Q	Grade
1	0.6
2	0.6
3	0.6
4	0.6
5	0.6
6	0.6
7	0.5
8	0.6
9	0.6
10	1.3

UNIVERSIDADE DE AVEIRO
DEPARTAMENTO DE ELECTRÓNICA TELECOMUNICAÇÕES E INFORMÀTICA

Machine Learning final exam - 9/June 2020 PART 1 (40 min)

Nº: 85129 Name:Gabriel Augusto Santos Silva

Instructions: You have 40 min. to write down your answers of the questions below. During this time, please**, keep switched on the camera of your PC**. Save and name the file with your answers as "ML_P1_XXXXX" and substitute XXXXX with your academic *(*mechanographic) number.
Send a **PDF** version of the file with your answers and a PDF file of the digitalized pages, you may have produced while solving the problems, to petia@ua.pt with **Subject: ML_P1 + your academic number**

Q1. Suppose you are running a company, and you want to develop learning algorithms to address the following problems:
Problem1: You have a large inventory of identical items. You want to predict how many of these items will sell over the next 3 months.
Problem 2: You would like software to examine individual customer accounts, and for each account decide if it has been hacked (compromised).
Should you treat these as classification or as regression problems:

A. Treat both as classification problems.
B. Treat Problem 1 as a classification problem, Problem 2 as a regression problem.
C. Treat Problem 1 as a regression problem, Problem 2 as a classification problem.
D. Treat both as regression problems.

Answer: C

Q2. You run gradient descent for a number of iterations with learning rate α=0.3 and compute the cost $J(\theta)$ over each iteration. You find that the value of $J(\theta)$ increases over time. Based on this, which one of the following conclusions seems most plausible?

A. Rather than use the current value of α, it would be better to try a larger value of α (say α=1).

B. Rather than use the current value of α, it would be better to try a smaller value of α (say α=0.1)

C. αdoes not affect the convergence of the cost function $J(\theta)$

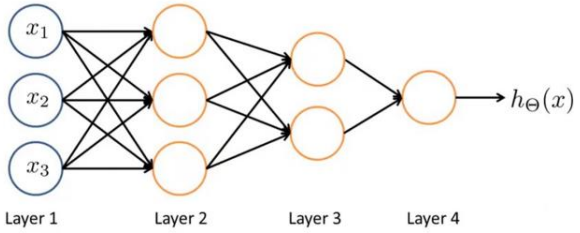
Answer: B

Q3. Suppose you have m=14 examples with n=3 features. What are the dimensions of the data matrix X, the output y and the vector of parameters θ when you implement linear regression.

A. X is 14x4, y is 14x1, θ is 4x1
B. X is 14x3, y is 14x1, θ is 3x3
C. X is 14x3, y is 14x1, θ is 3x1
D. X is 14x4, y is 14x4, θ is 4x4

Answer: A

Q4. How many parameter matrices has this neural network (NN)? What is the dimension of each matrix ? What is the total number of NN parameters ?



Answer: There are 3 parameters matrices Theta. Theta$^{(1)}$ is 3*4, Theta$^{(2)}$ is 2*4 and Theta$^{(3)}$ is 1*3, which gives a total of 12 + 8 + 3 = 23 parameters.

Q5. Suppose you have implemented regularized classifier to classify what object is in an image (i.e., to do object recognition). When you test this model on a new set of images, it makes unacceptably high errors. However, the model performs well (has low error) on the training set. Which of the following are promising steps to take? Check all that apply.

- **A.** Use a smaller set of features.
- **B.** Decrease the regularization parameter λ.
- **C.** Increase the train set examples.
- **D.** Increase the regularization parameter λ.

