Cairo University
Faculty of Computers and Artificial Intelligence

Midterm Exam

Department: CS
Course Name: Machine Learning
Course Code: CS467
Instructor(s): Dr. Hanaa Bayomi
Name:………………………………

Date: 19/11/2019
Duration: 1 hour
Total Marks: 20
ID:…………………………..

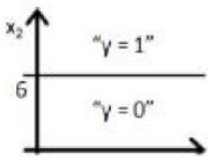*Question 1*   *[2.5 marks]*
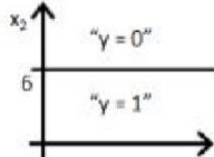
- **Answer the following Questions:**

[Suppose you train a logistic regression classifier, and the learned hypothesis function is

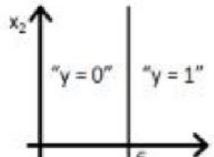$$h_\theta(x) = \sigma(\theta_0 + \theta_1 x_1 + \theta_2 x_2),$$

Where $\theta_0 = 6$; $\theta_1 = 0$; $\theta_2 = -1$. Which of the following represents the decision boundary for $h_\theta(x)$? Explain your answer.

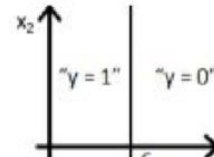A   B   C   D
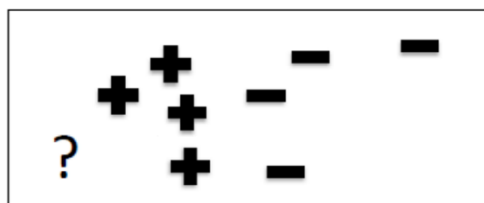
Correct answer is: B
Explanation :

Explanation: the answer is (B). We can rule out C and D because the decision boundary is independent of x_1. Additionally, $h_\theta(x)$ is smaller than 0.5 for $x_2 > 6$ (so the output is 0), and larger than 0.5 for $x_2 < 6$ (so the output is 1).

Question 2 Given the following data set   *[7 marks]*

Using KNN algorithm, For what (minimal) value of K will the query point "?" be negative? Explain your answer.

Value of K : 0

Explanation: in order to classify point "?" as a negative , should the majority of points class selected by K th nearest point should be negative and it will never happen in this case.

a) What is the difference between Parametric and Nonparametric Methods, show which of the following algorithms consider parametric and non-parametric?
(KNN - SVM - linear regression – Logistic regression- Multivariate Linear regression – Naïve Bayes -SVM with linear Kernel)

Parametric methods summarize data into a fixed number of parameters whose count does not increase as the number of data points increase.

Nonparametric methods do not summarize data into a fixed number of parameters. The number of parameters (if they're used at all) increases as the number of data points increase.

Parametric	Nonparametric
Linear Regression	KNN
Naïve bayes	
Multivariate Linear Regression	
Logistic Regression	
SVM	
SVM with linear Kernel	

Question 3   [6 marks]
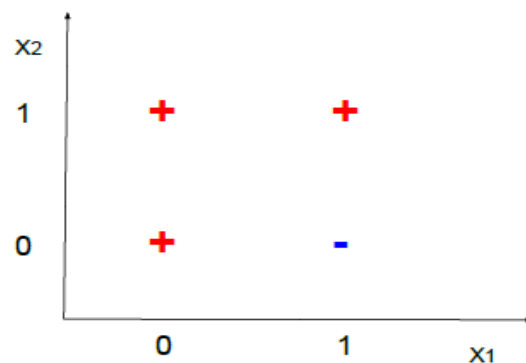
a) Suppose we learn a Naive Bayes classifier from the examples in Figure, using MLE as the training rule. Write down the probabilities.

Solution: $P(Y = 0) = 0.25, P(Y = 1) = 0.75$
$P(X_1 = 0 | Y = 1) = \frac{2}{3}$
$P(X_1 = 1 | Y = 1) = \frac{1}{3}$
$P(X_2 = 0 | Y = 1) = \frac{1}{3}$
$P(X_2 = 1 | Y = 1) = \frac{2}{3}$

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

        (a) Logistic regression   (b) SVM (quadratic kernel)
        (c) ID3 decision trees   (d) K nearest neighbors (k=3) classifier

Question 4   *[4.5 marks]*

1) Consider a naive Bayes classifier with 3 boolean input variables, $X1; X2$ and $X3$, 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]*

Solutions:

For a naive Bayes classifier, we need to estimate $P(Y=1)$, $P(X1 = 1/ y = 0)$; $P(X2 = 1/y = 0)$, $P(X3 = 1/y = 0)$, $P(X1 = 1/y = 1)$; $P(X2 = 1/y = 1)$; $P(X3 = 1/y = 1)$. Other probabilities can be obtained with the constraint that the probabilities sum up to 1.
So we need to estimate 7 parameters.

2) What will happen for bias and variance when the model complexity increase
As model complexity increases, bias will **decrease.** while variance will **increase.**

3) For each of the following learning problems, please indicate whether it is a prediction, regression or classification problem. (An explanation is not required.)

    (a)   A biologist has given different amounts of food to different rats in his laboratory. He has recorded the weight of each rat after two months. Now he wants to learn how the weight of the rats depends on the amount of food they get.

    (b)   Each spring a farmer counts the number of newborn sheep. Based on his counts of the previous years he wants to estimate the number of newborn sheep in the coming year.

    (c)   A computer program tries to determine whether a newspaper article is about politics based on the number of times the article contains the following words/phrases: 'law', 'sports', 'newspaper', 'hockey', 'elections', 'human rights' and 'party'.

Answers: (a) regression (b) prediction (c) classification