



Cairo University
Faculty of Computers and Information

جامعة القاهرة
جامعة القاهرة

Final Exam

الطالب
الطالب
الطالب

Department: Computer Sciences

Course Title: Machine Learning

Course Code: CS467

Semester: 2017/2018

Instructors: Dr Hanaa Bayoumi

Date: 21 Aug. 2017

Exam Duration: 2 Hours

60

Question	Mark	Signature
One		
Two		
Three		
Four		
Five		
Six		
Seven		
Eight		
Nine		
Total Marks		

Total Marks in Writing: _____

Attempt ALL Questions:

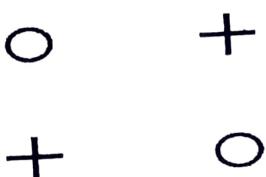
Question 1: Complete [15 marks]

- 1- Consider Y (output) Binary , X_i (features) continuous, $X = \langle X_1, X_2, \dots, X_n \rangle$, the number of estimated parameters in Naïve bayes will be.....and in logistic regression
- 2- For a neural network, which one of these structural assumptions is the one that most affects the trade-off between underfitting and overfitting
- 3- Logistic Regression transforms the output probability to be in a range of $[0, 1]$. the function is used by logistic regression to convert the probability in the range between $[0,1]$ is
- 4- (true / false) For small training sets, Naive Bayes generally is more accurate than logistic regression
- 5- (true / false) Hierarchical clustering methods require a predefined number of clusters, much like k-means
- 6- (true / false) The number of parameters in a parametric model is fixed, while the number of parameters in a nonparametric model grows with the amount of training data
- 7- (true / false) Support vector machines are resistant to outliers, i.e., very noisy examples drawn from a different distribution
- 8- As model complexity increases, bias will while variance will
- 9- Convolutional layer in convolution neural network used for while the fully connected layer used for
- 10- Performance of machine learning model depend on ,
- 11- and Consider examples of parametric models while and consider non parametric models in machine learning

Question 2 answer the following questions [19 marks]

- 1) Consider a naive Bayes classifier with 3 boolean input variables, X_1, X_2 and X_3 , and one Boolean output, Y . How many parameters must be estimated to train such a naive Bayes classifier? (you need not list them unless you wish to, just give the total) [3 marks]

- 2) Consider the following data set:



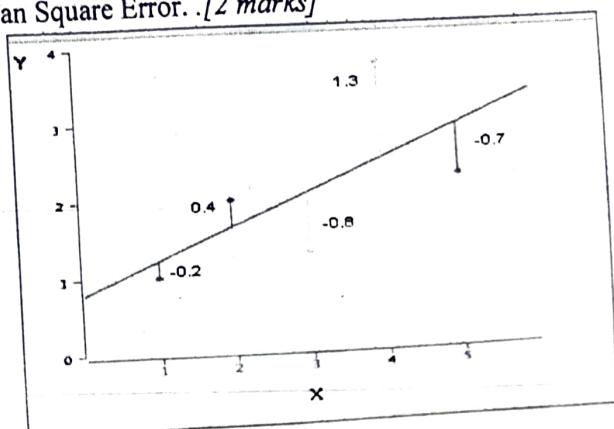
[2 marks]

Choose all of the classifiers that will achieve zero training error on this data set. (You may choose more than one.)

- (a) Logistic regression
(c) Depth-2 ID3 decision trees

- (b) SVM (quadratic kernel)
(d) K nearest neighbors ($k=3$) classifier

- 3) The graph below represents a regression line predicting Y from X . The values on the graph show the residuals (distance between real value and the estimated value) for each predictions value. Use this information to compute the Mean Square Error. [2 marks]



4) Give two advantages of hierarchical clustering over K-means clustering, and two advantages of K-means clustering over hierarchical clustering. [4 marks]

5) What is the difference between single link and complete link? [2 marks]

6) What is the importance of using Relue layer and Pooling layer in convolutional neural network?
[3 marks]

7) If you are given m data points, and use half for training and half for testing, the difference between training error and test error decreases as m increases. True or False? Explain. [3 marks]

Question 3 [6 marks]

In a convolutional neural network, consider RGB image with dimension 125×125 , suppose you use 10 filter with dimension 5×5 with stride 3 are used in convolution layer.

