CS 5805 **Homework 3**

Question 1 (Comparing Classifiers: kNN, Decision Trees, ANN) [5 points]

A realtor is studying housing values in Blacksburg and has given you a dataset with the following attributes: crime rate in the neighborhood, proximity to Virginia Tech, number of rooms per dwelling, age of unit, distance to elementary schools, and distance to restaurants. The target variable is the cost of the house (with two values: high and low). Given this scenario, indicate the choice of classifier for each of the following questions and give a brief explanation.

a) If the realtor wants a model that not only performs well but is also easy to interpret, which one would you choose between kNN, Decision Trees and ANN?

Decision Trees, because it handles different types of data without needing to scale and its clear boundaries are easy to understand

b) If you had to choose between ANN and Decision Trees, which one would you prefer for a classification problem where there are missing values in the training data?

Decision Trees; they easily handle missing values while ANN does not do that automatically

Question 2 (Comparing Classifiers: Logistic, DT, kNN) [9 points]

Figures A, B and C below show different scenarios with two target classes (represented as red circles and blue triangles). Please determine which of the following classifiers can work best on each of the figures: (1) Logistic regression model, (2) 1-Nearest Neighbor, and (3) Decision Tree with at most three internal nodes. Give a very brief justification for each case in the table below.

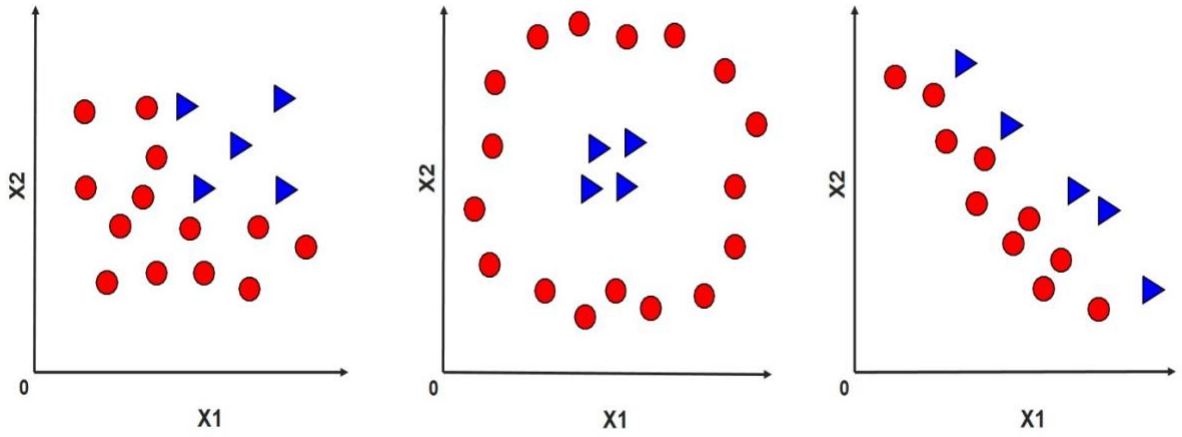


Figure A Figure B Figure C

Figure A	Decision tree; it can easily find the clear linear boundaries between red and blue
Figure B	1-nearest neighbor; boundary is non-linear so other two may not be suitable while 1NN would be 100% accurate on the set
Figure C	Logistic Regression; there is a clear linear boundary that logistic regression can easily identify

Question 3 (Comparing Classifiers) [12 points]

In each of the classification scenarios listed below, you are given a set of classifiers and a description of the classification scenario. For each scenario, state the choice of one classifier that is best suited and another one that is worst suited. Give a brief explanation to support your answer.

a) **Scenario:** The data has mixed attributes (some are continuous, some are categorical) and some attributes are redundant (duplicate copies of other attributes).
Classifiers for comparison: kNN, Decision Trees, Naïve Bayes, multi-layer ANN

Decision trees are best suited as they can handle both mixed attributes and missing attributes. The worst suited is kNN because it relies on distance, which requires a lot of scaling or normalization because of the mixed types.

b) **Scenario:** Attributes in the data are not discriminative by themselves (when considered individually) but are discriminative when used in non-linear combinations.
Classifiers for comparison: Naïve Bayes, linear SVM, logistic regression, kNN

kNN would be the best because it is able to capture nonlinear relationships. Naive Bayes would be the worst because it is unable to handle nonlinear relationships and assumes conditional independence, which is not the case here.

