kn}$

$P(Y=k)=F_Y(k)-F_Y(k-1)=(1-(1-p)^{kn})-(1-(1-p)^{(k-1)n})$

$=(1-p)^{n(k-1)}(1-(1-p)^n)$

3. Let $p'=1-(1-p)^n$. We can know $P(Y=k)=p'(1-p')^{(k-1)}$

$\boxed{\therefore\ Y\sim \text{Geometric}(p')}$

(c) We can take the length of sequence as the trial, so we will get

$$P(X_{s_2}=k) = C_k^n \, p_T^k (1-p_T)^{n-k} \quad \text{for } n \text{ trial}$$

k is number of T

Since $n=123$

$$\therefore \; P(X_{s_2}=k) = C_k^{123} \, P_T^k (1-P_T)^{123-k}$$

Problem3

Problem3.

(a) We need to proof $H(\frac{1}{2},\frac{1}{3},\frac{1}{6}) = H(\frac{1}{2},\frac{1}{2}) + \frac{1}{2}\cdot H(\frac{2}{3},\frac{1}{3})$

LHS: $H(\frac{1}{2},\frac{1}{3},\frac{1}{6}) = -(\frac{1}{2}\ln\frac{1}{2} + \frac{1}{3}\ln\frac{1}{3} + \frac{1}{6}\ln\frac{1}{6})$

RHS: $H(\frac{1}{2},\frac{1}{2}) = -(\frac{1}{2}\ln\frac{1}{2} + \frac{1}{2}\ln\frac{1}{2}) = \ln 2$

$H(\frac{2}{3},\frac{1}{3}) = -(\frac{2}{3}\ln\frac{2}{3} + \frac{1}{3}\ln\frac{1}{3})$

$$\begin{aligned}
H(\tfrac{1}{2},\tfrac{1}{2}) + \tfrac{1}{2}H(\tfrac{2}{3},\tfrac{1}{3}) &= \ln 2 - \tfrac{1}{2}(\tfrac{2}{3}\ln\tfrac{2}{3} + \tfrac{1}{3}\ln\tfrac{1}{3}) \\
&= \ln 2 - \tfrac{1}{3}(\ln 2 - \ln 3) - \tfrac{1}{6}\ln\tfrac{1}{3} \\
&= \tfrac{2}{3}\ln 2 + \tfrac{1}{3}\ln 3 - \tfrac{1}{6}\ln\tfrac{1}{3} \\
&= -\tfrac{1}{3}\ln\tfrac{1}{2} - \tfrac{1}{6}\ln\tfrac{1}{4} - \tfrac{1}{3}\ln\tfrac{1}{3} - \tfrac{1}{6}\ln\tfrac{1}{3} \\
&= -\tfrac{1}{3}\ln\tfrac{1}{2} - \tfrac{1}{3}\ln\tfrac{1}{3} - \tfrac{1}{6}\ln\tfrac{1}{12} \\
&= -\tfrac{1}{2}\ln\tfrac{1}{2} - \tfrac{1}{3}\ln\tfrac{1}{3} - \tfrac{1}{6}\ln\tfrac{1}{6}
\end{aligned}$$

∴ We can get LHS=RHS

(b) Because we only have 2 possible outcome with probability $p_1 = p$, $p_2 = (1-p)$. So, we can get

$$H(X) = -p\ln p - (1-p)\ln(1-p)$$

maximum: the maximum when the outcome are equally likely. ex. $p_1 = p_2 = 0.5$, we will get:

$$H_{max} = -0.5\ln 0.5 - 0.5\ln 0.5 = \ln 2.$$

minimum: the minimum when outcome are extreme value, like $p_1 = 0 / p_2 = 1$ or $p_1 = 1 / p_2 = 0$. Then, we will get:

$$H_{min} = 0.$$

(c). Because entropy is to estimate the uncertainty, so, we will get the maximuined entropy when the possible outcome are qually like, even the $n \geq 2$.

$\therefore$ When $p_1 = p_2 = p_3 \cdots = p_n = \frac{1}{n}$, the $H(X)$ is maximum $(= \ln n)$

(d) the entropy is minimum when the outcome is determinated.

$\therefore$ when one of outcome equal to 1 (ex. $p_1 = 1$, $p_2 = p_3 \cdots = p_n = 0$) the $H(X) = -(1\ln 1 + 0\ln 0 + \cdots) = 0$ is the minimum.

Problem4

Problem 4.

