CMP9794M Advanced Artificial Intelligence – Workshop Week 1

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. For the exercises with no answers provided, please show your answers to your lecturer and/or demonstrator.

Task 1: **Revise the following calculated probabilities**

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? Answer=4/34
b. What is the probability of getting a 2:1? Answer=10/34
c. What is the probability of getting a 2:2? Answer=12/34
d. What is the probability of getting a 3(pass)? Answer=5/34
e. What is the overall probability of students passing the module (not to fail it)? Answer=31/34

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)=31/34=0.9117
g. What is the probability of being Female and passing? P(pass,Female)=11/34=0.3235

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)=P(sunny, hot)+P(sunny,cold)=0.3+0.1=0.4
i. Calculate the marginal probability P(hot)= P(hot,sunny)+P(hot,rainy)=0.3+0.1=0.4
j. Calculate the conditional probability P(hot|sunny)=P(hot,sunny)/P(sunny)=0.3/0.4=0.75
k. Calculate the conditional probability P(rainy|cold)=P(rainy,cold)/P(cold)=0.5/0.6=0.833

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. P(x ∧ y)=P(x,y)=0.2
m. P(x)=P(x,y)+P(x, ⌐y)=0.2+0.3=0.5
n. P(x ∨ y)=P(x,y)+P(x, ⌐y)+P(⌐x,y)=0.2+0.3+0.4=0.9
o. P(y)=P(x,y)+P(⌐x,y)=0.2+0.4=0.6
p. P(x|y)=P(x,y)/P(y)=0.2/0.6=0.333
q. P(¬x|y)=P(-x,y)/P(y)=0.4/0.6=0.666
r. P(¬y|x)=P(¬y,x)/P(x)=0.3/0.5=0.6

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

s. P(sun)= 0.3+0.1+0.1+0.15=0.65
t. P(sun|winter)= P(sun,winter)/P(winter)=(0.1+0.15)/(0.1+0.05+0.15+0.2)=0.25/0.5=0.5
u. P(sun|winter, hot)= P(sun,winter,hot)/P(winter,hot)=0.1/(0.1+0.05)=0.1/0.15=0.666

Given the following probability distribution

Rash	Measles	P(X,Y)
r	m	P(r,m)=0.1
r	⌐m	P(r, ⌐m)=0.8
⌐r	m	P(⌐r,m)=0.01
⌐r	⌐m	P(⌐r, ⌐m)=0.09

What is the probability of not having measles given that a person has a rash? In other words, calculate $P(⌐m|r) =$

What is the probability of having measles given that a person has a rash? In other words, calculate $P(m|r)$=

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)=

Answer:
We know the following:
P(HamburgerEater|KJ) =0.9
P(KJ)=1/100000=0.00001
P(HamburgerEater)=0.5
Applying the Bayes rule we get

$$\mathrm{P(KJ|HamburgerEater)} = \frac{\mathrm{P(HamburgerEater|KJ)} * \mathrm{P(KJ)}}{\mathrm{P(HamburgerEater)}} = \frac{0.9 * 0.00001}{0.5} = 0.000018$$

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(¬t|¬d) =0.95. Pat wants to know the probability that she has the disease.

Calculate $P(d|t) =$

Task 3: **Naïve Bayes classification**

From Blackboard, download an implementation of the Naïve Bayes classifier discussed in the first lecture. Look for file `NB_Classifier_v1.py from` in the workshop materials of this week. In addition, download the data (`data-workshop-w1.zip`) to test your calculations.

a. Run the code from the command line (or IDE environment such as Spyder) 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 run it using probabilities and log probabilities. The code has a flag (`log_probabilities` initialised to False) 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 59. 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 53.