Question 4 (ANN) [6 points]

State whether the following statements are true/false, giving a one-line justification for your answer.

a) A multi-layer ANN can learn non-linear decision boundaries even if the activation function chosen at every layer is linear.
True. A single layer can only capture a linear boundary, but the use of multiple layers allows the ANN to capture nonlinear boundaries

b) In the back-propagation algorithm for training ANN models, to compute the gradients of loss w.r.t. weights at the $k+1^{th}$ layer, we need to first compute the gradients of loss w.r.t. weights at the k^{th} layer.
False. In back propagation, the gradients at layer k are calculated based on layer k+1

c) While applying an ANN model on a test instance, to compute the activations at nodes at the $k+1^{th}$ layer, we need to first compute the activations at nodes at the k^{th} layer.
True. The activations at layer k+1 are a function of the weights and the activations at layer k.

Question 5 (ANN and Logistic Regression) [4 points]

State two similarities and two differences between perceptron and logistic regression.

The first similarity between the two is that they are both limited to capturing linear relationships and boundaries. They are also similar in that they are used to predict if the output is either 0 or 1 given an input. The first difference is that perceptron uses a sign activation function while logistic regression uses a sigmoid/logistic function. Another difference is that perceptron minimizes error while logistic regression minimizes entropy.

Question 6 (Decision Trees) [4 points]

Imagine you are given the task to predict the educational qualification of each person using their demographic data with the following attributes: (1) Annual Income (real-valued), (2) Income Tax filed (real-valued), (3) Age (integer), (4) State of residence in the US (categorical), (5) Gender (categorical), (6) House Owner or not (Boolean) and (7) Height (in inches). Assume the target classes are (a) college degree and (b) without a college degree. State one strength and one weakness of decision trees for this task?

A strength is that they are able to handle the different types of attributes in the data. A weakness is that the tree could be overfitted as it would be unable to distinguish a relevant attribute(like income) from something irrelevant(height) if there is a random correlation between the irrelevant attribute and education

Practice Questions

Question 7 (Naïve Bayes)

Consider the data set shown below.

Instance	A	B	C	Class
1	0	0	1	−
2	1	0	1	+
3	0	1	0	−
4	1	0	0	−
5	1	0	1	+
6	0	0	1	+
7	1	1	0	−
8	0	0	0	−
9	0	1	0	+
10	1	1	1	+

(a) Estimate the conditional probabilities for **P**(**A** = 1|+), **P**(**B** = 1|+), **P**(**C** = 1|+), **P**(**A** = 1|−), **P**(**B** = 1|−), and **P**(**C** = 1|−).

(b) Use the conditional probabilities in part (a) to predict the class label for a test sample (**A** = 1**,B** = 1**,C** = 1) using the Naïve Bayes approach.

(c) Compare **P**(**A** = 1), **P**(**B** = 1), and **P**(**A** = 1**,B** = 1). State the relationships between **A** and **B**.

(d) Repeat the analysis in part (c) using **P**(**A** = 1), **P**(**B** = 0), and **P**(**A** = 1**,B** = 0).

(e) Compare **P**(**A** = 1**,B** = 1|**Class** = +) against **P**(**A** = 1|**Class** = +) and **P**(**B** = 1|**Class** = +). Are the variables conditionally independent given the class?

Question 8 (Naïve Bayes)

Instance	Distinguishing Attributes				Noise Attributes								Class Label	
	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10	X11	X12		
1	■	■					■						A1	Class A
2	■	■			■							■		
3	■	■								■				
4	■	■						■						
5			■	■							■		A2	
6			■	■	■		■							
7			■	■								■		
8			■	■					■					
9	■		■			■							B1	Class B
10	■		■				■							
11	■		■							■				
12	■		■			■						■		
13		■		■				■					B2	
14		■		■					■					
15		■		■	■						■			
16		■		■			■							

(a) Explain how Naïve Bayes performs on the data set above when we only consider two classes: A and B.

(b) If each class is further divided such that there are four classes (**A**1, **A**2, **B**1, and **B**2), will naive Bayes perform better?

