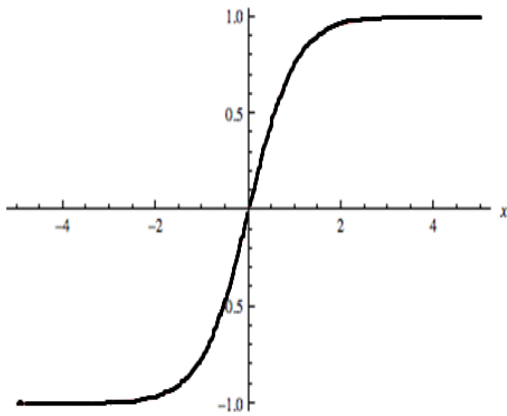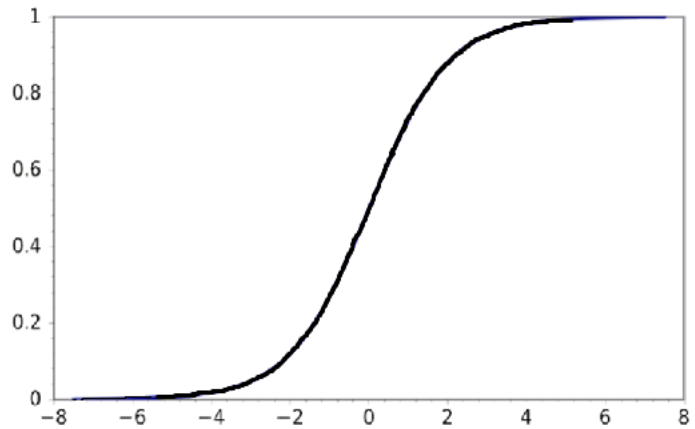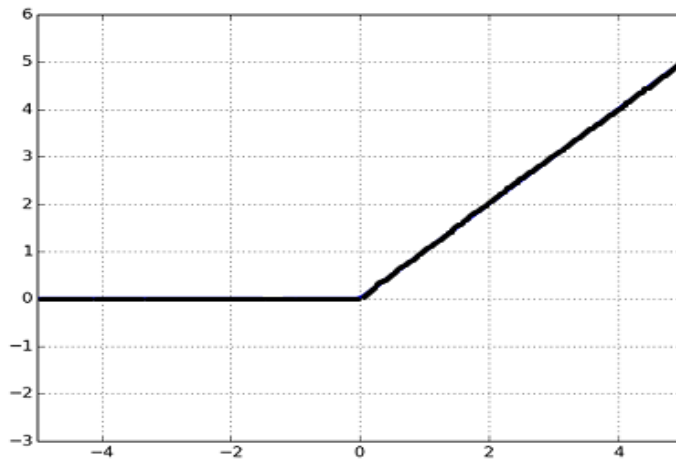# MACHINE LEARNING

1) Given below are three images (1,2,3). Which of the following option is correct for these images?

A)

B)

C)

A) 1 is tanh, 2 is ReLU and 3 is SIGMOID activation functions.

B) 1 is SIGMOID, 2 is ReLU and 3 is tanh activation functions.

C) 1 is ReLU, 2 is tanh and 3 is SIGMOID activation functions.

D) 1 is tanh, 2 is SIGMOID and 3 is ReLU activation functions.
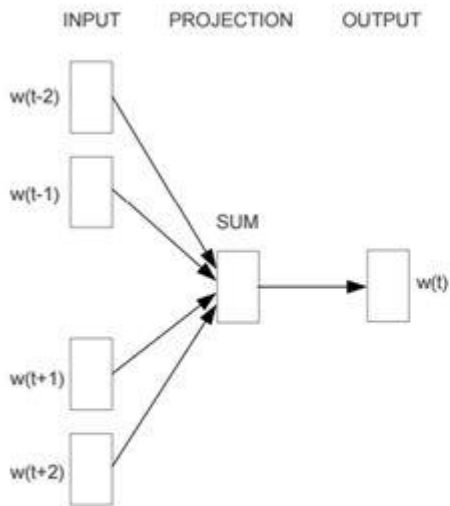
Solution: (D)

The range of SIGMOID function is [0,1].

The range of the tanh function is [-1,1].

