ML Final 2020 Answer (V2.0)

By Dr\ Hanaa Bayoumi

Solved by **Ahmed Sallam** if you find any mistakes, please contact me.

لا تنسونا من صالح دعائكم

Question 1:

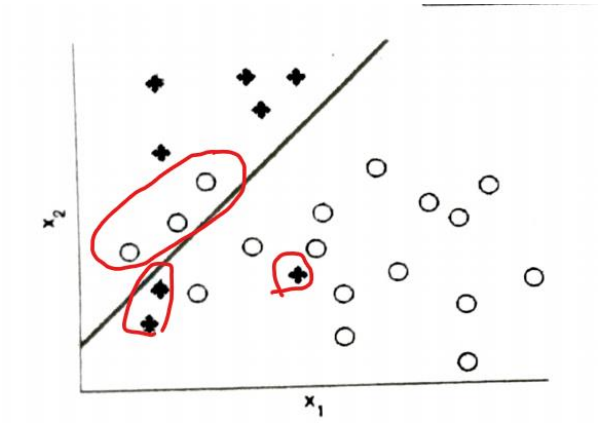
You will find Mistakes in red.

Question 1: [6 marks]

The following figures depict decision boundaries of classifiers obtained from three learning algorithms: decision trees, logistic regression, and nearest neighbor classification (in some order). Under each of the three plots, write the name of the learning algorithm and the number of mistakes it makes on the training data.

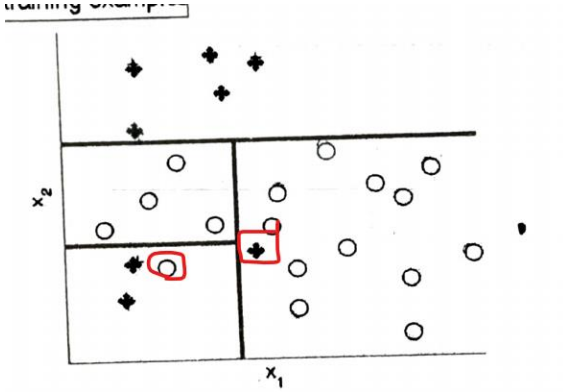
+	positive training examples
○	negative training examples

بيقولك معاك كذا plot ل decision boundaries of classifiers ودول بيمثلو logistic regression و KNN و Decision tree اكتب جمب كل plot اسم ال algorithm و كام عدد ال data الي ال model عملها wrong classification



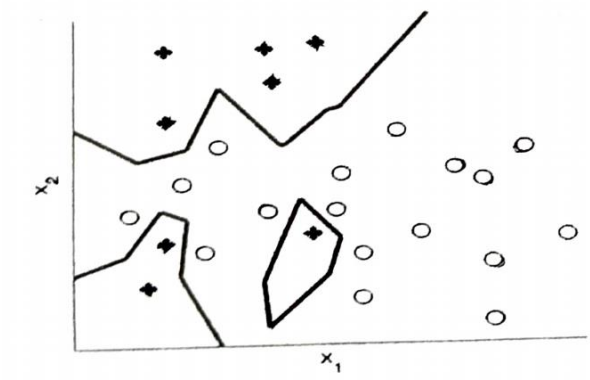
Logistic regression

Number of mistakes = 6



Decision tree

Number of mistakes = 2



KNN

No Mistake

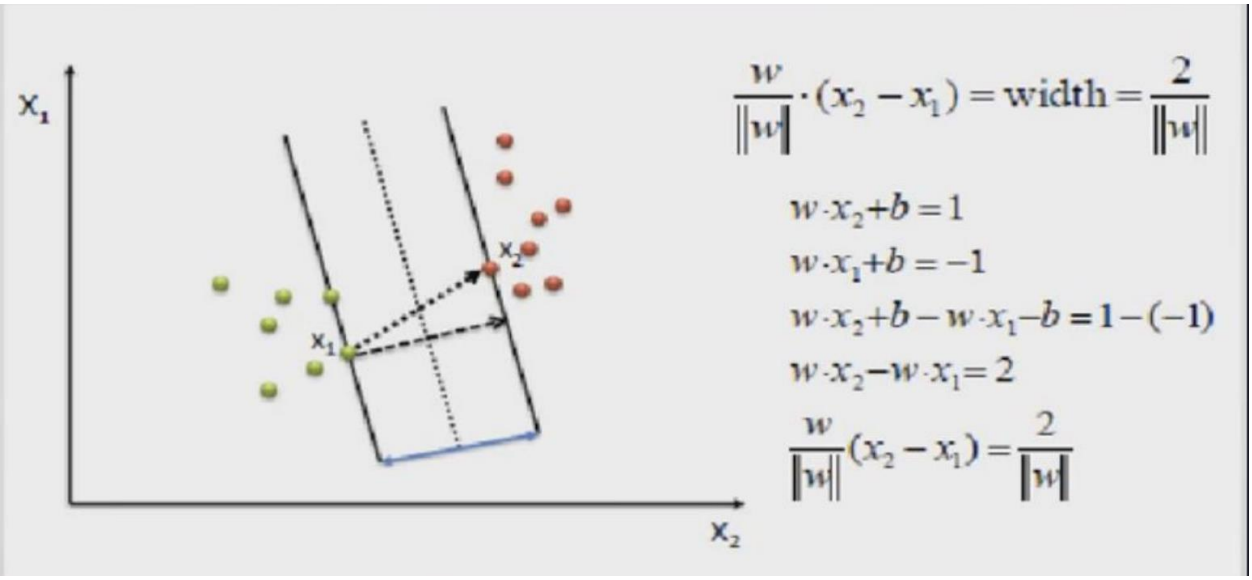
Question 2: CNN is out of course scope.

