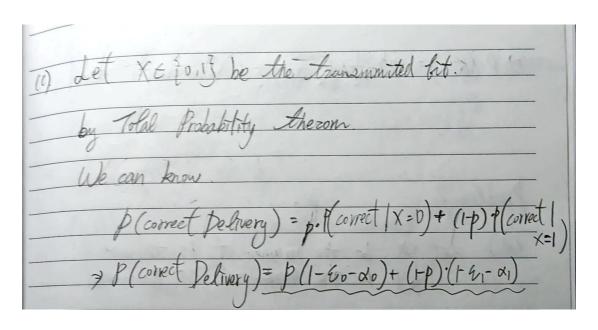
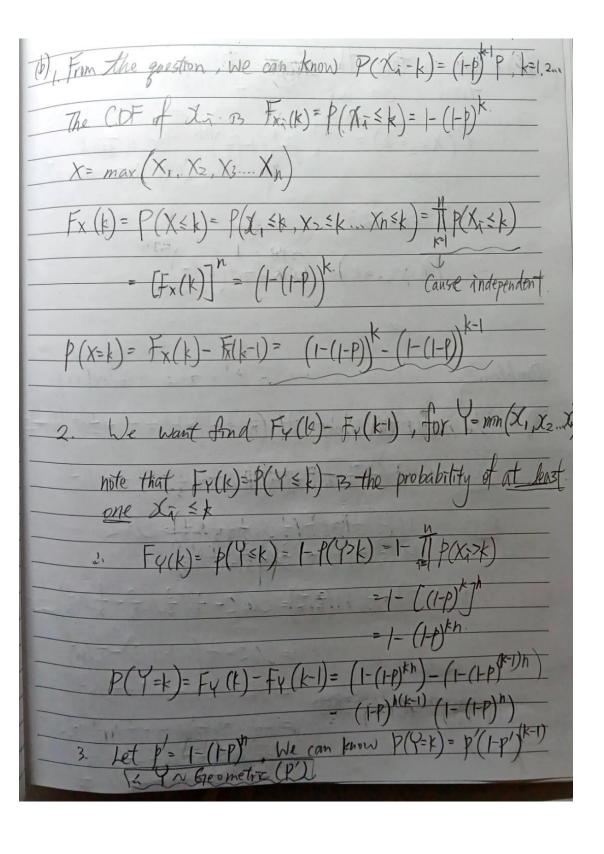
Probability HW2

109350008 張詠哲

Roblem 1.
(A). We need to proof P(ADB) = P(A) + P(B) by the problem statement we can know that:
by the problem statement we can know that:
p(A)=1-P, P(B)=p.E0+(1-p).E1
$P(A \cap B) = (1-p) \cdot \varepsilon,$
Z Wa need to substitude into independent condion.
(1-P). E1 = (1-p). (p. E0+(1P). E1)
If 1-p = 0 = p = 1.
=> 41= p.20+(1-p).21
= P.E. + E P.E.
*) p(20-61)=0 3. Q=61
If { = 1 conditions hold. We can say that
A & B are independent event.

(b). We need to prot P(ANB(C) = P(AIC) · P(BC) We can know P(c) = p. P(c 1 bit 0) + (1-p) · P(c 1 bit 1) = p(xot 20)+ (+p)(x,+ 6,1) P(A1C) = P(Anc) = (1-P)·(2+61)
P(C) P(B(c) = P(Bnc) = P(c) + (1-p)-(21)
P(C) P(ANBIC) = P(ANBIC) (1-4). EI , (1-P)·41 = (1-P)·(x+41) P·40+(1-P)·41
P(C) P(C) above equation, we can know that A &B are not conditionally independent given C, unless gome specific condition, such as 41=40=(1-01)

