MACHINE LEARNING FOUNDATIONS AND APPLICATIONS

Assignment 1 Due Date: 15 October, 2020

Total Marks: 50

[Instructions: Please show all the steps and calculations with proper explanation. Numerical accuracy is less important than methodological accuracy. You can submit through Moodle. You can work out on paper and submit scanned copies, or you can directly work out using LATEX. The submission filename should be <roll number>.pdf. Copying will result in 0 marks.]

1) Consider a 3-class classification problem with 3 features: 2 of which are binary and 1 discrete with 4 values. 200 training examples are provided. In the table below, the statistics of these examples are provided. Using them, construct a decision tree of depth 3, i.e. choose two features for splitting sequentially. Use the decision tree to compute the accuracy on the training set, i.e. how many of the 200 training examples are classified correctly. **[10 + 5 = 15]**

X1	X2	X3		#(Y=1)	#(Y=2)	#(Y=3)
1	1	A		15	0	0
1	2	A		15	0	0
2	2	A		2	9	1
2	1	A		3	5	0
1	1	B		0	10	4
1	2	B		0	10	1
2	2	B		8	2	4
2	1	B		7	3	1
1	1	C		0	6	0
1	2	C		0	9	0
2	2	C		1	0	14
2	1	C		0	0	20
1	1	D		0	2	15
1	2	D		1	3	14
2	2	D		1	0	9
2	1	D		0	0	5

2) Suppose you have been given only part of the above table, other rows are missing (X). Use a Naïve Bayes Classifier to make class label predictions for the feature values along those rows. Indicate the corresponding confidence values. **[10]**

X1	X2	X3		#(Y=1)	#(Y=2)	#(Y=3)
1	1	A		15	0	0
1	2	A		15	0	0
2	2	A		2	9	1
2	1	A		X	X	X
1	1	B		0	10	4
1	2	B		0	10	1
2	2	B		8	2	4
2	1	B		X	X	X
1	1	C		X	X	X
1	2	C		0	9	0
2	2	C		1	0	14
2	1	C		0	0	20
1	1	D		0	2	15
1	2	D		1	3	14
2	2	D		1	0	9
2	1	D		X	X	X

3) Given 20 training examples with 2 features – both continuous, construct the best decision stump (single split) with only 2 leaf nodes. Choose the feature and also the threshold. Plotting the points on 2D plane may help you. **[5]**

ID	X1	X2	Y		ID	X1	X2	Y
1	5	7	1		11	13	6	2
2	7	12	1		12	14	8	2
3	12	5	1		13	17	15	2
4	10	8	1		14	15	9	2
5	6	11	1		15	13	10	2
6	13	8	1		16	11	5	2
7	8	12	1		17	16	18	2
8	9	11	1		18	15	7	2
9	11	6	1		19	12	12	2
10	8	12	1		20	18	9	2

4) Use the same examples as above to construct a Naïve Bayesian Classifier where the Class-conditional distributions are Normal (for each feature). Estimate the parameters of the normal distributions from the data (sample mean and sample variance). For which feature values is your NBC least confident? **[5 + 5 = 10]**

5) (i) You are given a training set with N data-points of the form (xi, yi, wi), where xi is a D-dimensional vector, the label yi is real-valued, and wi denotes the weight. You are now required to fit a linear regression model to this data, of the form y=a'x + b as usual. However, fitting errors are now weight-dependent, i.e. the loss should be more for points with higher

weightage. Find the linear regression model in this situation, starting from the objective function and showing the necessary steps. **[5 marks]**

(ii) Suppose you are given N datapoints (xi, yi) where xi is a D-dimensional vector and the label yi is real-valued. One again you are required to fit a linear regression model to this data, but the vector "a" should be as close to a given vector "v" as possible in terms of Euclidean distance. Find the linear regression model in this situation, starting from the objective function and showing the necessary steps. **[5 marks]**