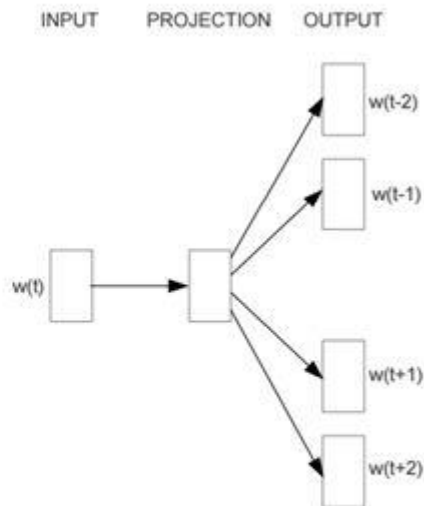The range of the RELU function is [0, infinity].

So Option D is the right answer.

MACHINE LEARNING

2) Skip gram model is one of the best models used in Word2vec algorithm for words embedding. Which one of the following models depicts the skip gram model?

INPUT	PROJECTION	OUTPUT
w(t-2)		
w(t-1)	SUM	
		w(t)
w(t+1)		
w(t+2)		

Model A

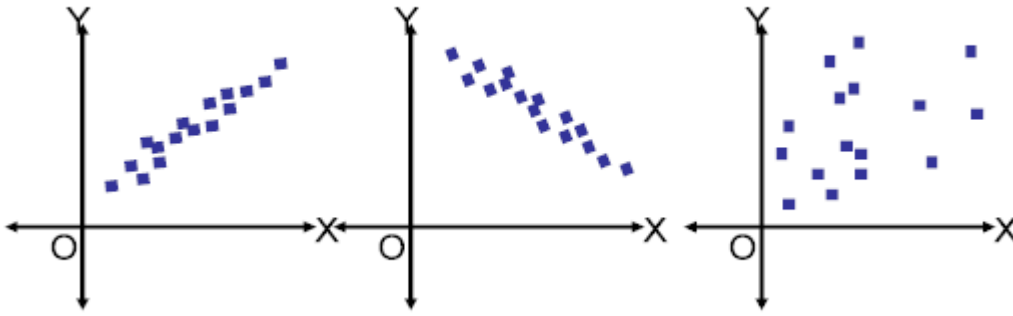| INPUT	PROJECTION	OUTPUT
		w(t-2)
		w(t-1)
w(t)		
		w(t+1)
		w(t+2)

Model B

A) A

B) B

C) Both A and B

D) None of these

Solution: (B)

Both models (model1 and model2) are used in Word2vec algorithm. The model1 represent a CBOW model where as Model2 represent the Skip gram model.

3) Given below are three scatter plots for two features (Image 1, 2 & 3 from left to right).

In the above images, which of the following is/are examples of multi-collinear features?

A) Features in Image 1

B) Features in Image 2

C) Features in Image 3

D) Features in Image 1 & 2

E) Features in Image 2 & 3

F) Features in Image 3 & 1

Solution: (D)

In Image 1, features have high positive correlation where as in Image 2 has high negative correlation between the features so in both images pair of features are the example of multicollinear features.

4) In previous question, suppose you have identified multi-collinear features. Which of the following action(s) would you perform next?

1. **Remove both collinear variables.**
2. **Instead of removing both variables, we can remove only one variable.**
3. **Removing correlated variables might lead to loss of information. In order to retain those variables, we can use penalized regression models like ridge or lasso regression.**

A) Only 1

B) Only 2

C) Only 3

D) Either 1 or 3

E) Either 2 or 3

Solution: (E)

You cannot remove the both features because after removing the both features you will lose all of the information so you should either remove the only 1 feature or you can use the regularization algorithm like L1 and L2.

MACHINE LEARNING

5) Given below is a scenario for training error TE and Validation error VE for a machine learning algorithm M1. You want to choose a hyper parameter (H) based on TE and VE.

H	TE	VE
1	105	90
2	200	85
3	250	96
4	105	85
5	300	100

 Which value of H will you choose based on the above table?

A) 1

B) 2

C) 3

D) 4

E) 5

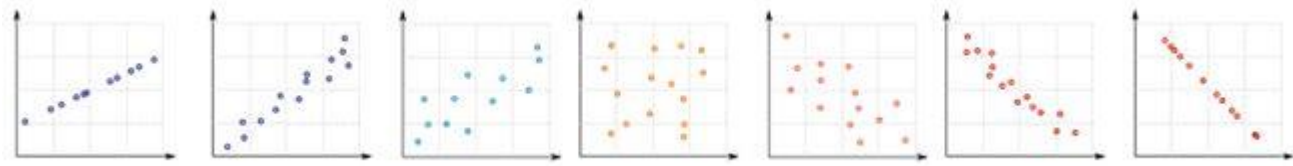**Solution: (D)**

Looking at the table, option D seems the best

6) Suppose you are given 7 Scatter plots 1-7 (left to right) and you want to compare Pearson correlation coefficients between variables of each scatter plot.

