第二次討論區(MAP & Naïve Bayes Classifier) **Due 11/12 星期日中午 12:00**

11/15 為第一次大考，範圍就是第一次＆第二次討論區的內容。因此盡早做完老師才可以回應。考試時，你可以帶一張 A4 大小的紙寫任何對考試有幫忙的秘方（只能寫一面），這張紙的右上方要有名字學號，考完交回來。不符合規定者，一旦發現老師立即沒收。

學號 最後一碼	1	2	3	4	5	6	7	8	9	0
題目號碼	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**

注意事項與第一次相同。記得分數是以次數記算。最後，老師花了很多時間心力準備這份討論區題目，但難免有失漏，請同學包涵且立即提醒老師。

1. Given that today's outlook is overcast, temperature is hot, humidity is high, and it is windy, is it suitable for playing golf based on the given observations? Answer the question by Naïve Bayes Classifier where features are independent and they are

Outlook 天氣 ∈ {rainy, sunny, overcast 陰天}

Temperature ∈ {hot, mild, cool}

Humidity ∈ {high, normal}

Windy ∈ {True, False}

For example, the first row tells us it is **not** suitable for playing golf when the outlook is rainy, temperature is hot, humidity is high and it is not windy.

	OUTLOOK	TEMPERATURE	HUMIDITY	WINDY	PLAY GOLF
0	Rainy	Hot	High	False	No
1	Rainy	Hot	High	True	No
2	Overcast	Hot	High	False	Yes
3	Sunny	Mild	High	False	Yes
4	Sunny	Cool	Normal	False	Yes
5	Sunny	Cool	Normal	True	No
6	Overcast	Cool	Normal	True	Yes
7	Rainy	Mild	High	False	No
8	Rainy	Cool	Normal	False	Yes
9	Sunny	Mild	Normal	False	Yes
10	Rainy	Mild	Normal	True	Yes
11	Overcast	Mild	High	True	Yes
12	Overcast	Hot	Normal	False	Yes
13	Sunny	Mild	High	True	No

2. To decide whether a person is ill, we use four features per subject to describe an individual person. These features are N (running nose), C (coughing), R (reddened skin), and F (fever), each of which can take the value of true ('+') or false ('--'). The classes are Yes (ill) or No (healthy). Based on these training samples, decide whether Alice is ill if she has only two symptoms, running nose and reddened skin, among four features.

ID	N	C	R	F	class
1	+	+	+	--	Y
2	+	+	--	--	Y
3	--	--	+	+	Y
4	+	--	--	--	N
5	--	--	--	--	N
6	--	+	+	--	N

3. Is today suitable for playing tennis? We classify whether the day is suitable for playing tennis, given the features of the day including Outlook (rainy, sunny, overcast), Temperature (hot, mild, cool), Humidity (high, normal) and Windy (strong, weak). And the frequency tables for each feature is given below.

Outlook	rainy	sunny	over't
Play	3	2	4
No play	2	3	0

Temp.	hot	mild	cool
Play	2	4	3
No play	2	2	1

Windy	strong	weak
Play	3	6
No play	3	2

Humid.	high	normal
Play	3	6
No play	4	1

Based on the data, decide whether it is suitable for playing tennis by Naïve Bayes Classifier when the outlook is sunny, temperature is cool, humidity is high, and it has strong wind.

4. To estimate the species (M or H) of an instance, we observe the attributes of its color, number of legs, height, smelly (yes or no). And we have collected the data given below. Decide the species of a new instance which has color green, 2 legs, and it is tall and smelly.

No.	Color	Legs	Height	Smelly	Species
1	white	3	short	yes	M
2	green	2	tall	no	M
3	green	3	short	yes	M
4	white	3	short	yes	M
5	green	2	short	no	H
6	white	2	tall	no	H
7	white	2	tall	no	H
8	white	2	short	yes	H

5. Consider a football game between two rival teams, A and B. Suppose team A wins 65% of the time and team B wins the remaining matches. Among the games won by team A, only 35% of them comes from playing at teams B's football field. On the other hand, 75% of the victories for team B are obtained while playing at home. If teams B is to host the next match between the two teams, who will emerge as the winner?

hint: define X_S = the game is playing at teams S's football field, Y_S = the team S wins the match. $S \in \{A, B\}$.

6. Predict the class of a new sample (0, 1, 0) if we have training data below:

ID	A	B	C	Class
1	**0**	**0**	**0**	**+**
2	**0**	**0**	**1**	**–**
3	**0**	**1**	**1**	**–**
4	**0**	**1**	**1**	**–**
5	**0**	**0**	**1**	**+**
6	**1**	**0**	**1**	**+**
7	**1**	**0**	**1**	**–**
8	**1**	**0**	**1**	**–**
9	**1**	**1**	**1**	**+**
10	**1**	**0**	**1**	**+**

7. Estimate p by the indicated method if Y is a Geometric RV (p) and Y = 3 in one experiment

(1) by ML if the possible values of $p \in \{1/2, 1/4, 1/5\}$;

(2) by MAP if the distribution of p is f(1/2) = f(1/5) = 0.3, f(1/4) = 0.4, and f(p) = 0 otherwise.

(3) Give a discrete probability distribution for p such that the MAP result p^ would be 1/5. Justify your answer.

(4) by MAP if the distribution of p is $f(p)=2p$ if $0 \leq p \leq 1$, and f(p)=0 otherwise.

Note. Make sure that you write the likelihood function.

8. An experiment is designed as toss a coin ten times and count the number of heads. Repeat the experiment three times and observe the result as 4, 6, 7. Based on this, estimate the probability of getting H, p, by

(a) by ML

(b) by ML if we know $p \in \{0.4, 0.6\}$

(c) by MAP if we know the distribution of p is $f(p)=3p^2$ if $0 \leq p \leq 1$, and f(p)=0 otherwise.

9. We are interested to know the defective rate p of a certain product in the manufacturing process. If randomly check 10 items and found out 2 defective ones.

(a) Write likelihood function and find p^ by MLE.

(b) If we believe that the defective rate p is a Beta (a, b) RV with parameters a = 4, b = 18. Find p^ by MAP after we have seen the examination result (2 defects among 10).

10. In a study of the number of customers entering a retail store per hour, the number of customers X follows a Poisson distribution with parameter λ (Poisson PMF is $f(x)= (\lambda^x/x!)\, e^{-\lambda}$ $\lambda > 0$, and $x = 0, 1, 2, \ldots$). In a 5 hours observation we found out data as: 5, 2, 3, 4, and 6.

(a) Write the likelihood function and use MLE method to estimate λ based on the observed data.

(b) If λ is an exponential RV with a mean of 4. Use the observed data and the prior, find MAP of λ.

(Note: exponential RV $f(x) = a\, e^{-ax}$ $x \geq 0$, and $f(x) = 0$ otherwise. $E(X) = \text{mean} = 1/a$)