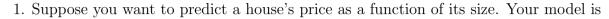
## Intro to Big Data Science — Spring 2021-2022

Name: A A T

ID No.: 12011327

Quiz 3

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



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

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

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

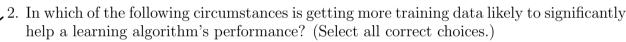
Finally, suppose you want to use feature scaling (without mean normalization). Which of the following choices for  $x_1$  and  $x_2$  should you use? (Note:  $\sqrt{1000} \approx 32$ )

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

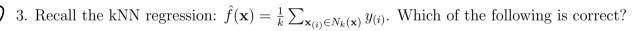
(B) 
$$x_1=32$$
 size,  $x_2=\sqrt{\text{size}}$ 

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

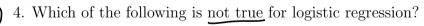
(D) 
$$x_1 = \text{size}/32$$
,  $x_2 = \sqrt{\text{size}}$ 



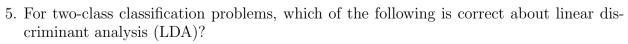
- (A) Algorithm is suffering from high bias.
- (B) Algorithm is suffering from high variance.
- (C) CV error is much larger than training error.
- (D) CV error is about the same as training error.



- (A) Small k may lead to large bias and small variance
- (B) Large k may lead to small bias and large variance
- (C) The model may be overfitted for too large value of k
- (D) Appropriate selection of k by cross-validation could avoid overfitting



- (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.



- (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.





- 6. Suppose you have a classification problem. The misclassification error is defined as  $\frac{1}{m}\sum_{i=1}^{n} \operatorname{err}(h_{\theta}(x^{(i)}), y^{(i)})$ , 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)
- 7. (Optional) For elastic net,

$$\min_{\mathbf{w}} \left[ (\mathbf{y} - \mathbf{X}\mathbf{w})^T (\mathbf{y} - \mathbf{X}\mathbf{w}) + \lambda_1 \|\mathbf{w}\|_1 + \lambda_2 \|\mathbf{w}\|_2^2 \right]$$

By introducing augmented  $\tilde{\mathbf{X}}$  and  $\tilde{\mathbf{y}}$  (you have to give the exact form for  $\tilde{\mathbf{X}}$  and  $\tilde{\mathbf{y}}$  in terms of  $\mathbf{X}$ ,  $\mathbf{y}$ , and  $\lambda_2$ ), can you convert it into a standard Lasso?