COMP47460 Tutorial

Naïve Bayes Classifiers

1. Consider the dataset below, which contains 10 examples describing weather conditions. The goal is to predict whether or not an individual will go swimming based on these conditions.

	Rain Recently (RR)	**Rain Today (RT)**	**Temp (T)**	**Wind (W)**	**Sunshine (S)**	**Swimming**
1	Moderate	Moderate	Warm	Light	Some	Yes
2	Light	Moderate	Warm	Moderate	None	No
3	Moderate	Moderate	Cold	Gale	None	No
4	Moderate	Moderate	Warm	Light	None	Yes
5	Moderate	Light	Cold	Light	Some	No
6	Heavy	Light	Cold	Moderate	Some	Yes
7	Light	Light	Cold	Moderate	Some	No
8	Moderate	Moderate	Cold	Gale	Some	No
9	Heavy	Heavy	Warm	Moderate	None	Yes
10	Light	Light	Cold	Light	Some	No

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

b) Based on the contingency table, classify the two new examples below using Naïve Bayes.

	Rain Recently (RR)	**Rain Today (RT)**	**Temp (T)**	**Wind (W)**	**Sunshine (S)**	**Swimming**
X1	Heavy	Moderate	Warm	Light	Some	???
X2	Light	Moderate	Warm	Light	Some	???

2. Consider the following dataset, which contains examples describing several cases of sunburn.

	Name	**Hair**	**Height**	**Build**	**Lotion**	**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) Provide the contingency table of conditional and prior probabilities that would be used by Naïve Bayes to build a classifier for this dataset.

b) 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.

	Hair	**Height**	**Build**	**Lotion**	**Result**
X	blonde	average	heavy	no	???

3. Consider the following dataset that aims to predict the risk of a loan application based on 3 features describing each applicant: credit history, debt, and income. Applications are assigned to 3 different risk classes: low, medium, high.

	Credit History	**Debt**	**Income**	**Risk**
1	bad	low	0to30	high
2	bad	high	30to60	high
3	bad	low	0to30	high
4	unknown	high	30to60	high
5	unknown	high	0to30	high
6	good	high	0to30	high
7	bad	low	over60	medium
8	unknown	low	30to60	medium
9	good	high	30to60	medium
10	unknown	low	over60	low
11	unknown	low	over60	low
12	good	low	over60	low
13	good	high	over60	low
14	good	high	over60	low

a) Calculate the contingency table that would be used by Naïve Bayes to build a classifier using this training data.

b) Based on the contingency table, predict a risk level for the new loan application X below.

	Credit History	**Debt**	**Income**	**Risk**
X	bad	low	30to60	???