Intro to Big Data Science — Spring 2023-2024

Name: ____________________ **ID No.:** ________________

Quiz 3

To receive credit, this worksheet MUST be handed in at the end of the class.

1. Suppose you want to predict a house's price as a function of its size. Your model is

$$f_\theta(x) = w_0 + w_1 * \text{size} + w_2 * \sqrt{\text{size}},$$

where $\theta = (w_0, w_1, w_2)$. Suppose size ranges from 1 to 1000 (feet^2). You will implement this by fitting a model

$$f_\theta(x) = w_0 + w_1 x_1 + w_2 x_2.$$

Finally, suppose you want to use feature scaling (without mean normalization). Which of the following choices for x_1 and x_2 should you use?(Note: $\sqrt{1000} \approx 32$)

   (A) x_1=size, $x_2 = 32\sqrt{\text{size}}$

   (B) x_1=32 size, $x_2 = \sqrt{\text{size}}$

   (C) x_1=size/1000, $x_2 = \sqrt{\text{size}}/32$

   (D) x_1=size/32, $x_2 = \sqrt{\text{size}}$

2. Which of the following is not true for logistic regression?

   (A) Logistic regression can deal with problems which are not linearly separable.

   (B) MLE can be used to estimate the parameters in logistic regression.

   (C) Logistic regression is used for classification.

   (D) Logistic regression is used for regression.

3. For two-class classification problems, which of the following is correct about linear discriminant analysis (LDA)?

   (A) LDA can do better than Bayes classifier.

   (B) LDA assumes the different covariance matrix for two classes.

   (C) LDA classifies samples based on their projection on the coordinate axis.

   (D) LDA assumes the conditional Gaussian distributions for each class.

4. Suppose you have a classification problem. The misclassification error is defined as $\frac{1}{m}\sum_{i=1}^{n} \text{err}(h_\theta(x^{(i)}), y^{(i)})$, and the cross validation misclassification error is similarly defined, using the CV examples $(x^{(1)}), y^{(1)}), \ldots, (x^{(n_{cv})}), y^{(n_{cv})})$. Suppose your training error is 0.10, and your CV error is 0.30. What problem is the algorithm most likely to be suffering from?

   (A) High bias (overfitting)

   (B) High bias (underfitting)

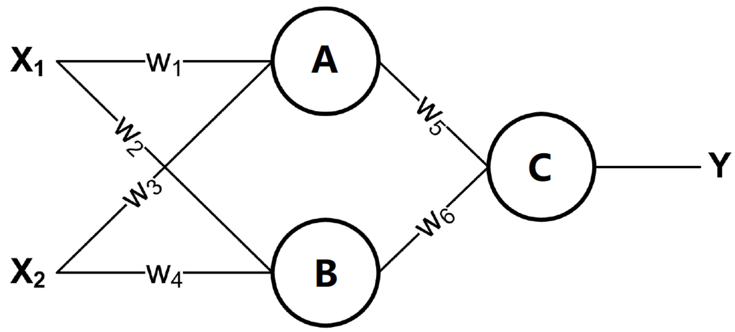
   (C) High variance (overfitting)

   (D) High variance (underfitting)

5. Please give the THREE BIG NAMES of scientists contributing greatly to deep learning.

6. Consider a two-layer neural network to learn a function $f : \mathbf{x} \mapsto y$ where $\mathbf{x} = (x_1, x_2)^T$ consists of two attributes. The weights $w_1, \ldots, w_6$ can be arbitrary. There are **three** units (namely, A, B and C) in this network. We have two possible choices for the function implemented by each unit in this network:

   - **S**: signed sigmoid function $S(a) = \text{sign}(\frac{1}{1+\exp(-a)} - 0.5)$;
   - **L**: linear function $L(a) = ca$ with a given constant c.



   (a) Assign proper activation functions (**S** or **L**) to each unit in the following graph so this neural network simulates a linear regression: $y = \beta_1 x_1 + \beta_2 x_2$. And derive β_1 and β_2 in terms of $w_1, \ldots, w_6$.

   (b) Assign proper activation functions (**S** or **L**) to each unit in the following graph so this neural network simulates a binary logistic regression classifier: $y = \arg\max_y \text{P}(Y = y|\mathbf{X} = \mathbf{x})$, where $\text{P}(Y = 1|\mathbf{X} = \mathbf{x}) = \frac{\exp(\beta_1 x_1 + \beta_2 x_2)}{1+\exp(\beta_1 x_1 + \beta_2 x_2)}$ and $\text{P}(Y = -1|\mathbf{X} = \mathbf{x}) = \frac{1}{1+\exp(\beta_1 x_1 + \beta_2 x_2)}$. And derive β_1 and β_2 in terms of $w_1, \ldots, w_6$.

   (c) Assign proper activation functions (**S** or **L**) to each unit in the following graph so this neural network simulates a boosting classifier which combines two logistic regression classifiers, $f_1 : \mathbf{x} \mapsto y_1$ and $f_2 : \mathbf{x} \mapsto y_2$, to produce its final prediction: $y = \text{sign}(\alpha_1 y_1 + \alpha_2 y_2)$. Use the same definition in Step 2 for f_1 and f_2. And derive α_1 and α_2 in terms of $w_1, \ldots, w_6$.