(a). i). We need to proif $E(X)=\frac{1}{p}$

We know the $X \sim \text{Geometric}(p)$

So, the $P(X=k) = (1-p)^{k-1}p$, $k=1,2,3\ldots$

$E(X) = \sum_{k=1}^{\infty} k\,P(X=k) = \sum_{k=1}^{\infty} k(1-p)^{k-1}p$

Let $t=(1-p)$

We can get $\sum_{k=1}^{\infty} k(1-p)^{k-1}p = p\sum_{k=1}^{\infty} k\,t^{k-1}$

We know the $\sum_{k=1}^{\infty} t^k = \frac{1}{(1-t)}$ (for $|t|<1$)

$$\frac{d}{dt}\left(\sum_{k=1}^{\infty} t^k\right) = \frac{d}{dt}\left(\frac{1}{(1-t)}\right)$$

$$\Rightarrow \sum_{k=1}^{\infty} k\,t^{k-1} = \frac{1}{(1-t)^2}$$

$$\therefore p\sum_{k=1}^{\infty} k(1-p)^{k-1} = p\cdot\frac{1}{(1-(1-p))^2} = p\cdot\frac{1}{p^2} = \frac{1}{p} \ \#$$

ii) The moment-generating function $M_X(t) = E[e^{tX}]$

$$M_X(t) = \sum_{k=1}^{\infty} e^{tk} P(X=k) = \sum_{k=1}^{\infty} e^{tk} p(1-p)^{k-1}$$

$$\sum_{k=1}^{\infty} e^{tk} p(1-p)^{k-1} = e^t p \sum_{k=1}^{\infty} [e^t(1-p)]^{k-1}$$

Let $w = e^t(1-p)$, we can get:

$$e^t p \sum_{k=1}^{\infty} [e^t(1-p)]^{k-1} = e^t p \frac{1}{(1-w)} \quad (\text{for } |w|<1)$$

$$= \frac{p \cdot e^t}{(1-e^t(1-p))}$$

$\because |w|<1 \Rightarrow |e^t(1-p)|<1 \Rightarrow e^t < \frac{1}{(1-p)} \Rightarrow t < \ln(1-p)$

(b).

1. $Var[Z]$, $Var[Z] = E[Z^2] - (E[Z])^2$

$$E[Z] = \sum_{n=1}^{\infty} (-1)^n \sqrt{n} \cdot \frac{6}{(\pi n)^2}$$

$\because (-1)^n \quad \therefore E[Z] = 0$

$$E[Z^2] = \sum_{n=1}^{\infty} n \cdot \frac{1}{(\pi n)^2} = \frac{6}{\pi} \sum_{n=1}^{\infty} \frac{1}{n} = \infty$$

$\therefore Var[Z] = E[Z^2] - (E[Z])^2 = \infty - 0 = \infty$ (doesn't exist)

2. $$\sum_{n=1}^{\infty} z_n^3 P(Z=z_n) = \sum_{n=1}^{\infty} (-1)^n n^{\frac{3}{2}} \cdot \frac{6}{(\pi n)^2} = \infty$$

It is also diverage, cause the $n^{\frac{3}{2}}$.

So, the $\sum_{n=1}^{\infty} z_n^3 P(Z=z_n)$ doesn't exist

3. by the statment of (2). We can know that

$$E[Z^3] = \sum_{n=1}^{\infty} z_n^3 P(Z=z_n) = \infty \quad \text{doesn't exist.}$$

Problem5

(a) I achieve the 0.9892 accuracy with a 70%/30% partition of the dataset and a uniform prior

```
# TODO: Test the spam filter by using
y_pred = nb_model.predict(X_test_matrix)
accuracy = accuracy_score(y_test, y_pred)
print(f"ACC: {accuracy:.4f}")
```

```
ACC: 0.9892
```

(b)

The following table shows the experiment result of 2 different priors and 1 different partition of dataset.

Differences	Accuacy
Prior [0.6, 0.4]	0.9833
Prior [0.7, 0.3]	0.9856
Partition 8:2	0.9839

We can see that both different prior and dataset partition have similar accuracy values compare with the result in (a) (0.9892), which indicate that the prior and the dataset partition don't have a major impact in this dataset. I think the reason is that:

1. The dataset is already well-balanced or the model is robust to slight changes in the priors. So, the choice of priors may not significantly impact the final accuracy. (But if the dataset were imbalanced, choosing a prior closer to the actual class distribution could improve the performance.)

2. More training data (like 80/20 split) usually helps the model perform better. But the difference is minimal, indicating that the original 70/30 split was already enough to learn well from the data. Therefore, the results also show that small changes in the partition don't significantly affect the performance.