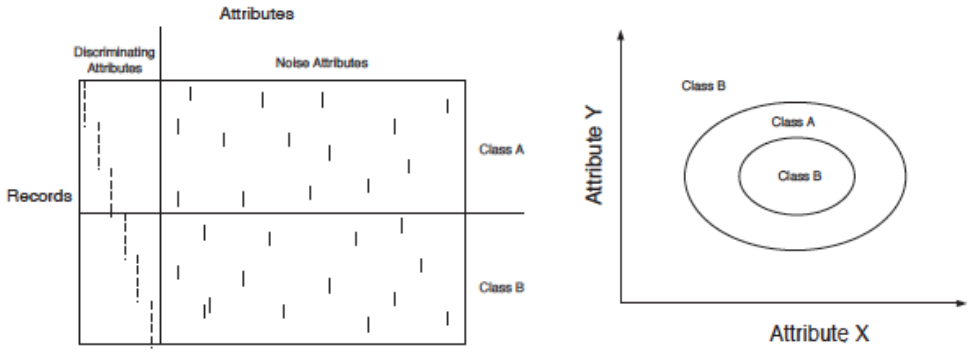
(c) How will a decision tree perform on this data set (for the two-class problem)? What if there are four classes?

Question 9 (Comparing classifiers: DT, kNN)

Suppose you have a dataset with many different types of variables (categorical, continuous, binary, etc.) and many missing values and a number of redundant attributes. Also, the data contains a large number of attributes out of which only some are discriminative. You can select one of the following classifiers to run on this dataset: decision tree and kNN. Which one will you choose and why? Also explain why the other two wouldn't work as well.

Question 10 (Comparing Classifiers: DT, NB and kNN)

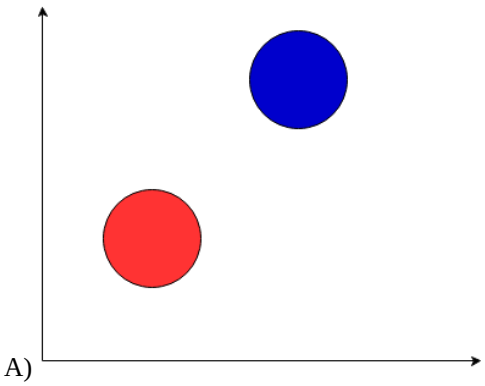
Given the data sets shown in figure below, explain how the decision tree, Naïve Bayes (NB), and k-nearest neighbor (k-NN) classifiers would perform on these two data sets.



(a) Synthetic data set 1. **(b) Synthetic data set 2.**

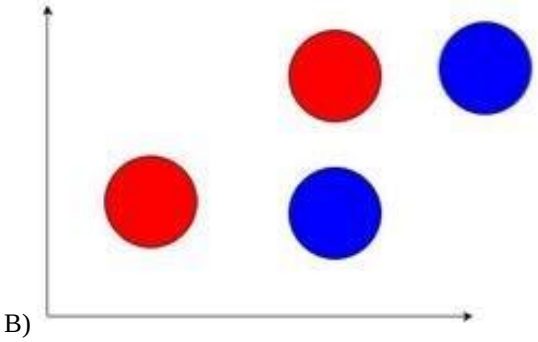
Question 11 (Naïve Bayes)

For any two-dimensional dataset, we can use visual inspection to determine whether Naïve Bayes can correctly estimate the class labels at different points, thereby assessing the suitability of naïve Bayes for classification on the given dataset. Determine whether the Naïve Bayes classifier would be suitable for each of the datasets shown in the figures below. The two axes in each of the figures represent the two attributes, and the target binary classes are red and blue. The red and blue patches are of the equal density of points and there are an equal number of red and blue points in all subfigures, and hence the prior probabilities of both classes is 0.5.

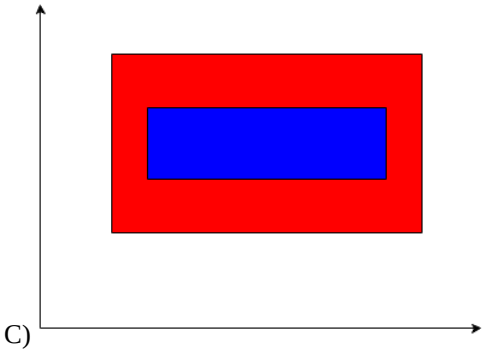


A)

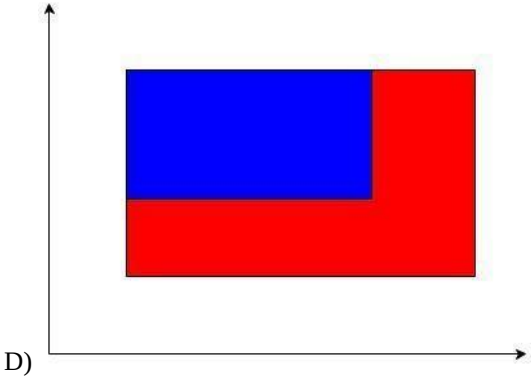
B)



