Midterm sample questions

Machine Learning Applications, Fall 2019

- These are sample questions for the midterm exam. The official midterm exam will include 10 questions. (it will test the content of the Module up to Week 7 included)
- The mid-term exam will last 2h
- The midterm will be on **November 4th, from 9am to 11am, room GEMS0016. You must be there at 8.40am**
- **You are only allowed to use a simple calculator for this exam, one paper notebook (with your handwritten notes) and a printed copy of any material from Sulis/Module_material (slides, etc.) you find to be useful for the mid-term exam. You cannot use your mobile phone, your laptop or any other device with internet access and you cannot use books.**
- You need to give brief and clear explanations for full credits.

Question 1

For data D and discrete variable θ, say whether or not the following equations must always be true.

a

$$\sum_{\tilde{\theta}\in\mathrm{dom}(\theta)} p(\theta = \tilde{\theta}|D) = 1 \qquad \text{is it always true?}$$

b For some $\tilde{\theta} \in \mathrm{dom}(\theta)$,

$$\sum_{d\in\mathrm{dom}(D)} p(\theta = \tilde{\theta}|D = d) = 1 \qquad \text{is it always true?}$$

c For some $\tilde{\theta} \in \mathrm{dom}(\theta)$,

$$\sum_{d\in\mathrm{dom}(D)} p(\theta = \tilde{\theta}|D = d)p(D = d) = 1 \qquad \text{is it always true?}$$

Answer

The correct answer is (a).
Explanation: a conditional probability mass function of θ given D is a probability mass function and so the sum over all elements in $\mathrm{dom}(\theta)$ must be equal to 1.

Question 2

Consider a spam filter (bag-of-words model) and denote with S a binary variable that assumes values S=1 (if the email is spam) and S=0 if the email is not-spam. For a dataset $D = \{1, 1, 1, 0, 0, 0, 0, 0\}$, where 1 means the email is Spam and 0 means not-Spam, let θ be the probability $p(S = 1)$ and assume that the observations (data) are conditional independent given θ, what is the maximum likelihood estimator of θ.

Answer

The answer is: the MLE of θ is $\hat{\theta} = \frac{3}{8}$.
Explanations: there are in total 8 emails and 3 emails are spam.

Question 3

Consider a spam filter (bag-of-words model) and denote with S a binary variable that assumes values S=1 (if the email is spam) and S=0 if the email is not-spam. For a dataset $D = \{1, 1, 1, 0, 0, 0, 0, 0\}$, where 1 means the email is Spam and 0 means not-Spam, let θ be the probability $p(S = 1)$ and assume that the observations (data) are conditional independent given θ, what is the likelihood $p(D|\theta)$?

Answer

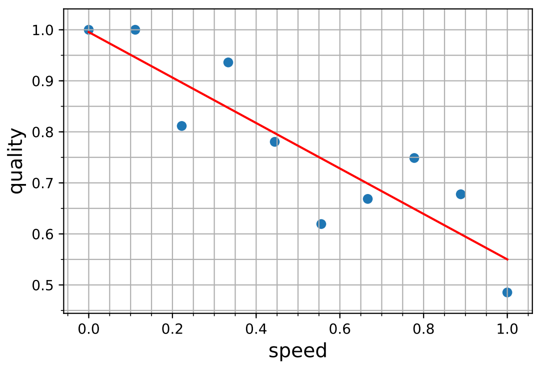
The answer is

$$p(D|\theta) = \theta^3(1 - \theta)^5$$

Explanation: since the observations (data) are conditional independent given θ, then the probability of observing 3 Spam emails is θ^3 and 5 not-spam emails is $(1 - \theta)^5$ and, therefore, $p(\{1, 1, 1, 0, 0, 0, 0, 0\}|\theta) = \theta^3(1 - \theta)^5$.

Question 4

A glass company uses a linear regression model to predict the quality of its glass as a function of the speed of the manufacturing process (input: speed, output:quality). Quality, denoted as y, is a real variable between 0 and 1, where 1 means high quality and 0 low quality. To fulfil the customer requirement, the quality y must be mandatorily greater than 0.95.
According to the prediction of the linear regression model,



what is the maximum speed that allows the Company to satisfy the Customer's requirement.