Question 3: in SVM show how to express the margin in terms of weight (W)

ببساطه اكتب ال proof ل margin بالنسبه ل Weight (W)

• The margin is the distance between the decision boundary (hyperplane) and the nearest data point from either class. The margin is typically expressed in terms of the weight vector (W) and bias term (b) of the hyperplane.

• Hyperplane = $\bar{W} \bullet \bar{X} + b = 0$



خلي بالك لازم ترسم ال Hyperplane وتحدد ال x1,x2

Question 4:

1) Give One Sentence Reason

a) we might prefer Decision Tree learning over Logistic Regression for a particular learning task.

1) when the relationship between input features and the target variable is nonlinear and involves complex interactions, as Decision Trees can capture such nonlinearity and interactions more effectively.

2)If we want our model to produce results and rules easily interpreted by humans.

١- لما العلاقه بين ال features و ال target تبقي nonlinear وفيها تعقيدات كثير ال DT هتكون أكثر فعاليه في التعامل مع ال data دي

٢- لو احنا عايزين ال model يطلع نتائج وقواعد سهل الانسان الطبيعي يفهمها.

b) we might prefer Logistic Regression over Naive Bayes for a particular learning task.

• If we have small training data Logistics Regression will be better, and if we know that the conditional independence assumptions made by Naive Bayes are not true for our problem.

١ – لو عندنا data صغيرة ال Logistic Regression هيتعامل معاه احسن

٢- conditional independence ممكن ال Naïve يطلعها غلط فهنلجئ ل Logistic Regression

c) we need re-estimate probabilities (smoothing) in Naive Bayes classifier.

• to prevent zero probabilities and improve the model's Accuracy by avoiding situations where a feature in the test data was not observed in the training data, leading to a probability of zero and causing the entire classification to be based on that feature.

- بنحتاج نعمل re-estimate probabilities (smoothing) عشان ننمنع أي احتمال انه يطلع بصفر ونحسن ال models accuracy عشان نتنجب موقف ان تكون فيه feature موجوده في test data ومش موجوده في train data وده هيسبب ظهور احتمالات صفرية

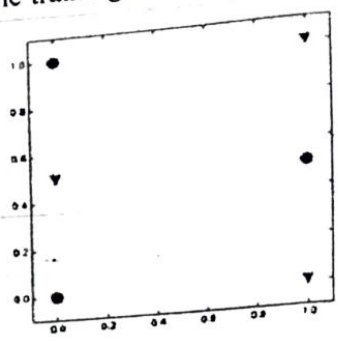
2) You trained a binary classifier model, which gives very high accuracy on the training data, but much lower accuracy on validation data. Which of the following statement(s) are true?

This is an instance of overfitting.	(T)
This is an instance of under fitting.	(F)
The training and testing examples are sampled from different distributions.	(T)

3) CNN is out of course scope.

4) Check all the binary classifiers that are able to correctly separate the training data (circles vs. triangles) given in the following Figure.

Logistic Regression	(X)
SVM with kerne	(T)
Decision tree	(T)
3-nearest-neighbor classifier (with Euclidean distance).	(X)



- Logistic regression: logistic regression has linear decision boundaries, so it may not be able to correctly separate the training data because the data is not linearly separable.

- SVM with kernel: it can be a non-linear decision boundary. It can capture more complex relationships in this data, making it suitable for the classes that are not linearly separable.

- Decision tree: can be model complex decision boundaries by making splits along different features. They can accommodate non-linear relationships in the data.

