DS 301: Statistical and Mathematical Methods for Data Science

Midterm 1 Solutions

Fall 2019

QUESTION 1 **(Marks: 7+3)**

In a study it is found that 40% tall people are obese. An overall 20% people are tall.

a. If a person is selected at random then what is the probability that this person is **both tall and not obese**? Show formula and working.

SOLUTION

Let T = A person is tall
O = A person is obese
$\neg$ denote the not operator

$P(T) = 0.2$
$P(0) = 0.4$
$P(T, \neg O) = P(\neg O|T)P(T)$
$= (1-0.4)(0.2)$
$= 0.12$

b. Suppose that the proportion of obese people is 0.1 then can we say that height and obesity are independent? Show all working to make your decision? (A simple yes or no IS NOT ACCEPTABLE)

SOLUTION

as $P(O) = 0.1 \neq P(O|T) = 0.4$ the two random variables are not independent

QUESTION 2 **(Marks: 10)**

Suppose that 70% cold drinks get sold on a sunny day. Only 20% cold drinks get sold on a non-sunny day. The probability of a sunny day is 0.4. What is the overall fraction of cold drinks that get sold? A mathematical formula and working both are required.

SOLUTION

Let D = cold drinks are sold
S = it is a sunny day

$P(D|S) = 0.7$
$P(D|\neg S) = 0.2$
$P(S) = 0.4$

$P(D) = P(D|S)P(S)+P(D|\neg S)P(\neg S)$
$= 0.4$ (plug in all values, working is not shown)

QUESTION 3 **(Marks: 10)**

Find $P(C=1|\mathbf{x}=(1,0))$ using naive Bayes' assumption. Show all working and formulas.

x_1	1	0	1	0	1
x_2	0	0	1	1	0

C	1	1	1	0	0

SOLUTION
P(C=1|**x**=(1,0)) = P(**x**=(1,0)|C=1)P(C=1) / P(**x**=(1,0))

P(**x**=(1,0)) = P(**x**=(1,0)|C=1)P(C=1) + P(**x**=(1,0)|C=0)P(C=0) = 11/30 (working not shown. Plug in values)

Hence
P(C=1|**x**=(1,0)) = 8/11 (plug in values. Working not shown here)

QUESTION 4
What is the unbiased estimate of covariance between x and y?

x	-2	-1	3	0
y	0	-3	4	2

Answer: ____5_____

b. The correlation of a random variable z with itself is: **__1____**

c. If A and B are mutually exclusive exhaustive events then: P(A∪B) = **__1__** P(A,B) = **____0_____**

d. We have two independent variables x and y. var(x)=4 and var(y) = 9. A point in 2D space is written as (x,y). What is the Mahalanobis distance between (1,3) and (2,5)?

Answer: **_____5/6_____**

e. Given the following table of joint distribution of x and y (each cell indicates P(x,y)):

	y = 0	y = 1
x = 0	0.2	0.4
x = 1	0.1	0.3

P(x=0|y=1) = **_____4/7_____**