Answer

The answer is 0.1. Explanation: from the prediction (red line), the speed must be less than 0.1 to satisfy the 0.95 quality requirement.

Question 5

Assume you know the following joint distribution for three binary variables A, B, C:

A	B	C	$Prob$
0	0	0	0.168
1	0	0	0.032
0	1	0	0.252
1	1	0	0.048
0	0	1	0.072
1	0	1	0.128
0	1	1	0.108
1	1	1	0.192

What is the probability $p(B = 0, C = 0) =$?

Answer

The answer is

$$p(B = 0, C = 0) = 0.2$$

Explanation: it follows by

$$p(B = 0, C = 0) = p(A = 0, B = 0, C = 0) + p(A = 1, B = 0, C = 0)$$

Question 6

Assume you know the following joint distribution for three binary variables A, B, C:

A	B	C	$Prob$
0	0	0	0.168
1	0	0	0.032
0	1	0	0.252
1	1	0	0.048
0	0	1	0.072
1	0	1	0.128
0	1	1	0.108
1	1	1	0.192

What is the probability $p(B=0|C=0)=?$

Answer

The answer is

$$p(B=0|C=0)=\frac{p(B=0,C=0)}{p(C=0)}=\frac{0.2}{0.5}=\frac{2}{5}$$

Explanation: by Bayes' rule

$$p(B=0|C=0)=\frac{p(B=0,C=0)}{p(C=0)}$$

where

$$p(B=0,C=0)=p(A=0,B=0,C=0)+p(A=1,B=0,C=0)=0.2$$

$$p(B=1,C=0)=p(A=0,B=1,C=0)+p(A=1,B=1,C=0)=0.3$$

and

$$p(C=0)=p(B=0,C=0)+p(B=1,C=0)=0.5$$

and so

$$p(B=0|C=0)=\frac{p(B=0,C=0)}{p(C=0)}=\frac{0.2}{0.5}=\frac{2}{5}$$

Question 7

Consider two binary variables A and B and the conditional probability

$$p(A=1|B=0)=0.8$$

and the following possible values of $p(A=1|B=1)$. In what case we can say that the two variables are independent.

a

$$p(A=1|B=1)=0.8$$

b

$$p(A=1|B=1)=0.2$$

c

$$p(A=1|B=1)=0.5$$

d None of the three cases

Answer

The answer is (a): Explanation $p(A = 1|B = 0) = p(A = 1|B = 1)$ to have independence.

Question 8

Consider the following dataset

money	*spam*
1	1
0	1
0	0
1	1
0	0
1	1
1	0
1	1

where spam=1 means that the email is spam and money=1 means that the email includes the word 'money'.
The maximum likelihood estimator of $p(spam = 1)$ is $\frac{5}{8}$, what is the maximum likelihood estimator of $p(money = 1|spam = 1)$?

Answer

The answer is $\frac{4}{5}$.
Explanation: there are 5 cases where $spam = 1$ and $money = 1$ in 4/5 of these cases.

Question 9

Assume you know the following joint distribution for three binary variables $Money, Rise, Spam$:

$Money$	$Rise$	$Spam$	$Prob$
0	0	0	0.208
1	0	0	0.056
0	1	0	0.312
1	1	0	0.084
0	0	1	0.052
1	0	1	0.084
0	1	1	0.078
1	1	1	0.126

Compute the posterior probability

$$p(Spam = 1|Money = 1, Rise = 0) =?$$

Answer

The answer is

$$p(Spam = 1|Money = 1, Rise = 0) = 0.6$$

Explanation: it can be derived from Bayes' rule

$$p(Spam = 1|Money = 1, Rise = 0) = \frac{p(Spam = 1, Money = 1, Rise = 0)}{p(Money = 1, Rise = 0)}$$

with

$$p(Money = 1, Rise = 0) = p(Spam = 0, Money = 1, Rise = 0)+p(Spam = 1, Money = 1, Rise = 0)$$