- 3-nearest-neighbor classifier: may not be able to correctly separate this training data because it relies on the local neighborhood of points. In this case where different classes are mixed closely, it might make errors in classification.

5) Describe the difference between parametric methods and nonparametric methods.

Parametric methods make assumptions about the functional form of the underlying data distribution and have a fixed number of parameters, while nonparametric methods do not make explicit assumptions about the distribution and can adapt to more complex patterns without a predetermined number of parameters.

6) What is the similarity and difference between feature selection and dimensionality reduction?

Feature selection involves choosing a subset of relevant features from the original feature set, while dimensionality reduction aims to transform the data into a lower-dimensional space, preserving essential information by combining or projecting the original features.

7) In the K-nearest neighbor classifier, which of the following statement(s) are true?

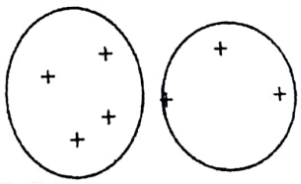
a) A KNN is supervised classifier	(T)
b) The hyper parameter K in KNN is typically set to an odd number	(T)
c) When K is set to an extremely large number, it is more likely that the classifier will overfit than underfit.	(F)
d) Both KNN and K-means are unsupervised learning techniques	(F)
e) Increase k in a k-nearest neighbor classifier increase bias	(T)

The k, represent the number of closest neighbors that you are comparing, right? So, no matter if you have 2 or n classes, if you choose an even k, there is a risk of a tie in the decision of which class you should set a new instance. This is why the k is usually odd

if the value of k is too high, then it can underfit the data

if k value is too small then it can overfit the data

8) Are the two clusters shown below well separated? Circle an answer: Yes **No**
Now in one or two sentences justify your answer.



It is not well separated because some points in each cluster are closer to points in another cluster than to points in the same cluster.

9) For the methods below, indicate whether the method is parametric or nonparametric using a "P" for parametric and a "N" for nonparametric:

Method	P/N
Linear Regression	parametric
k-Nearest Neighbor	non-parametric
Support Vector Machines	
Multivariate Linear Regression	parametric
Logistic Regression	parametric
Perceptron	parametric
Multilayer Feed-Forward Neural Network	parametric
K-Means	non-parametric

Linear SVM = parametric
Non-Linear SVM = non-parametric

Notic:

Parametric model:

- Model fit the data exactly

- this models have a parameters that model are try to find and calculate them exactly

Like Linear Regression W.X+b=0

Non- Parametric model:

- The data tell you what the fit method look like.

- they have a parameters but we don't know how many of them, the data will tell the model how many of them.

10) Most machine learning approaches use training sets, test sets and validation sets to derive models. Describe the role each of the three sets plays!

- اشرحهم بالطريقه الاحسن ليك والي انت فاهمها

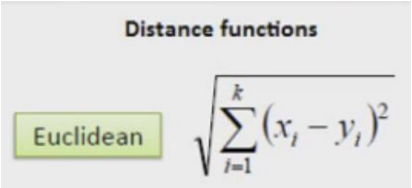
Training Set: is used to train the machine learning model. It consists of a labeled dataset where the algorithm learns the patterns, relationships, and features present in the data. The model adjusts its parameters during training to minimize the difference between its predictions and the actual labels in the training set.

Validation Set: is used to fine-tune the hyperparameters of the model and to assess its performance during training. The model is not directly trained on the validation set, but its performance on this set helps in selecting the best model architecture, tuning parameters, and preventing overfitting. It provides an unbiased evaluation before testing on unseen data.

Test Set: is a completely independent dataset that the model has not seen during training or validation. It is used to evaluate the final performance of the trained model. its simulates real-world scenarios where the model encounters new, unseen data. Evaluating on it provides an unbiased estimate of the model's generalization performance and helps assess its ability to make accurate predictions on new, unseen examples.

Question 5: Use complete-link agglomerative clustering to group the data described by the following distance matrix. Show the dendrograms.

Notice: *complete-link agglomerative* is we consider the distance between one cluster and another cluster to be equal *to the greatest distance* from any member of one cluster to any member of the other cluster.



Solution:

	A	B	C	D
A	0	1	4	5
B		0	2	6
C			0	3
D				0

First, we choose the minimum distance to choose the clusters (A, B) = 1.

