Machine Learning

Exercise Sheet on Bayesian Learning

4.1 Suppose Manchester United were about to play Arsenal in the Premiership and you assess the probability of Manchester United winning to be 0.7. You also feel that if they did win, there is a probability of 0.9 that your local pub will be packed with fans celebrating their club's victory. Alternatively, if they loose, you believe that there is still a probability of 0.6 that the pub will be packed with fans albeit for drowning their sorrows. As someone not interested in football, you enjoy a quiet day at work and on the way home, notice a big crowd in the pub. Can you assume that Manchster United won the game? If so, what is the probability of them having won the game?

4.2 A forgetful nurse is supposed to give Mr. Smith a pill each day. The probability that she will forget to give the pill on a given day is 0.3. If he recieves the pill, the probability that he will die is 0.1. If he does not recieve the pill, the probability he will die is 0.8. Mr Smith dies today. Use Baye's Rule to compute the probability that the nurse forgot to give him the pill.

4.3 Based on statistics of gold mining in the area, the probability of gold being present in the grounds of the University of Warwick is 0.1, the probability of coal is 0.3 and the probability of neither is 0.6. If gold is present, a geological test will give a positive result with probability 0.8. If only coal is present, the test will give a positive result 0.4 and if neither is present the test will still give a positive result with a probability of 0.2. Given a positive result, what is the probability of Gold being found on campus? You may assume that the presence of one mineral makes it impossible for the other mineral to occur.

4.4 A screening test for Meningitis is known to provide a positive result 95% of the time when a patient with Meningitis is tested, while it gives a negative result for 70% of patients tested who are not suffering from the disease. National statistics suggest that 5% of the population get the disease.

1. Given that a patient has tested positive, use Bayes Rule to decide whether the patient has Meningitis or not.
2. If the doctor orders a second test that returns a negative result, how would this affect the probabilities associated with Meningitis and $\neg$Meningitis?

4.5 The prior probability of it raining in the UK is 0.8. Mr. Jones, the new weather man at the BBC, has devised a new set of experiments that allow him to predict whether it will rain on the following day or not, given the current weather. The experiment is fairly accurate, returing a positive result 75% of the time when it actually rains the next day and only 15% of the time does it return a positive result when it doesn't actually rain the next day.

Given a negative result today, should Mr. Jones predict no rain for the following day? Show your calculations to support your answer.

4.6 Suppose the probability of contracting a rare tropical disease Chickungunya on a holiday to India is 0.0001. Fred, on his return to the United Kingdom from a Jungle Safari in India, feels unwell and decides to see his doctor. Fred complains to his doctor of *joint pain*. The doctor had heard that there had been some cases of Chickungunya reported in India and also knew that 64% of people with Chickungunya complained of joint pain. There are other causes of Joint Pain however and the doctor estimates that the probability of *joint pain* in the absence of Chickungunya is 0.6. The doctor ordered a very reliable test used to diagnose Chickungunya that returned a *positive* result 99% of the time when the patient had Chickungunya and only returned a *positive* result 4% of the time when the patient tested was not suffering from the disease. The test result returned for Fred was *positive*.

1. Based on the evidence to date can the doctor conclude that Fred has Chickungunya?
2. Realising that there are no other symptoms of the disease or any other more accurate tests available, the doctor decides to ask for the test to be repeated. Once again a positive test result is obtained. How, if at all, will this affect his belief in the diagnosis of Chickungunya?
3. Can the strategy of repeating the same test over and over again improve his belief in a positive diagnosis of Chickengunya? If so, what is the minimum number of times that the test must be carried out, assuming a positive result each time, for the belief in Chickungunya to be greater than the belief in *not Chickungunya*?

4.7
1. Given the data below, using the Naïve Bayes algorithm, compute the probability of a customer, with Handset = Old, Time Since Customer (in Years) >2.5 and Age≤ 55, churning.

Handset	Time Since Customer (in Years)	Age (in Years)	Churned
New	>2.5	≤ 55	Yes
New	>2.5	>55	No
New	[1,2.5]	>55	No
New	<1	>55	Yes
New	<1	≤ 55	Yes
Old	[1,2.5]	>55	Yes
New	[1,2.5]	≤ 55	No
Old	>2.5	>55	Yes
New	>2.5	≤ 55	No
New	>2.5	≤ 55	No

2. Given that the data set being used to estimate the underlying probability distribution is very small, assume a uniform distribution for each of the input attributes, conditioned on the class label and compute the m-estimate for the various probabilities. What effect does the size of m have on the probability of churning for the customer above?

4.8 You are given the data below

x	y	Classification
5	6	+
5	7	+
3	6	+
4	5	+
1	7	-
3	4	-
2	6	-
8	1	-

Table 1: Training Data Set

Using the Naïve Bayes algorithm, compute the probability of a customer, with X = 7 and Y = 4 being a positive example. Assume for the purposes of this exercise, that the attributes X and Y are continuous and that their underlying denisty function is Gaussian given the class label. Note that the Equation for a

Gaussian density function is:

$$N(x;\mu,\sigma) = \frac{1}{\sqrt{2\pi}\sigma} e^{\frac{-(x-\mu)^2}{2\sigma^2}}$$

4.9 Given the following data:

2.8, 4.2, 5.3, 5.5, 2.1

1. What is the log-likelihood of data set having been generated by a mixture of two Gaussian distributions with means 2.3 and 6.8 and standard deviations 1.8 and 2.2 respectively. Assume that there is an equal prior probability for a data point being generated from either of the two components in the mixture.

   Note that the Equation for a Gaussian density function is:

$$N(x;\mu,\sigma) = \frac{1}{\sqrt{2\pi}\sigma} e^{\frac{-(x-\mu)^2}{2\sigma^2}}$$

2. Can you suggest an improved estimate for the component densities given the data? If so, show how you arrived at the estimate.