Assignment 2 (Questions)
Introduction to Machine Learning
Prof. B. Ravindran

1. Let $A^{m \times n}$ be a matrix of real numbers. The matrix AA^T has an eigenvector x with eigenvalue b. Then the eigenvector y of A^TA which has eigenvalue b is equal to

   (a) x^TA
   (b) A^Tx
   (c) x
   (d) Cannot be described in terms of x

2. Let $A^{n \times n}$ be a row stochaistic matrix - in other words, all elements are non-negative and the sum of elements in every row is 1. Let b be an eigenvalue of A. Which of the following is true?

   (a) $|b| > 1$
   (b) $|b| \leq 1$
   (c) $|b| \geq 1$
   (d) $|b| < 1$

3. Let u be a $n \times 1$ vector, such that $u^Tu = 1$. Let I be the $n \times n$ identity matrix. The $n \times n$ matrix A is given by $(I - kuu^T)$, where k is a real constant. u itself is an eigenvector of A, with eigenvalue -1. What is the value of k?

   (a) -2
   (b) -1
   (c) 2
   (d) 0

4. Which of the following are true for any $m \times n$ matrix A of real numbers.

   (a) The rowspace of A is the same as the columnspace of A^T
   (b) The rowspace of A is the same as the rowspace of A^T
   (c) The eigenvectors of AA^T are the same as the eigenvectors of A^TA
   (d) The eigenvalues of AA^T are the same as the eigenvalues of A^TA

5. Consider the following 4 training examples

   - $x = -1, y = 0.0319$
   - $x = 0, y = 0.8692$
   - $x = 1, y = 1.9566$
   - $x = 2, y = 3.0343$

   We want to learn a function $f(x) = ax + b$ which is parametrized by (a, b). Using squared error as the loss function, which of the following parameters would you use to model this function.

(a) $(1,1)$

(b) $(1,2)$

(c) $(2,1)$

(d) $(2,2)$

6. You are given the following five training instances

- $x_1 = 2, x_2 = 1, y = 4$
- $x_1 = 6, x_2 = 3, y = 2$
- $x_1 = 2, x_2 = 5, y = 2$
- $x_1 = 6, x_2 = 7, y = 3$
- $x_1 = 10, x_2 = 7, y = 3$

We want to model this function using the K-nearest neighbor regressor model. When we want to predict the value of y corresponding to $(x_1, x_2) = (3, 6)$

(a) For $K = 2, y = 3$

(b) For $K = 2, y = 2.5$

(c) For $K = 3, y = 2.33$

(d) For $K = 3, y = 2.666$

7. Bias and Variance can be visualized using a classic example of a dart game. We can think of the true value of the parameters as the bull's-eye on a target, and the arrow's value as the estimated value from each sample. Consider the following situations, and select the correct option(s)

(a) Player 1 has low variance compared to player 4

(b) Player 1 has higher variance compared to player 4

(c) Bias exhibited by player 2 is more than that done by player 3.

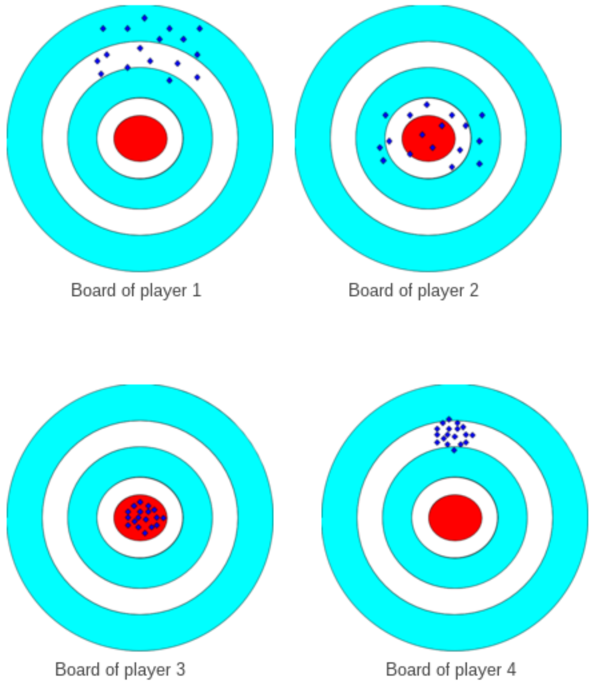

Figure 1: Figure for Q7

8. Choose the correct option(s) from the following.

   (a) When working with a small dataset, one should prefer low bias/high variance classifiers over high bias/low variance classifiers.

   (b) When working with a small dataset, one should prefer high bias/low variance classifiers over low bias/high variance classifiers.

   (c) When working with a large dataset, one should prefer high bias/low variance classifiers over low bias/high variance classifiers.

   (d) When working with a large dataset, one should prefer low bias/high variance classifiers over high bias/low variance classifiers.

9. Consider a modified k-NN method in which once the k nearest neighbours to the query point are identified, you do a linear regression fit on them and output the fitted value for the query point. Which of the following is/are true regarding this method.

   (a) This method makes an assumption that the data is locally linear.

   (b) In order to perform well, this method would need dense distributed training data.

   (c) This method has higher bias compared to K-NN

   (d) This method has higher variance compared to K-NN

10. The Singular Value Decomposition (SVD) of a matrix R is given by USV^T. Consider an orthogonal matrix Q and $A = QR$. The SVD of A is given by $U_1 S_1 {V_1}^T$. Which of the

following is/are true?
Note-There can be more than one correct option.

(a) $U = U_1$

(b) $S = S_1$

(c) $V = V_1$

11. Assume that the feature vectors defining the training data are not all linearly independent. What happens if we apply the standard linear regression formulation considering all feature vectors?

(a) The coefficients $\hat{\beta}$ are not uniquely defined.

(b) $\hat{y} = X\hat{\beta}$ is no longer the projection of y into the column space of X.

(c) X is full rank.

(d) $X^T X$ is singular.