- a) what is the number of estimated parameters in this layer and the resulting feature map of these filters over the image?
- b) Suppose you use pad with three pixels border in this case what is the output of the convolutional layer?

Question 4

[10 marks]

In this problem, you are asked to apply hierarchical clustering to a set of seven data points, with dissimilarities between the points given by the matrix. These dissimilarities are actually distances between the points.

	X	Y
P1	0.5	2
P2	2.5	3
P3	4.2	0.7
P4	5.5	0.3
P5	4.8	3.5
P6	7	2.5
P7	8.5	2.8

	p ₁	p ₂	p ₃	p ₄	p ₅	p ₆	p ₇
p ₁	0	2.24	3.92	5.28	4.55	6.52	8.04
p ₂	2.24	0	2.86	4.04	2.35	4.53	6.00
p ₃	3.92	2.86	0	1.36	2.86	3.33	4.79
p ₄	5.28	4.04	1.36	0	3.28	2.66	3.91
p ₅	4.55	2.35	2.86	3.28	0	2.42	3.77
p ₆	6.52	4.53	3.33	2.66	2.42	0	1.53
p ₇	8.04	6.00	4.79	3.91	3.77	1.53	0

- a) Apply basic agglomerative hierarchical clustering to this data using the single link notion of dissimilarity between clusters. Visualize the result as a dendrogram.

s)
Question 5

[10 marks]

Consider the following data set consisting of the scores of two variables on each of seven individuals:
You will simulate the k-means algorithm to identify TWO clusters in the data.

Subject	A	B
1	1.0	1.0
2	1.5	2.0
3	3.0	4.0
4	5.0	7.0
5	3.5	5.0
6	4.5	5.0
7	3.5	4.5

Suppose you are given initial assignment cluster center as {cluster1: #1}, {cluster2: #2} – the first data point is used as the first cluster center and the forth point as the second cluster center.



Cairo University
Faculty of Computers and Information

Final Exam

الطالب رقم

الرقة انتقام والرابعة

Department: Computer Sciences
Course Title: Machine Learning
Course Code: CS467
Semester: 2018/2019
Instructors: Dr Hanaa Bayoumi
Date: 6 Jan. 2019
Exam Duration: 2 Hours

تصميم الحاسوب

60

Question	Mark	Signature
One		
Two		
Three		
Four		
Five		
Six		
Seven		
Eight		
Nine		
Total Marks		

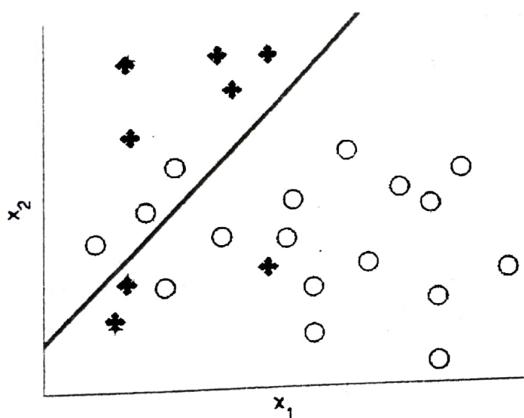
Total Marks in Writing: _____

attempt all questions

Question 1: [6 marks]

