CMP9794M Advanced Artificial Intelligence – Workshop Week 2

Summary: In this workshop you will carry out calculations to get familiarised with the fundamentals of probabilistic reasoning. Then you will work on an implementation of the Naïve Bayes classifier, which you will apply to two different datasets to analyse probabilistic predictions.

Task 1: **Exercises on probabilistic reasoning**

Assume we have gathered the following statistics about student marks of a particular module:

Mark	1st	2:1	2:2	3(pass)	Fail
Num Students	4	10	12	5	3

a. What is the probability of getting a 1st?
b. What is the probability of getting a 2:1?
c. What is the probability of getting a 2:2?
d. What is the probability of getting a 3(pass)?
e. What is the overall probability of students passing the module (i.e. not to fail it)?

Now assume that we are only looking at whether students passed or failed this module. We have the following statistics per gender:

	pass	Fail
Male	20	2
Female	11	1

f. What is the probability of passing the module from this table? P(pass)=
g. What is the probability of being Female and passing the module? P(pass,Female)=

Given the following joint probability table:

	sunny	Rainy
hot	0.3	0.1
cold	0.1	0.5

h. Calculate the marginal probability P(sunny)=
i. Calculate the marginal probability P(hot)=
j. Calculate the conditional probability P(hot|sunny)=
k. Calculate the conditional probability P(rainy|cold)=

Given the following probability distribution

X	Y	P(X,Y)
x	y	0.2
x	¬y	0.3
¬x	y	0.4
¬x	¬y	0.1

l. Calculate $P(x \wedge y)$=
m. Calculate $P(x)$=
n. Calculate $P(x \vee y)$=
o. Calculate $P(X)$=
p. Calculate $P(Y)$=
q. Calculate $P(x|y)$=
r. Calculate $P(\neg x|y)$=
s. Calculate $P(\neg y|x)$=

Given the following probability distribution

S	T	W	Probability
summer	hot	sun	0.30
summer	hot	rain	0.05
summer	cold	sun	0.10
summer	cold	rain	0.05
winter	hot	sun	0.10
winter	hot	rain	0.05
winter	cold	sun	0.15
winter	cold	rain	0.20

t. Calculate P(sun)?
u. Calculate P(sun|winter)?
v. Calculate P(sun|winter, hot)?

Task 2: **Exercises using the Bayes rule**

a. Consider the following fictitious scientific information. Doctors find that people with the Kreuzfeld-Jacob disease (KJ) almost invariably ate hamburgers, thus P(HamburgerEater|KJ) = 0.9. The probability of an individual having KJ is rather low, about 1/100, 000. Assuming eating lots of hamburgers is rather widespread, say P(HamburgerEater) = 0.5, what is the probability that a Hamburger Eater will have the KJ disease? i.e., P(KJ|HamburgerEater)=

b. Pat goes in for a routine health check and takes some tests. One test for a rare genetic disease comes back positive. The disease (d) is potentially fatal. She asks around and learns that rare means $P(d)=1/10000$. The test (t) is very accurate $P(t|d) = 0.99$ and $P(\neg t|\neg d) = 0.95$. Pat wants to know the probability that she has the disease, i.e., calculate $P(d|t)$ =

UNIVERSITY OF LINCOLN

Task 3: **Naïve Bayes classification**

Implement—using the Python programming language—the Naïve Bayes classifier discussed in the first lecture. Download the data from blackboard to test your calculations.

a. During the actual workshop you will be provided with an example implementation of it, which you will be able to run it from the command line (or equivalent from an IDE environment) as follows:
```
python NB_Classifier_v1.py play_tennis-train.csv play_tennis-test.csv
python NB_Classifier_v1.py lung_cancer-train.csv lung_cancer-test.csv
```

b. Inspect the code and find how to run it using probabilities and log probabilities. Hint: the code has a flag that you can use to set that configuration. Did you get the probabilities and log probabilities shown in the lecture slides?

c. The implementation provided via Blackboard does not properly implement the zero estimates issue raised in lecture slide 58. It only uses a very small probability to do that. Extend this implementation using the hyperparameter $l = 1$ (called Laplacian smoothing) and implementing the corresponding equation in slide 58.