Question 10

Assume you know the following joint distribution for three binary variables $Money, Rise, Spam$:

$Money$	$Rise$	$Spam$	$Prob$
0	0	0	0.208
1	0	0	0.056
0	1	0	0.312
1	1	0	0.084
0	0	1	0.052
1	0	1	0.084
0	1	1	0.078
1	1	1	0.126

Compute the marginal probability

$$p(Rise = 0) = ?$$

Answer

The answer is

$$p(Rise = 0) = 0.4$$

Explanation

$$p(Rise = 0) = \sum_{m,s \in \{0,1\}} P(Money = m, Rise = 0, Spam = s)$$

Question 11

Consider the experiment of throwing 2 fair dice.

a Find the probability that both dice show the same face.

b Find the same probability, given you know that the sum of the dice is not greater than 4.

Answer

(a) The probability that both dices show the same face is

$$p(x = 1, y = 1) + p(x = 2, y = 2) + \cdots + p(x = 6, y = 6) = \frac{1}{6}$$

(b) Let A be the event that the dice show the same face, and B the event that the sum is not greater than 4. Then B = {(1, 1), (1, 2), (1, 3)(2, 1), (2, 2), (3, 1)}, and $A \cap B$ = {(1, 1), (2, 2)}. Hence, $P(A|B) = 2/6 = 1/3$. We can also solve it by applying Bayes' rule:

$$p(x = i, y = i|x+y \leq 4) = \frac{p(x + y \leq 4|x = i, y = i)p(x = i, y = i)}{p(x + y \leq 4)} = \frac{\frac{2}{6}\frac{1}{36}}{\frac{1}{6}} = \frac{2}{36}$$

Note then

$$p(x + y \leq 4|x = i, y = i) = 1$$

for $i = 1, 2$ and zero otherwise. Therefore

$$p(x = 1, y = 1|x+y \leq 4) = \frac{p(x + y \leq 4|x = 1, y = 1)p(x = 1, y = 1)}{p(x + y \leq 4)} = \frac{\frac{1}{36}}{\frac{1}{6}} = \frac{1}{6}$$

and

$$p(x = 2, y = 2|x+y \leq 4) = \frac{p(x + y \leq 4|x = 1, y = 1)p(x = 1, y = 1)}{p(x + y \leq 4)} = \frac{\frac{1}{36}}{\frac{1}{6}} = \frac{1}{6}$$

and

$$p(x = i, y = i|x + y \leq 4) = 0 \text{ for } i \geq 3$$

Hence,

$$p(x = 1, y = 1|x + y \leq 4) + p(x = 2, y = 2|x + y \leq 4) = \frac{1}{3}$$

Question 12

Consider the experiment of throwing 2 dice. Denote with x the outcome for the first dice and y the ourcome for the second dice. Assume that the joint probability mass function of x, y is:

$$p(x = i, y = i) = 0 \text{ and } p(x = i, y = j) = \frac{1}{30}$$

for $i, j = 1, 2, 3, 4, 5, 6$ and $i \neq j$.

a Compute the marginal probability mass functions $p(x)$ and $p(y)$.

b Are the two variables x and y independent?

c By applying Bayes' rule, compute the probability that $x + y \geq 11$ given that you know that x is even.

Answer

a $p(x=1) = \sum_{\mathsf{y}=1}^{6} p(x=1, y=\mathsf{y}) = \frac{5}{30} = \frac{1}{6}$, similarly $p(x=2) = p(x=3) = \cdots = p(x=6) = 1/6$. Same for $p(y)$.

b By definition, two variables are independent if

$$p(x=\mathsf{x}, y=\mathsf{y}) = p(x=\mathsf{x})p(y=\mathsf{y})$$

for all $\mathsf{x}, \mathsf{y} = 1, 2, \ldots, 6$. In this case,

$$p(x=1, y=1) = 0 \neq p(x=1)p(y=1) = \frac{1}{36}$$

and, therefore, the variables are dependent. Note in fact that, when $x=i$ then the probability that $y=i$ is zero for $i=1,2,\ldots,6$.