Answer: A,D

Q6. Consider the following training set. You want to fit a linear regression model $h_\theta(x) = \theta_0 + \theta_1 x_1 + \theta_2 x_2$, but first you need to scale the features applying both division by *max-min* and *mean* normalization. What will be the normalized value of the first feature of the first example $x_1^{(1)}$?

x1	x2	y
69	4761	78
72	5184	74
94	8836	87
89	7921	96

Answer: mean of x1 = (69+72+94+89)/4= 81 . max-min = 94-69= 25 . x1$^{(1)}$ = |(69- 81)|/25 = 0.48

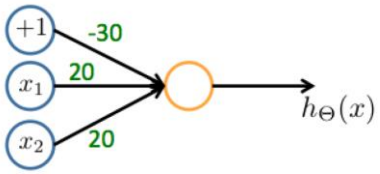
Q7. You have trained an anomaly detection system for fraud detection, to flag anomalies when *p(x) < epsilon*. However, on the cross-validation set it mis-flagging many good transactions as fraudulent. What should you do?

A. Decrease *epsilon*

B. Increase *epsilon*

Answer: A

Q8. Consider the following simple Neural Network that takes two binary inputs x_1 (0 or 1) and x_2 (0 or 1) and applies sigmoid function. Which of the following logical functions will compute it (approximately)? **Justify why ?**



A. XOR (exclusive OR)

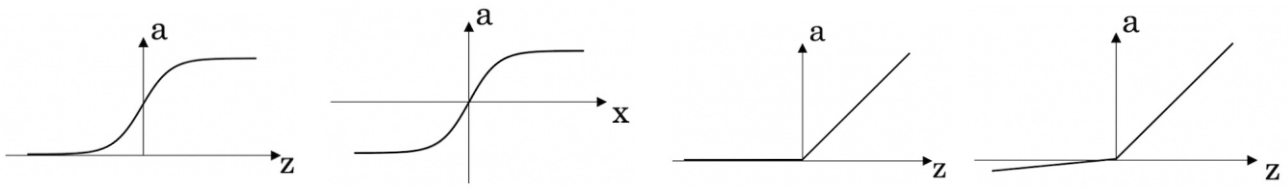
C. AND

B. OR

D. NAND (i.e. NOT AND)

Answer: C

Q9. Which one of these functions is the prefered and most widely used in Convolutional Neural Networks (CNN)? What is the name of the function?



A) B) C) D)

Answer: C, it's called a RELU (Rectified Linear Unit)

Q10 The following animal data set has 4 Features (Give Birth, Can Fly, Live in Water, Have legs) and 2 Classes (mammals and non-mammals) . Apply Naïve Bayes classifier to decide the class of the new example?

Name	Give Birth	Can Fly	Live in Water	Have Legs	Class
human	yes	no	no	yes	mammals
python	no	no	no	no	non-mammals
salmon	no	no	yes	no	non-mammals
whale	yes	no	yes	no	mammals
frog	no	no	sometimes	yes	non-mammals
komodo	no	no	no	yes	non-mammals
bat	yes	yes	no	yes	mammals
pigeon	no	yes	no	yes	non-mammals
cat	yes	no	no	yes	mammals
leopard shark	yes	no	yes	no	non-mammals
turtle	no	no	sometimes	yes	non-mammals
penguin	no	no	sometimes	yes	non-mammals
porcupine	yes	no	no	yes	mammals
eel	no	no	yes	no	non-mammals
salamander	no	no	sometimes	yes	non-mammals
gila monster	no	no	no	yes	non-mammals
platypus	no	no	no	yes	mammals
owl	no	yes	no	yes	non-mammals
dolphin	yes	no	yes	no	mammals
eagle	no	yes	no	yes	non-mammals

New example

Give birth	Can Fly	Live in water	Have legs	Class
No	Yes	No	Yes	Answer: ?

Answer:

P(x_new/Mammals)= 1/7 * 1/7 * 5/7 * 5/7 = 0.0104

P(x_new/Non-mammals)= 12/13 * 3/13 * 6/13 * 9/13 = 0.0681

P(x_new/M)P(M) = 0.0104*(7/20) = 0.0036
P(x_new/N)P(N) = 0.0681*(13/20) = 0.044

x_new is assigned to Non_mammal because P(x_new/N)P(N) > P(x_new/M)P(M)