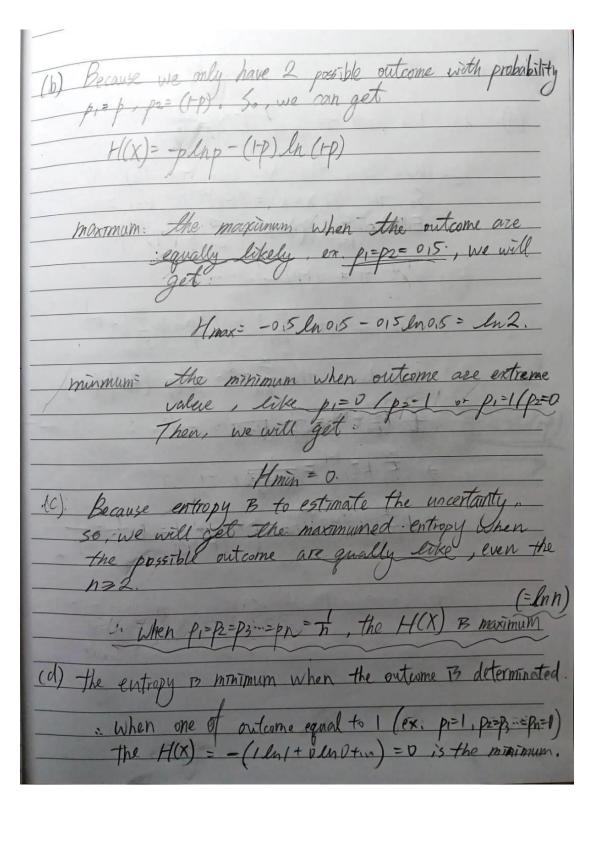




(c) We can take the length of seamonce as the trial, so we will get $P(Xs=k) = C^{n} p_{7}^{k} (+P_{7})^{nk} \text{ for } n \text{ trial}$ k = n pumber of TSince n = 123 $P(Xs=k) = C^{n} P_{7}^{k} (-P_{7}^{n})$

Problem 3.

(a) We need to proof
$$H(\frac{1}{5},\frac{1}{5}) = H(\frac{1}{5},\frac{1}{5}) + \frac{1}{2} \cdot H(\frac{1}{5},\frac{1}{5})$$
 $2H5: H(\frac{1}{7},\frac{1}{3},\frac{1}{6}) = -(\frac{1}{7}\ln\frac{1}{7} + \frac{1}{5}\ln\frac{1}{7} + \frac{1}{5}\ln\frac{1}{6})$
 $RHS: H(\frac{1}{7},\frac{1}{7}) = -(\frac{1}{7}\ln\frac{1}{7} + \frac{1}{5}\ln\frac{1}{7}) = \ln 2$
 $H(\frac{1}{3},\frac{1}{3}) = -(\frac{1}{2}\ln\frac{1}{3} + \frac{1}{3}\ln\frac{1}{3})$
 $= \ln 2 - \frac{1}{5}(\ln 2 - \ln 3) - \frac{1}{6}\ln\frac{1}{3}$
 $= \ln 2 - \frac{1}{3}(\ln 2 - \ln 3) - \frac{1}{6}\ln\frac{1}{3}$
 $= -\frac{1}{3}\ln\frac{1}{2} - \frac{1}{6}\ln\frac{1}{3} - \frac{1}{6}\ln\frac{1}{3}$
 $= -\frac{1}{3}\ln\frac{1}{3} - \frac{1}{6}\ln\frac{1}{3} - \frac{1}{6}\ln\frac{1}{3}$
 $= -\frac{1}{3}\ln\frac{1}{3} - \frac{1}{6}\ln\frac{1}{3} - \frac{1}{6}\ln\frac{1}{3}$
 $= -\frac{1}{3}\ln\frac{1}{3} - \frac{1}{6}\ln\frac{1}{3} - \frac{1}{6}\ln\frac{1}{3}$