Which of the following is in the right order?

1. 1<2<3<4
2. 1>2>3 > 4
3. 7<6<5<4
4. 7>6>5>4

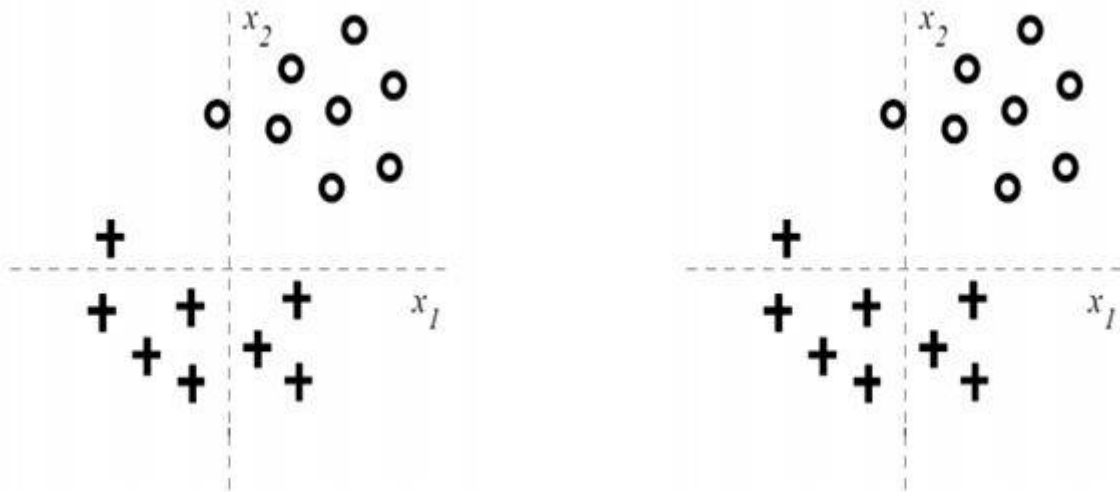A) 1 and 3

B) 2 and 3

C) 1 and 4

D) 2 and 4

Solution: (B)

from image 1to 4 correlation is decreasing (absolute value). But from image 4 to 7 correlation is increasing but values are negative (for example, 0, -0.3, -0.7, -0.99).

MACHINE LEARNING

7) Suppose you are given the below data and you want to apply a logistic regression model for classifying it in two given classes.

You are using logistic regression with L1 regularization.

$$\sum_{i=1}^{n} \log P(y_i|x_i, w_0, w_1, w_2) - C(|w_1| + |w_2|).$$

Where C is the regularization parameter and w1 & w2 are the coefficients of x1 and x2.

Which of the following option is correct when you increase the value of C from zero to a very large value?

A) First w2 becomes zero and then w1 becomes zero

B) First w1 becomes zero and then w2 becomes zero

C) Both becomes zero at the same time

D) Both cannot be zero even after very large value of C

Solution: (B)

By looking at the image, we see that even on just using x2, we can efficiently perform classification. So at first w1 will become 0. As regularization parameter increases more, w2 will come more and more closer to 0.

8) Imagine you are working on a project which is a binary classification problem. You trained a model on training dataset and get the below confusion matrix on validation dataset.

n=165	Predicted: NO	Predicted: YES
Actual: NO	50	10
Actual: YES	5	100

Based on the above confusion matrix, choose which option(s) below will give you correct predictions?

1. **Accuracy is ~0.91**
2. **Misclassification rate is ~ 0.91**
3. **False positive rate is ~0.95**
4. **True positive rate is ~0.95**

A) 1 and 3

B) 2 and 4

C) 1 and 4

D) 2 and 3

Solution: (C)

The Accuracy (correct classification) is (50+100)/165 which is nearly equal to 0.91.

The true Positive Rate is how many times you are predicting positive class correctly so true positive rate would be 100/105 = 0.95 also known as "Sensitivity" or "Recall"