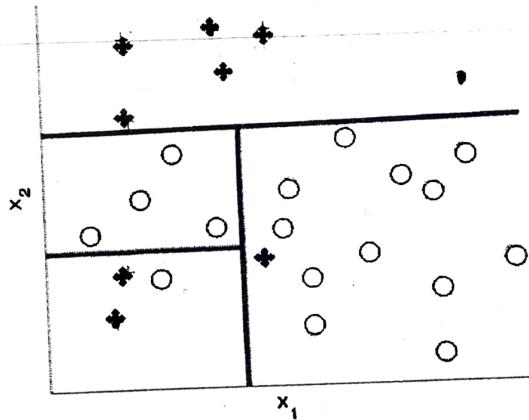
The following figures depict decision boundaries of classifiers obtained from three learning algorithms: decision trees, logistic regression, and nearest neighbor classification (in some order). Under each of the three plots, write the name of the learning algorithm and the number of mistakes it makes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

+ positive training examples
○ negative training examples



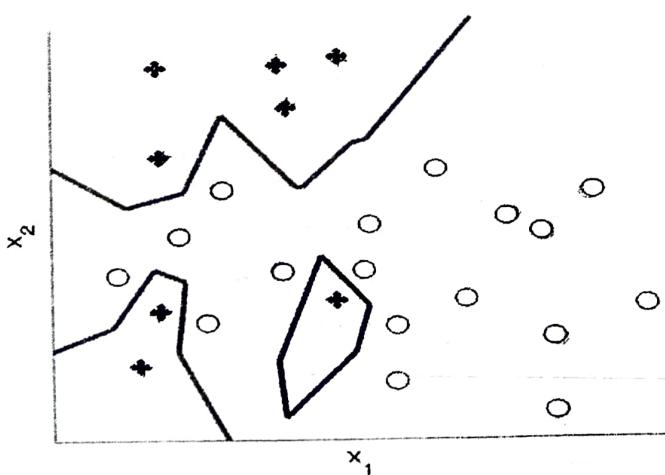
Classifier name:

Number of mistakes:



Classifier name:

Number of mistakes:



Classifier name:

Number of mistakes:

Question 2 answer the following questions [12 marks]

We will use the dataset below to learn a decision tree which predicts if people pass machine learning (Yes or No), based on their previous GPA (High, Medium, or Low) and whether or not they studied.

GPA	Studied	Passed
L	F	F
L	T	T
M	F	F
M	T	T
H	F	T
H	T	T

a) What is the entropy $H(\text{Passed})$?

b) What is the entropy $H(\text{Passed} \mid \text{GPA})$?

c) What is the entropy $H(\text{Passed} \mid \text{Studied})$?

- d) Draw the full decision tree that would be learned for this dataset.

Question 3 [5 marks]

In SVM show how to express the margin in terms of weight (W) (show all steps)

Question 4 Answer the following questions [20 marks]

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

2) Why we need re-estimate probabilities (smoothing) in Naïve Bayes classifier?

