Q	Grade
11	0.6
12	0.6
13	0.6
14	0.6
15	0.6
16	0.6
17	0.6
18	0.6
19	0.6
20	1.3

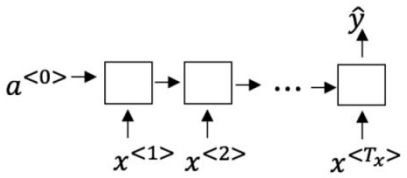
UNIVERSIDADE DE AVEIRO
DEPARTAMENTO DE ELECTRÓNICA TELECOMUNICAÇÕES E INFORMÀTICA

Machine Learning final exam - 9/June 2020 PART 2 (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_P2_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_P2 + your academic number**

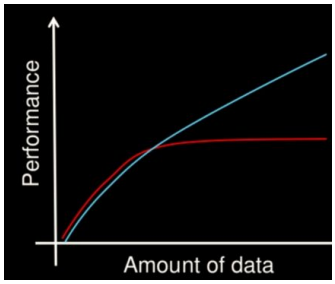
Q11. To which of the following tasks the Recurrent Neural Network (RNN) shown below is more suitable? Check all that apply.



- **A.** Speech recognition (input an audio clip and output a transcript)
- **B.** Sentiment classification (input a piece of text and output a 0/1 to denote positive or negative sentiment)
- **C.** Image classification (input an image and output a label).
- **D.** Gender recognition from speech (input an audio clip and output a label indicating the speaker's gender)

Answer: A,B

Q12. Which of the following assumptions are correct regarding the meaning of the red and blue curves ?



A. Red curve (conventional machine learning), blue curve (deep learning).
B. Red curve (training data performance), blue curve (test data performance) .
C. Red curve (deep learning), blue curve (conventional machine learning).
D. Red curve (test data performance), blue curve (training data performance) .

Answer: A

Q13. Which of the following statements regarding Softmax Regression are true? Check all that apply.

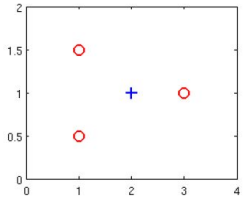
A. Softmax Regression is suitable for both classification and regression problems in machine learning.
B. Softmax Regression will output a k dimensional vector, whose elements are between 0 and 1.
C. Softmax Regression is more suitable than Logistic Regression for mutually exclusive classes.
D. Softmax Regression is recommended only in deep learning architectures.

Answer: B,C

Q14. Suppose you have the following training set and fit a logistic regression classifier

$h_\theta(x) = g(\theta_0 + \theta_1 x_1 + \theta_2 x_2)$.

x_1	x_2	y
1	0.5	0
1	1.5	0
2	1	1
3	1	0



Which of the following is true ? Check all that apply.

A. The cost function is convex, so the gradient descent will converge to the global minimum.
B. Adding (creating) new features may increase how well the classifier separates the data.
C. Instead of logistic regression better apply softmax regression.
D. The data are non-linearly separable, add a regularization term to improve the classification.

Answer: A,B

Q15. You have implemented regularized classifier to predict what items customers will purchase on a web shopping site. However, when you test your hypothesis on a new set of customers, it makes unacceptably large errors in its predictions. Furthermore, the hypothesis performs poorly on the training set. Which of the following might be promising steps to take? Check all that apply.

- **A.** Use more training examples.
- **B.** Use a smaller set of features.
- **C.** Increase the regularization parameter λ.
- **D.** Use additional features.

Answer: D

Q16. You have trained a binary classifier and obtained the following confusion matrix. Compute the classifier's accuracy (in %) and the F1 Score?

	Actual Class 1	Actual Class 2
Predicted class 1	890	85
Predicted class 2	10	15

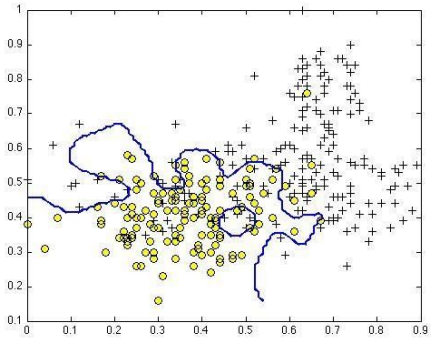
Answer:

accuracy = (890+15) / (890+85+10+15) = 90.5 %

recall = 890/(890+10) = 98.9%
precision = 890/(890+85) = 91.3%
f1_score = (2*Recall*Precision)/(Recall + Precision) = 94.9 %

Q17. Suppose you have trained an SVM classifier with a Gaussian kernel, and it learned the following decision boundary on the training set. The SVM's performance on a cross validation set is poor. Should you increase or decreasing C? Increase or decrease σ?

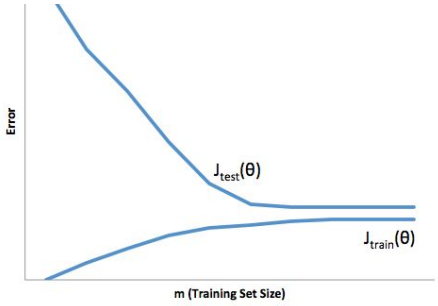
- **A.** It would be reasonable to decrease C and decrease σ.
- **B.** It would be reasonable to increase C and increase σ.



C. It would be reasonable to decrease C and increase σ.
D. It would be reasonable to increase C and decrease σ.

Answer: C

Q18. You train a learning algorithm, and find that it makes unacceptably high errors on test data. You plot the learning curve, and obtain the figure below. Is the algorithm suffering from high bias, high variance, or neither?



A. High bias

B. Neither

C. High variance

Answer: A

Q19. Suppose you have a dataset with n = 10 features and m = 5000 examples. After training a logistic regression classifier with gradient descent, you find that it has underfit the training set and does not achieve the desired performance on the training or cross validation sets. Which of the following might be promising steps to take? Check all that apply.

A. Reduce the number of examples in the training set.

B. Add (or create) new features.

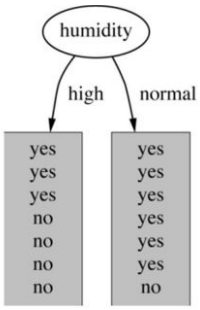
C. Use a different optimization method since using gradient descent to train logistic regression might result in a local minimum.

D. Try using a more complex model.

Answer: B,D

Q20. You want to build a decision tree for the weather data based on which to decide Play golf or Not. Compute the information gain for the split at the node for humidity.

OUTLOOK	TEMP	HUMIDITY	WINDY	PLAY
Sunny	Hot	High	False	No
Sunny	Hot	High	True	No
Overcast	Hot	High	False	Yes
Rainy	Mild	High	False	Yes
Rainy	Cool	Normal	False	Yes
Rainy	Cool	Normal	True	No
Overcast	Cool	Normal	True	Yes
Sunny	Mild	High	False	No
Sunny	Cool	Normal	False	Yes
Rainy	Mild	Normal	False	Yes
Sunny	Mild	Normal	True	Yes
Overcast	Mild	High	True	Yes
Overcast	Hot	Normal	False	Yes
Rainy	Mild	High	True	No%%



Answer: H(pre-split) = -((9/14)*log2(9/14) + (5/14)*log2(5/14)) = 0.94029

H(high) = -((3/7)*log2(3/7) + (4/7)*log2(4/7)) = 0.98523

H(normal) = -((6/7)*log2(6/7) + (1/7)*log2(1/7)) = 0.59167

H(post-split) = (7/14)*H(high) + (7/14)*H(normal) = 0.78845

Information gain = H(pre-split) - H(post-split) = 0.94029 - 0.78845 = 0.15184