c We apply Bayes' rule:

$$p(x+y \geq 11 | x \text{ is even}) = \frac{p(x \text{ is even} | x+y \geq 11) p(x+y \geq 11)}{p(x \text{ is even})}$$

where

$$p(x \text{ is even}) = \frac{1}{2} \text{ we know it from the marginal}$$

and

$$p(x \text{ is even} | x+y \geq 11) = \frac{1}{2}$$

and

$$p(x+y \geq 11) = \frac{2}{30}$$

Therefore, we have that

$$p(x+y \geq 11 | x \text{ is even}) = \frac{\frac{2}{30}\frac{1}{2}}{\frac{1}{2}} = \frac{1}{15}$$

Question 13

In a binary transmission channel, the bit 1 is transmitted with probability 2/3 and the bit 0 with probability 1/3. The channel is noisy so the conditional probability of receiving a 1 when a 1 was sent is 0.95, the conditional probability of receiving a 0 when a 0 was sent is 0.90. Given that a 1 is received, what is the probability that a 1 was transmitted?

Answer

Let B be the event that a 1 was sent, and A the event that a 1 is received. Then, $p(A = 1|B = 1) = 0.95$, and $p(A = 0|B = 0) = 0.90$. Thus, $p(A = 0|B = 1) = 0.05$ and $p(A = 1|B = 0) = 0.10$. Moreover, $p(B) = 2/3$ and $p(B = 0) = 1/3$. By Bayes' rule:

$$\begin{aligned} p(B = 1|A = 1) &= \frac{p(A = 1|B = 1)p(B = 1)}{p(A = 1|B = 0)p(B = 0) + p(A = 1|B = 1)p(B = 1)} \\ &= \frac{0.95 \cdot \frac{2}{3}}{0.10 \cdot \frac{1}{3} + 0.95 \cdot \frac{2}{3}} = 0.95 \end{aligned}$$

Question 14

In a regression problem, you have trained a system to approximate a function from x to y. What is the mean square error of the predicted output over the given test set?

test set	prediction
$x = 5, y = 9$	10
$x = 4, y = 6$	4
$x = 2, y = 5$	5
$x = 3, y = 4$	3

Answer

Answer: the MSE is equal to

$$MSE = \frac{1 + 4 + 0 + 1}{4} = \frac{6}{4}$$

Explanation: the MSE is defined as

$$MSE = \frac{1}{n} \sum_{i=1}^{N} (y_i - \hat{y}_i)^2$$

where $\hat{y}_i$ is the predicted output at the input value x_i and n is the number of observations.

Question 15

You have a training set, a test set and a machine learning algorithm (for instance Multinomial Naive Bayes or linear-regression). Answer as true/False

- zero training set error indicates good generalization performance. T/F
- a method that has higher test set error compared to its training set error has overfit to the training set T/F
- More complex models with larger number of parameters may fit the training data well, but they are more likely to overfit compared to smaller models. T/F

Answer

The correct answers are False, False, and True.
Explanation: a lookup table has training error zero, but it cannot predict any new instance (so the generalization error will be very high).

It is not necessarily true that a method that has higher test set error compared to its training set error has overfit to the training set.

We saw it in polynomial regression: more complex models are more likely to overfit.

Question 16

Consider a binary classification problem. We have two classes (C1 and C2) and we know that C1 class is a-priori more probable: probability is 0.7. That is there are more instances of class C1 than instances of class C2 (as an example you can think about the name-gender problem where the number of female names is larger than the number of male names).
In this classification problem, what would be the average accuracy if you pick a label randomly (you select C1 and C2 each with a probability of 0.5) for a given instance x?

Answer

The correct answer is 0.5.
Explanation: if the class of the instance x is C1, on average the accuracy of the random guesser will be 0.5. If the class of the instance x is C2 on average the accuracy of the random guesser will be 0.5. The average accuracy will then be

$$0.7 \cdot 0.5 + 0.3 \cdot 0.5 = 0.5$$