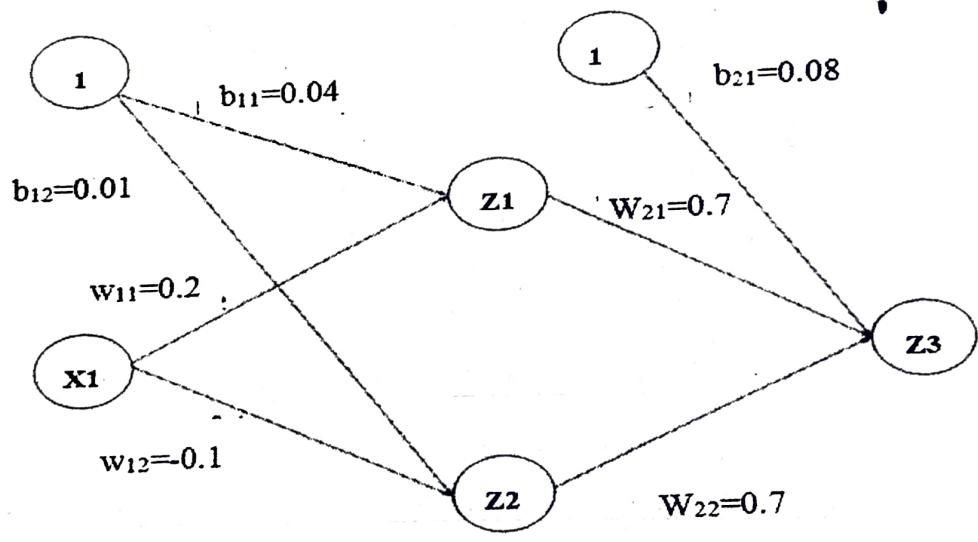
- 3) What is the similarity and difference between feature selection and dimensionality reduction?
- 4) True/ false, a single perceptron can only compute linear variation AND, OR, and XOR?
Explain in one sentence
- 5) When it is possible to run Gradient Descent algorithm, what is guaranteed by the algorithm (1 sentence)? And what isn't guaranteed by the algorithm (1 sentence)?
- 6) In the K-nearest neighbor classifier, which of the following statement(s) are true?
- a) A KNN is supervised classifier ()
 - b) The hyper parameter K in KNN is typically set to an odd number ()
 - c) When K is set to an extremely large number, it is more likely that the classifier will overfit than underfit. ()
 - d) Both KNN and K-means are unsupervised learning techniques ()
 - e) Increase k in a k-nearest neighbor classifier increase bias ()

- 7) 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	
k-Nearest Neighbor	
Support Vector Machines	
Multivariate Linear Regression	
Logistic Regression	
Perceptron	
Multilayer Feed-Forward Neural Network	
K-Means	

Question 5

[7 marks]



Consider the neural network architecture shown above for a 2-class (0,1) classification problem the values of weights and biases are shown in the figure , we define

$$a_1 = w_{11}x_1 + b_{11}$$

$$a_2 = w_{12}x_1 + b_{12}$$

$$a_3 = w_{21}z_1 + w_{22}z_2 + b_{21}$$

$$z_1 = \text{relu}(a_1)$$

$$z_2 = \text{relu}(a_2)$$

$$z_3 = \sigma(a_3), \sigma(x) = \frac{1}{1+e^{-x}}$$

Use these information for answer the following questions

- a) For $x_1=0.3$, compute z_3 in terms of e (show all steps)
- b) Which class does the network predict the given data point ($x_1=0.3$), i.e. y' . Note that $y'=1$ if $z_3>0.5$ else $y'=0$

Question 6

[6 marks]

In a convolutional neural network, consider RGB image with dimension 125×125 , suppose you use 5 filters with dimension 5×5 with stride 3 are used in convolution layer.

- a) what is the number of estimated parameters in this layer and the resulting feature map of these filters over the image?
- b) Suppose you use pad with three pixels border in this case what is the output of the convolutional layer?

Question 7**[4 marks]**

Consider a data set of English words, in which all words have the same length. The distance measure between pairs of data points (words) is defined as the number of substitutions in letters required to transform one string to another. This is a modified version of the edit distance between the two strings. For example, the distance between **baba** and **bebe** is 2; since we have to substitute two letters (**a** and **a** with **e** and **e**) in the first word to transform it into the second.

Consider the following words:

{Liller, Lisler, Killen, millen, biller, booled}

These words have the following modified edit distances in the distance matrix:

	Liller	Lisler	Killen	millen	biller	booled
Liller	0	1	2	2	1	4
Lisler	1	0	3	3	2	4
Killen	2	3	0	1	2	4
millen	2	3	1	0	2	4
biller	1	2	2	2	0	3
booled	4	4	4	4	3	0

We now perform bottom-up hierarchical clustering using single or complete linkage, and at an intermediate stage, we have the following three clusters:

C1 { Liller , Lisler }

C2 { Killen , millen }

C3 { biller , booled }

[6 pts] Which cluster will be merged with C1 in the next iteration? Circle the correct option, and then justify your answer.