9) Suppose, we were plotting the visualization for different values of C (Penalty parameter) in SVM algorithm. Due to some reason, we forgot to tag the C values with visualizations. In that case, which of the following option best explains the C values for the images below (1,2,3 left to right, so C values are C1 for image1, C2 for image2 and C3 for image3) in case of rbf kernel.
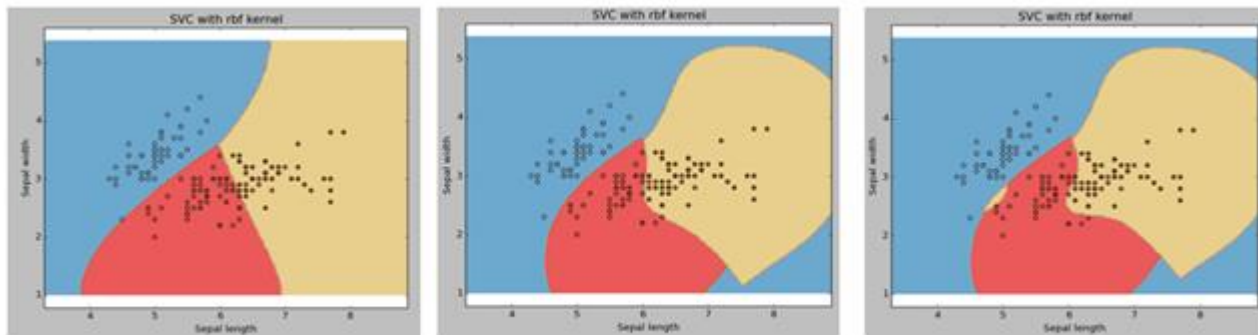
A) C1 = C2 = C3

B) C1 > C2 > C3

C) C1 < C2 < C3

D) None of these

Solution: (C)

Penalty parameter C of the error term. It also controls the trade-off between smooth decision boundary and classifying the training points correctly. For large values of C, the optimization will choose a smaller-margin hyper plane. .

10) Which of the following algorithm would you take into the consideration in your final model building on the basis of performance?

Suppose you have given the following graph which shows the ROC curve for two different classification algorithms such as Random Forest(Red) and Logistic Regression(Blue)
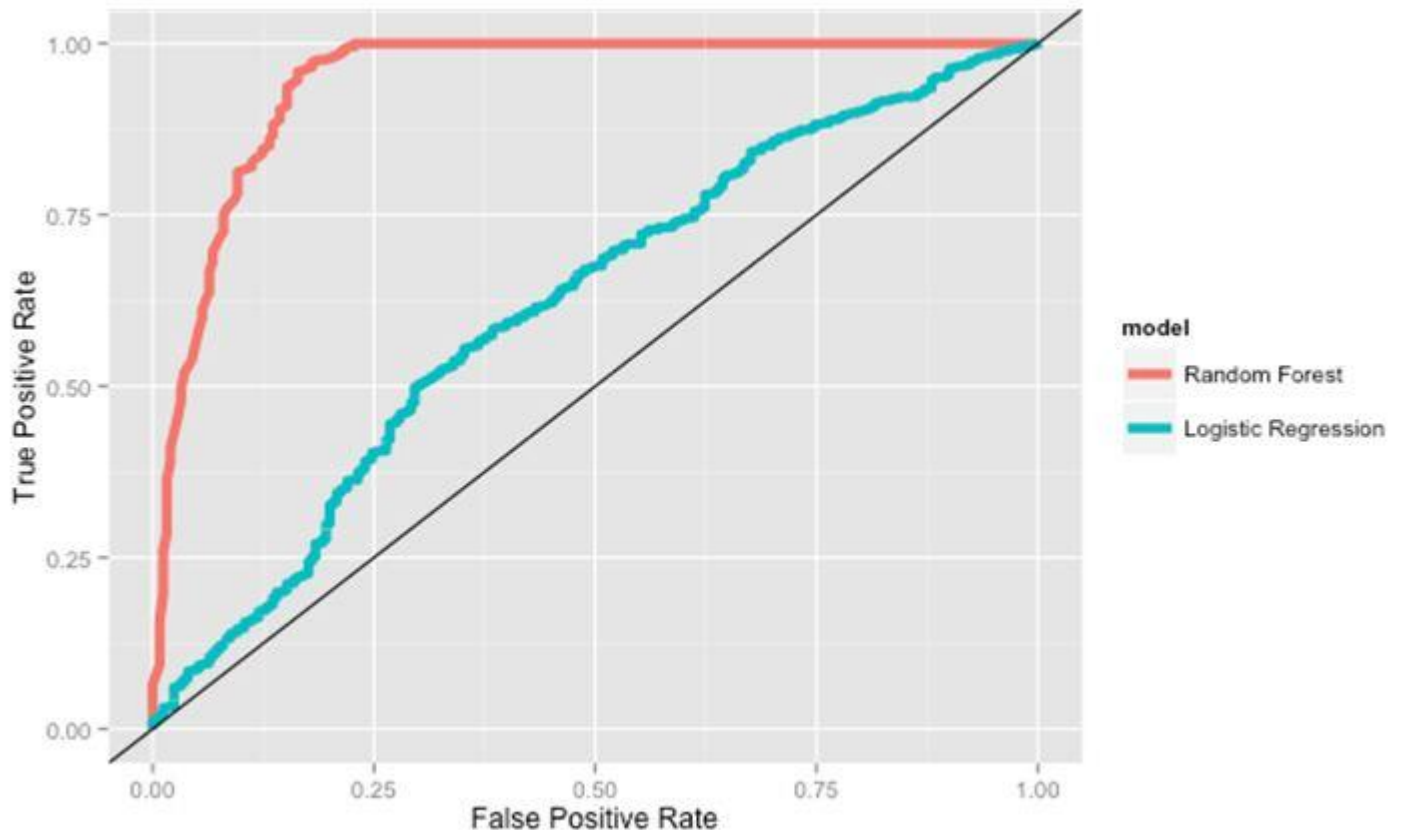
A) Random Forest
B) Logistic Regression
C) Both of the above
D) None of these