C)



D)



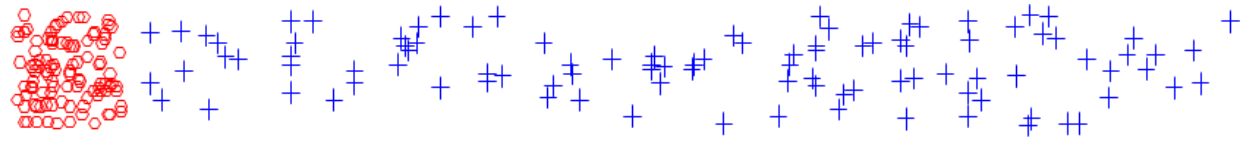
Question 12 (Comparing Classifiers)

a) A realtor is studying housing values in the suburbs of Boston, and has given you a dataset with the following attributes: crime rate, proximity to Charles River, number of rooms per dwelling, age of unit, distance to five Boston employment centers, pupil-teacher ratio by town, house value (the target variable with values *high* and *low*). The realtor would like you to build a classification model that not only performs well, but is also easy to interpret. Between ANN and decision tree classifiers, which one would you pick. Indicate your choice of classifier and briefly explain why the other is not a good choice.

b) Consider the problem of predicting whether a person is a good credit risk given the following attributes: hair color, income, weight, time in current job, marital status, height, age, and birth month. If you had to choose between a Bayesian classifier and logistic regression, which would you prefer? Indicate your choice of classifier and briefly explain why the other one may fail?

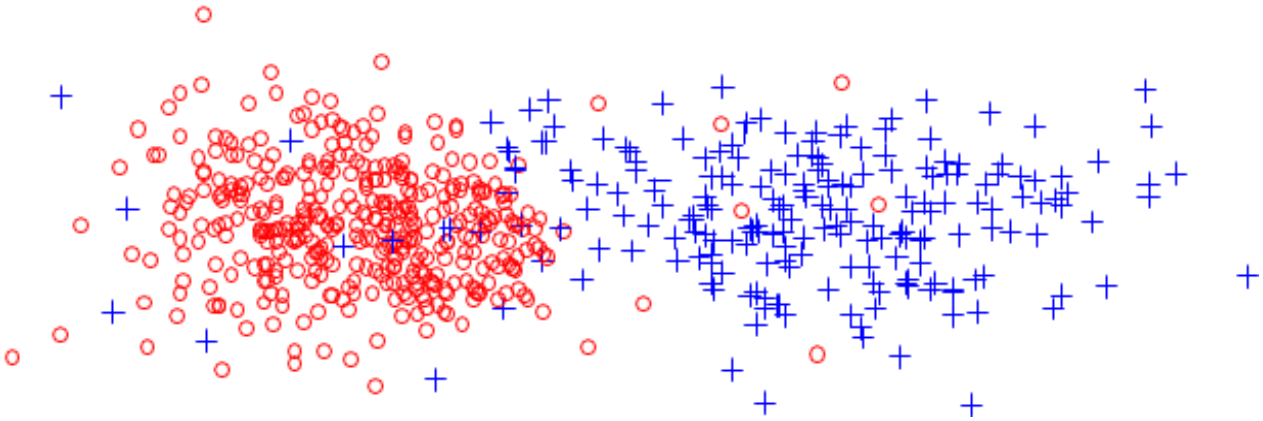
Questions 13 (Comparing Classifiers)

For each of the two given scenarios, make a right choice of K for the KNN classifier in order to obtain better performance with a brief explanation.



100 instances 100 instances

(a) K = 1 or K=5 or K = 50?



(b) K = 1 or K = 5 or K=50?

Question 14 (Comparing Classifiers: DT, Naïve Bayes, kNN, SVM)

Consider the following classification methods: Decision Trees, Support Vector Machines (SVM), Naïve Bayes, k-Nearest Neighbor (kNN).

In the following scenarios, choose which of the above classifiers is best suited or worst suited, depending on the question. Give a brief explanation.

a) Scenario: The number of attributes is very large, all attributes are numeric, but only a subset of them are relevant to discriminate between the classes.
Classifier to Choose: **Worst suited.**

b) Scenario: Computation time for training is to be minimized.
Classifier to Choose: **Best suited.**

c) Scenario: Data contains missing values during testing.
Classifier to Choose: **Best suited.**