Distance between Cluster (A, B) and C = Max (AC, BC) = (4, 2) = 4

Distance between Cluster (A, B) and D = Max (AD, BD) = (5, 6) = 6

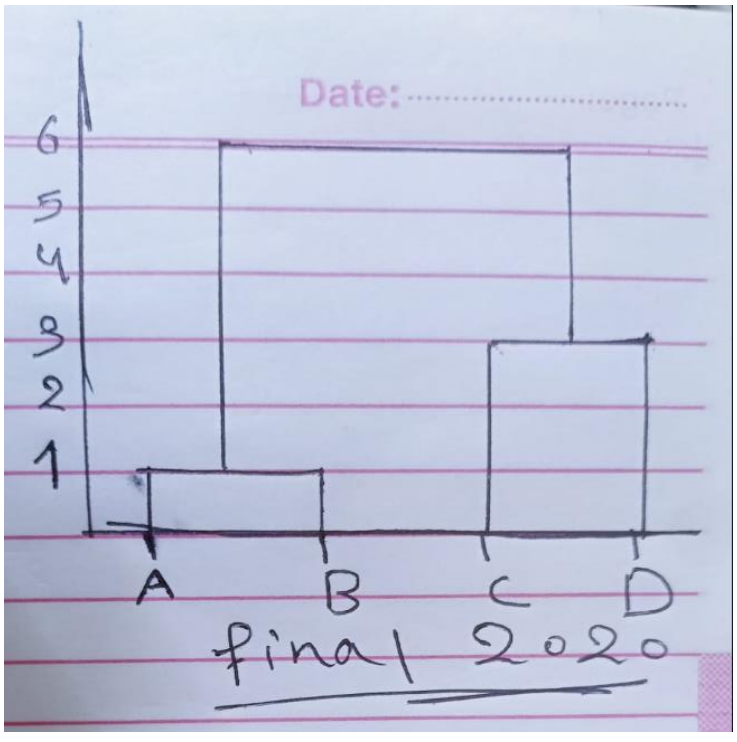
	(A, B)	C	D
(A, B)	0	4	6
C		0	3
D			0

Second, we choose the minimum distance to choose the new clusters (C, D) = 3

Distance between Cluster (C, D) and (A, B) = Max (C, (A, B), D, (A, B)) = (4, 6) = 6

	(A, B), (C, D)
(A, B), (C, D)	0

Dendrograms:



Question 6:

You are a robot in an animal shelter, and must learn to discriminate Dogs from Cats. You choose to learn a Naive Bayes classifier. You are given the following examples:

Example	Sound	Fur	Color	Class
Example #1	Meow	Coarse	Brown	Dog
Example #2	Bark	Fine	Brown	Dog
Example #3	Bark	Coarse	Black	Dog
Example #4	Bark	Coarse	Black	Dog
Example #5	Meow	Fine	Brown	Cat
Example #6	Meow	Coarse	Black	Cat
Example #7	Bark	Fine	Black	Cat
Example #8	Meow	Fine	Brown	Cat

a) Construct a classifier using Naive byes to discriminate Dogs from Cats.

$P(Cat) = \frac{4}{8}$ $P(Dog) = \frac{4}{8}$

Sound	Dog	Cat
Meow	$\frac{1}{4}$	$\frac{3}{4}$
Bark	$\frac{3}{4}$	$\frac{1}{4}$

Fur	Dog	Cat
Coarse	$\frac{3}{4}$	$\frac{1}{4}$
Fine	$\frac{1}{4}$	$\frac{3}{4}$

Color	Dog	Cat
Brown	$\frac{2}{4}$	$\frac{2}{4}$
Black	$\frac{2}{4}$	$\frac{2}{4}$

b) Consider a new example (Sound=Bark ∧ Fur=Coarse ∧ Color=Brown) Which Class belong to ?

P(Class=Dog)×P(Sound=Bark|Class=Dog)×P(Fur=Coarse|Class=Dog)×P(Color=Brown|Class=Dog)

$= \frac{4}{8} \times (\frac{3}{4} \times \frac{3}{4} \times \frac{2}{4}) = \frac{9}{64}$

P(Class=Cat)×P(Sound=Bark|Class=Cat)×P(Fur=Coarse|Class=Cat)×P(Color=Brown|Class=Cat)=

$\frac{4}{8} \times (\frac{1}{4} \times \frac{1}{4} \times \frac{2}{4}) = \frac{1}{64}$

Normalize the probabilities:

$$P(\text{Class=Dog} \mid \text{Sound=Bark} \wedge \text{Fur=Coarse} \wedge \text{Color=Brown}) = \frac{\frac{9}{64}}{\frac{9}{64} + \frac{1}{64}} = \frac{9}{10}$$

$$P(\text{Class=Cat} \mid \text{Sound=Bark} \wedge \text{Fur=Coarse} \wedge \text{Color=Brown}) = \frac{\frac{1}{64}}{\frac{9}{64} + \frac{1}{64}} = \frac{1}{10}$$

So, We label (Sound=Bark ∧ Fur=Coarse ∧ Color=Brown) as Dog

Done