- (a) C2 with either single or complete linkage
- (b) C3 with either single or complete linkage
- (c) C2 with single-linkage or C3 with complete-linkage
- (d) C3 with single-linkage or C2 with complete-linkage



Cairo University
Faculty of Computers and Artificial Intelligence

Final Exam

اسم الطالب _____
نوع الطالب _____

Department: Computer science

Course Title: Machine Learning

Course Code: CS467

Semester: first - 2019/2020

Instructor: Dr. Hanaa Bayomi

Date: 5/1/2020

Exam Duration: 2 Hours

تَعْلِيمُ الْحَاسِبَاتِ

الصِّرَاقَةِ

تعليمات هامة

- حيازة telephones المحمول مفتوحا داخل لجنة الامتحان يعتبر حالة غير تستوجب العقاب وإذا كان ضروري الدخول بالموصل فيوضع مغلقا في الحقيبة.
- لا يسمح بدخول ساعة الأذن أو الليتوثر.
- لا يسمح بدخول أي كتب أو ملازم أو أوراق داخل اللجنة والمخالفة

60

Question	Mark	Signature
One		
Two		
Three		.
Four		
Five		
Six		
Seven		
Eight		
Nine		
Ten		
Total Marks		

Marks in Writing: _____

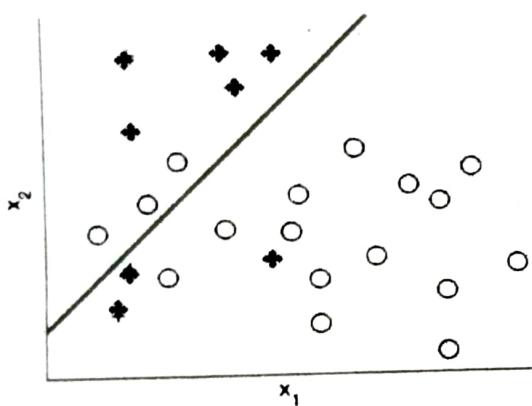
جامعة الملك عبد الله للعلوم والتقنية

Attempt ALL Questions:

Question 1: [6 marks]

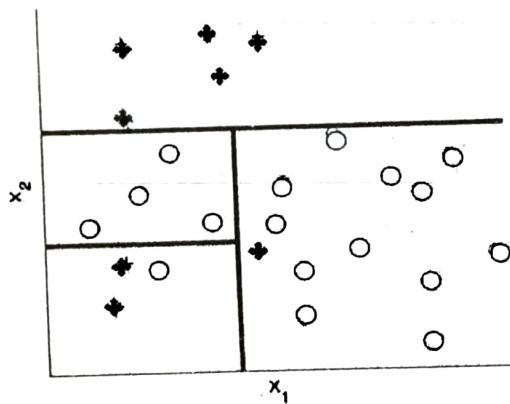
The following figures depict decision boundaries of classifiers obtained from three learning algorithms: decision trees, logistic regression, and nearest neighbor classification (in some order). Under each of the three plots, write the name of the learning algorithm and the number of mistakes it makes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

+	positive training examples
○	negative training examples



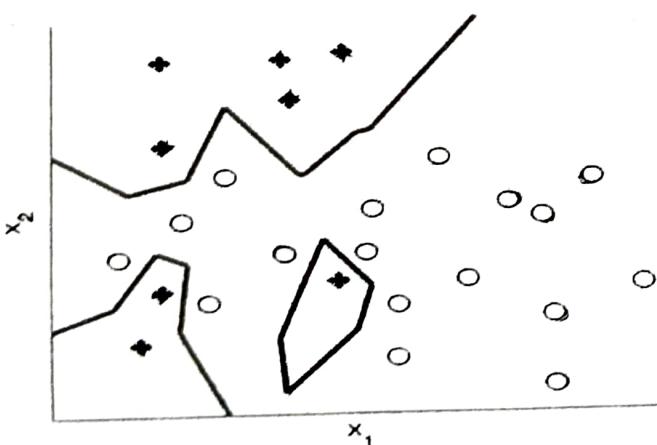
Classifier name:

