PHY 654 quiz, 29th August 2024

The exam consists of 4 pages, not including this cover page. Please go through your copy to make sure that all pages have been printed.

The first part of the exam consists of short questions (MCQ). Choose exactly one option as the answer to MCQ. If more than one option is chosen you will get zero, even if one of the chosen options is correct.
In the regime where *one* answer is chosen, the following rule applies:
For selecting correct answer: $+1$
For selecting wrong answer: -0.5
Note the negative marking in MCQ.

The second part of the exam consists of longer questions.

There are 20 points in total in this quiz. Good luck!

Name: _____

Roll Number: _____

Page	Points	Score
1	4	
2	3	
3	9	
4	4	
Total:	20	

IIT Kanpur
PHY 654
Machine learning in particle physics
Quiz (I)

Short Questions

1. (1 point) What is the role of gradient descent in linear regression?
   A. Gradient descent removes outliers from the dataset to help the model make better predictions.
   B. Gradient descent is an iterative process that finds the best values of parameters that minimize the loss.
   C. Gradient descent helps to determine what type of loss to use when training a model.
   D. None of the above.

2. (1 point) Why does a model need to be trained before it can make predictions?
   A. A model doesn't need to be trained. Models are available on most computers.
   B. A model needs to be trained to learn the mathematical relationship between the features and the label in a dataset.
   C. A model needs to be trained so it won't require data to make a prediction.
   D. None of the above.

3. (1 point) A logistic regression model with three features has the following bias and weights:

$$b = 1$$
$$w_1 = 2,\ w_2 = -1,\ w_3 = 5$$

   The following input feature values are given: $x_1 = 0$, $x_2 = 10$, $x_3 = 2$. What is the logistic regression prediction (y_{pred}) for these input values?
   A. 0.268
   B. 0.500
   C. 0.731
   D. 98.00

4. (1 point) Suppose your model is overfitting. Which of the following is NOT a valid way to try and reduce the overfitting?
   A. Increase the amount of training data.
   B. Try a regularizing technique.
   C. Decrease the model complexity.
   D. Improve the optimisation algorithm being used for error minimisation.

5. (1 point) Let's say we have a neural network that can neither fit the training dataset nor generalize to new data (i.e test dataset). Which of the following is true?

   A. This is happening due to overfitting.

   B. This is happening due to underfitting.

   C. It is overfitting in the training data and underfitting in the test data.

   D. It is underfitting in the training data and overfitting in the test data.

6. (1 point) In a binary-classification problem, let's say we are using logistic regression and the following log-loss:

$$\text{Loss}(h_\theta(x), y) = \begin{cases} -\log h_\theta(x) & \text{if} \quad y = 1 \\ -\log(1 - h_\theta(x)) & \text{if} \quad y = 0 \end{cases}$$

   where, $h_\theta(x)$ is the hypothesis.
   Which of the following is incorrect?

   A. If $h_\theta(x) = y$, then $\text{Loss}(h_\theta(x), y)$ is 0, for both $y = 0$ case and $y = 1$ case.

   B. If $y = 0$, then $\text{Loss}(h_\theta(x), y) \to \infty$ as $h_\theta(x) \to 1$.

   C. Regardless of whether $y = 0$ or $y = 1$, if $h_\theta(x) = 0.5$, then $\text{Loss}(h_\theta(x), y) > 0$.

   D. If $y = 0$, then $\text{Loss}(h_\theta(x), y) \to \infty$ as $h_\theta(x) \to 0$.

7. (1 point) You are evaluating the performance of two binary classification models: Model A and Model B. You found that Model A has an AUC of 0.5. Model B's predictions are made completely randomly. Which of the following statements is true?

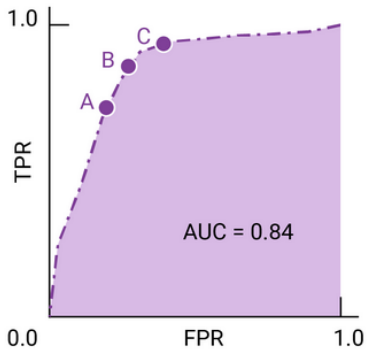
   A. Model A performs better than Model B.

   B. Model B performs better than Model A.

   C. Model A and Model B perform equally well.

   D. None of the above

8. (1 point) Imagine a situation where it's better to allow some spam to reach the inbox than to send a business-critical email to the spam folder. You've trained a spam classifier for this situation where the positive class is spam and the negative class is not-spam. Which of the following points on the ROC curve for your classifier is preferable?



A. Point A

B. Point B

C. Point C

Long Questions

9. (2 points) What distinguishes a supervised approach from an unsupervised approach in ML?

10. (2 points) What is a confusion matrix in the context of binary classification?

11. You want to build a ML model to predict whether an object reconstructed in the CMS detector is a photon (label = 1) or a jet (label = 0). You have 5 input features (i.e. $n = 5$) and 9000 training examples (i.e $m = 9000$). You decide to train a logistic regression that outputs a probability $y_{\text{pred}} \in [0, 1]$ that the object is a photon, where the predicted label is chosen to be 1 when $y_{\text{pred}} \geq 0.5$, and 0 otherwise.

    (a) (2 points) How many weights(w) and bias(b) parameters does the logistic regression model have?

    (b) (2 points) After training your logistic regression model, you decide to train a single hidden layer neural network with 6 hidden units on the same dataset. How many weights and biases does this neural network have?

(c) (3 points) Let's assume that you remembered to use a sigmoid activation in your neural network's output layer, but you forgot to use non-linear activation function for the hidden layer. Mistakenly, you used an identity function as the activation function of hidden layer. In that case, show mathematically that your implementation is equivalent to logistic regression.

(d) (1 point) Now assume that, in your neural network, you have remembered to use a non-linear activation function in the hidden layer. But in output layer, instead of using $y_{\text{pred}} = \text{sigmoid}(z)$, you have used $y_{\text{pred}} = \text{sigmoid}(\text{ReLU}(z))$. So, mistakenly, you have used two activation functions, instead of one, in the output layer. What problem are you going to encounter because of this?

End of exam.