(a). i). We need to prif $E(x) = p$. We know the $X \sim Geometric(p)$ So. the $P(X=k) = (1-p)^{k-1}p$, $k \ge 1, \ge 3$. $E(x) = \sum_{k=1}^{\infty} k P(X=k) = \sum_{k=1}^{\infty} k (1-p)^{k-1}p$. 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 $\sum_{k=1}^{\infty} t^k = (1-t)$ $\frac{d}{dt} \left(\sum_{k=1}^{\infty} t^k\right) = \frac{d}{dt} \left(1-t\right)^2$	Proble	mT.
Let $t = (I-p)$ We can get $\sum_{k=1}^{\infty} k(I-p)^{k+1} p = p \sum_{k=1}^{\infty} kt^{k+1}$ We know the $\sum_{k=1}^{\infty} t^{k} = I$ At $(\sum_{k=1}^{\infty} t^{k}) = A$ $(I-t)$ $\sum_{k=1}^{\infty} kt^{k+1} = I$ $\sum_{k=1}^{\infty} kt^{k+1} = I$ $\sum_{k=1}^{\infty} kt^{k+1} = I$ $\sum_{k=1}^{\infty} kt^{k+1} = I$	(a)	i). We need to print E(X)= ?.
Let $t = (I-p)$ We can get $\sum_{k=1}^{\infty} k(I-p)^{k+1} p = p \sum_{k=1}^{\infty} kt^{k+1}$ We know the $\sum_{k=1}^{\infty} t^{k} = I$ At $(\sum_{k=1}^{\infty} t^{k}) = A$ $(I-t)$ $\sum_{k=1}^{\infty} kt^{k+1} = I$ $\sum_{k=1}^{\infty} kt^{k+1} = I$ $\sum_{k=1}^{\infty} kt^{k+1} = I$ $\sum_{k=1}^{\infty} kt^{k+1} = I$		We know the Xn Geometric (p)
Let $t = (I-p)$ We can get $\sum_{k=1}^{\infty} k(I-p)^{k+1} p = p \sum_{k=1}^{\infty} kt^{k+1}$ We know the $\sum_{k=1}^{\infty} t^{k} = I$ At $(\sum_{k=1}^{\infty} t^{k}) = A$ $(I-t)$ $\sum_{k=1}^{\infty} kt^{k+1} = I$ $\sum_{k=1}^{\infty} kt^{k+1} = I$ $\sum_{k=1}^{\infty} kt^{k+1} = I$ $\sum_{k=1}^{\infty} kt^{k+1} = I$		So, the P(X=k)=(1-p) p., k=1, 2,3
We can get $\sum_{k=1}^{\infty} k(l-p)^k p = p \sum_{k=1}^{\infty} kt^{k-1}$ We know the $\sum_{k=1}^{\infty} t^k = \frac{1}{(l-t)}$ $\frac{d}{dt} \left(\sum_{k=1}^{\infty} t^k \right) = \frac{d}{dt} \left(\frac{1}{(l-t)} \right)$ $\frac{d}{dt} \left(\sum_{k=1}^{\infty} t^k \right) = \frac{d}{dt} \left(\frac{1}{(l-t)^2} \right)$		(P) (P)
We know the $\frac{2}{k+1}t^{k} = \frac{1}{(1-t)}$ $\frac{d}{dt}\left(\frac{2}{k+1}t^{k}\right) = \frac{d}{dt}\left(\frac{1}{(1-t)}\right)$ $\frac{2}{k+1}t^{k} = \frac{1}{(1-t)^{2}}$		Let t = (1-p)
We know the $\frac{2}{k+1}t^{k} = \frac{1}{(1-t)}$ $\frac{d}{dt}\left(\frac{2}{k+1}t^{k}\right) = \frac{d}{dt}\left(\frac{1}{(1-t)}\right)$ $\frac{2}{k+1}t^{k} = \frac{1}{(1-t)^{2}}$		we can get Exk(+P)+p=p = ktk-1
$\frac{d}{dt} \left(\frac{\mathcal{E}}{k} t^{k} \right) = \frac{d}{dt} \left(\frac{1}{(1-t)} \right)$ $\frac{\mathcal{E}}{k!} k^{2} t^{k} = \frac{1}{(1-t)^{2}}$		0
$\frac{2}{2} \left \frac{k}{k} \right ^{2} = \frac{1}{(1-t)^{2}}$	1111	
		$\frac{d}{dt} \left(\sum_{k=1}^{\infty} t^k \right) = \frac{d}{dt} \left(\frac{1}{1-t} \right)$
		2 E KT KT = 1
2 DE K (+P) = p= p= -		
KS1 1-1-11-11-11-11-11-11-11-11-11-11-11-1	THE REAL PROPERTY.	= PEX K (+P) = p. (1-(+p))= p. ================================

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

 $M_X(t) = 2e^{tR}P(X=k) = 2e^{tR}P(T=k)$
 $X = e^{tR}P(T=k) = e^{tR}P(T=k)$
 $X = e^{tR$

[b).

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

$$E[Z] = \sum_{n=1}^{\infty} (-1)^n \int_{-\infty}^{\infty} \frac{b}{(\pi n)^2},$$

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

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

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

$$2. \quad \sum_{n=1}^{\infty} Z_n^3 p(Z = Z_n) = \sum_{n=1}^{\infty} (-1)^n n^{\frac{1}{2}} \cdot \frac{b}{(\pi n)^3} = \infty$$

$$= \sum_{n=1}^{\infty} Also \text{ diverage , cause the } (n^{\frac{1}{2}}).$$

$$So, the \quad \sum_{n=1}^{\infty} Z_n^3 p(Z = Z_n) \text{ doesn't exist.}$$

$$3. \quad by \text{ the statement of } (2). \text{ We can know that }$$

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

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.