Number of mistakes:



Classifier name:

Number of mistakes:

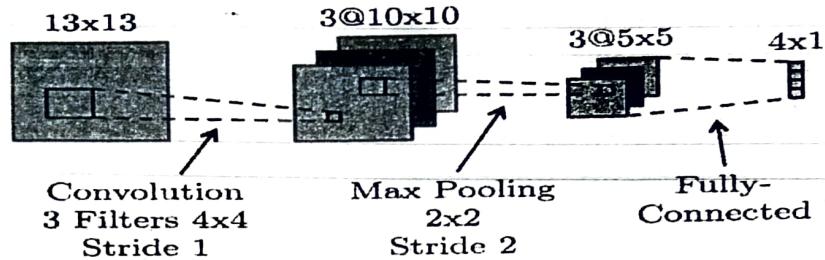


Classifier name:

Number of mistakes:

Question 2 answer the following questions [11 marks]

Below is a diagram of a small convolutional neural network that converts a 13×13 image into 4 output values. The network has the following layers/operations from input to output: convolution with 3 filters, max pooling, ReLu, and finally a fully-connected layer. For this network we will not be using any bias parameters (b). Please answer the following questions about this network.



- (a) [2 pts] How many weights in the convolutional layer do we need to learn?
- (b) [2 pts] How many ReLu operations are performed on the forward pass?
- (c) [2 pts] How many weights do we need to learn for the entire network?
- (d) [2 pts] What are the convolutional and fully connected layers in convolution neural network used for?
- (e) [3 pts] What are the disadvantages of a fully-connected neural network compared to a convolutional neural network with the same size layers?

Question 3 [5 marks]

In SVM show how to express the margin in terms of weight (W)

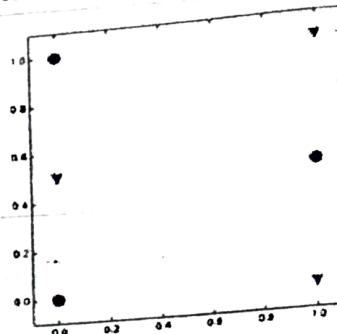
Question 4 Answer the following questions [22 marks]

- 1) Give one sentence reason why:
 - a) we might prefer Decision Tree learning over Logistic Regression for a particular learning task.
 - b) We might prefer Logistic Regression over Naive Bayes for a particular learning task.
 - c) we need re-estimate probabilities (smoothing) in Naïve Bayes classifier.
- 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.
 - This is an instance of under fitting.
 - The training and testing examples are sampled from different distributions.

- 3) What is the importance of using Relue layer in convolutional neural network?
It is preferred to apply Relue function before or after the pooling layer (justify your answer).

- 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	()
SVM with kerne	()
Decision tree	()
3-nearest-neighbor classifier (with Euclidean distance).	()



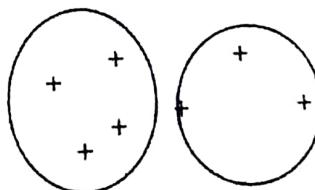
- 5) Describe the difference between parametric methods and nonparametric methods.

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

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

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

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



- 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	
k-Nearest Neighbor	
Support Vector Machines	
Multivariate Linear Regression	
Logistic Regression	
Perceptron	
Multilayer Feed-Forward Neural Network	
K-Means	

- 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! [4]

Question 5

[8 marks]

Use complete link agglomerative clustering to group the data described by the following distance matrix. Show the dendograms.

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

Question 6 [8 marks]

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

- Construct a classifier using Naïve Bayes to discriminate Dogs from Cats.
- Consider a new example (Sound=Bark ^ Fur=Coarse ^ Color=Brown), which class belong to?

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