Machine Learning & Machine Learning (extended)

Practice Exercise Sheet – Bayesian Classification

Question 1: A patient is tested by a lab test for a disease that has prevalence of 1 in 1000 in the population. The lab test has a false-positive rate of 1% and a false negative rate of 1%.

a) If the lab test result is positive, what is the probability that the patient actually has the disease?

b) Is it more probable that the patient has the disease or not?

c) Would the answers to a) and b) differ if a maximum likelihood versus maximum a posteriori hypothesis estimation method is used? Comment on your answer.

Question 2: Consider the following past examples available to train a robot cleaner to predict whether or not an office contains recycling bin.

	Status	Floor	Department	Office	Recycle Bin?
1	Faculty	4	CS	Medium	Y
2	Student	4	EE	Large	Y
3	Staff	5	CS	Medium	N
4	Student	3	EE	Small	Y
5	Staff	4	CS	Medium	N

How would a naïve Bayes classifier classify the following instance?

	Status	Floor	Department	Office	Recycle Bin?
6	Student	4	CS	Small	?

Question 3: Assume we have a data set described by the following three variables (i.e. attributes):

Hair = {B,D}, where B=blonde, D=dark.

Height = {T,S}, where T=tall, S=short.

Country = {G,P}, where G=Greenland, P=Poland.

You are given the following training data set (Hair, Height, Country):

(B,T,G)	(B,T,G)	(B,T,P)
(D,T,G)	(D,T,G)	(B,T,P)
(D,T,G)	(D,T,G)	(B,T,P)
(D,T,G)	(D,T,G)	(D,T,P)
(B,T,G)	(B,T,G)	(D,T,P)
(B,S,G)	(B,S,G)	(D,S,P)
(B,S,G)	(B,S,G)	(B,S,P)
(D,S,G)	(D,S,G)	(D,S,P)

We want to answer the following question: If you observe a new individual who is tall with blonde hair, what is the most likely country of origin?

a) Find the maximum a posteriori (MAP) estimate to the above question, using the Naïve Bayes assumption. Show all of your working.

b) Find the maximum a posteriori (MAP) estimate to the above question, without using the Naïve Bayes assumption. Show all of your working.

c) Find the Maximum Likelihood (ML) estimate to the above question, using the Naïve Bayes assumption.

d) Find the Maximum Likelihood (ML) estimate to the above question, without using the Naïve Bayes assumption.

e) Which of the above methods would you trust in each of the following situations, and why?

- a large number of training examples described by a small number of attributes
- a small number of training examples described by a large number of attributes

f) Explain how would you solve the same question if instead of blonde/dark we would have a continuous valued measurement of the hair colour, and instead of tall/short we would have the actual height in centimetres?

Question 4: We have the following past examples data available about the students' performance.

Student ID	First-class last year?	Gender	Hard-working	Drinks?	First class next year?
1	Y	M	N	Y	Y
2	Y	M	Y	N	Y
3	N	F	Y	N	Y
4	N	M	N	Y	N
5	Y	F	Y	Y	Y
6	N	M	Y	Y	N

Can we predict if students 7 & 8 will earn first-class grade next year?

Student ID	First-class last year?	Gender	Hard-working	Drinks?	First class next year?
7	N	Y	N	Y	?
8	N	N	Y	Y	?

Build a Bayesian classifier to find the answer, show all the working.