COMP30120 Tutorial

Decision Trees and Naïve Bayes Classifiers

1. Consider the following dataset S, which contains observations of several cases of sunburn:

	Name	Hair	Height	Weight	Dublin	Result
1	Sarah	blonde	average	light	no	sunburned
2	Dana	blonde	tall	average	yes	none
3	Alex	brown	short	average	yes	none
4	Annie	blonde	short	average	no	sunburned
5	Emily	red	average	heavy	no	sunburned
6	Pete	brown	tall	heavy	no	none
7	John	brown	average	heavy	no	none
8	Katie	brown	short	light	yes	none

a) What is the entropy of this dataset with respect to the target class label *Result*?

b) Construct the decision tree that would be built with Information Gain for this dataset. Show your work for selection of the root feature in your tree.

c) Using your decision tree from (b), how would you classify the following example X:

Hair=blonde, Height=average, Weight=heavy, Dublin=no

d) Use Naïve Bayes to give the likelihood that the result for the given example X is “sunburned”. Then indicate what prediction Naïve Bayes would make. Provide the probability table for each of the predicting features.

2. Consider the following dataset that helps to predict the *Risk* of a loan application based on the applicant's *Credit History*, *Debt* and *Income*.

	Credit History	**Debt**	**Income**	**Risk**
1	bad	low	0to15	high
2	bad	high	15to35	high
3	bad	low	0to15	high
4	unknown	high	15to35	high
5	unknown	high	0to15	high
6	good	high	0to15	high
7	bad	low	over35	moderate
8	unknown	low	15to35	moderate
9	good	high	15to35	moderate
10	unknown	low	over35	low
11	unknown	low	over35	low
12	good	low	over35	low
13	good	high	over35	low
14	good	high	over35	low

a) Which one of the predicting features would be selected by ID3 at the root of a decision tree? Explain your answer. Show all the steps of the calculations.

b) What is the main problem with the Information Gain criterion for feature selection in decision trees?

c) Provide the contingency table of conditional and prior probabilities that would be used by Naïve Bayes to build a classifier for this dataset.

d) Write down the rule used by Naïve Bayes to classify examples, and apply it to the following example. Which class will be returned by Naïve Bayes for the following example X?

CH=bad, Debt=low, Income=15to35