Solution: A

Since, Random forest has largest AUC given in the picture so I would prefer Random Forest

Consider the following figure for answering the next few questions. In the figure, X1 and X2 are the two features and the data point is represented by dots (-1 is negative class and +1 is a positive class). And you first split the data based on feature X1(say splitting point is x11) which is shown in the figure using vertical line. Every value less than x11 will be predicted as positive class and greater than x will be predicted as negative class.

11) How many data points are misclassified in above image?

A) 1
B) 2
C) 3
D) 4
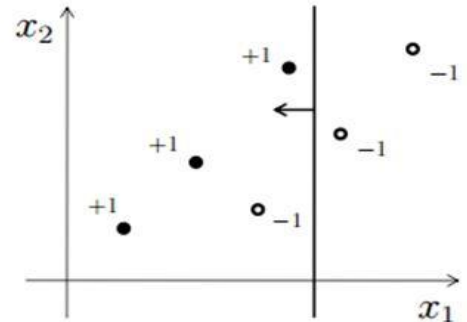
Solution: A

Only one observation is misclassified, one negative class is showing at the left side of vertical line which will be predicting as a positive class.

12) Which of the following splitting point on feature x1 will classify the data correctly?

A) Greater than x11
B) Less than x11
C) Equal to x11
D) None of above

Solution: D

If you search any point on X1 you won't find any point that gives 100% accuracy.

13) If you consider only feature X2 for splitting. Can you now perfectly separate the positive class from negative class for any one split on X2?
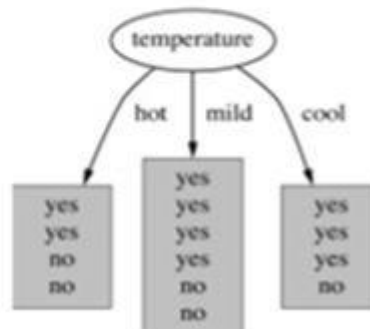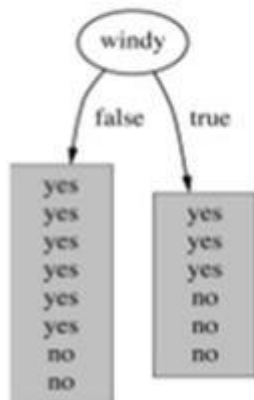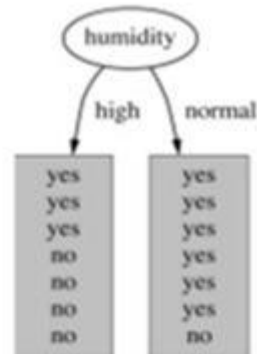
A) Yes
B) No

Solution: B

It is also not possible.

MACHINE LEARNING

14) Suppose you are building random forest model, which split a node on the attribute, that has highest information gain. In the below image, select the attribute which has the highest information gain?

A) Outlook
B) Humidity
C) Windy
D) Temperature

Solution: A

Information gain increases with the average purity of subsets. So option A would be the right answer

15) Suppose you have given the following scenario for training and validation error for Gradient Boosting. Which of the following hyper parameter would you choose in such case?

| Scenario	Depth	Training Error	Validation Error
1	2	100	110
2	4	90	105
3	6	50	100
4	8	45	105
5	10	30	150

A) 1
B) 2
C) 3
D) 4

Solution: B

Scenario 2 and 4 has same validation accuracies but we would select 2 because depth is lower is better hyper parameter.

MACHINE LEARNING

Suppose you have given the following 2-class data where "+" represent a positive class and "" is represent negative class.
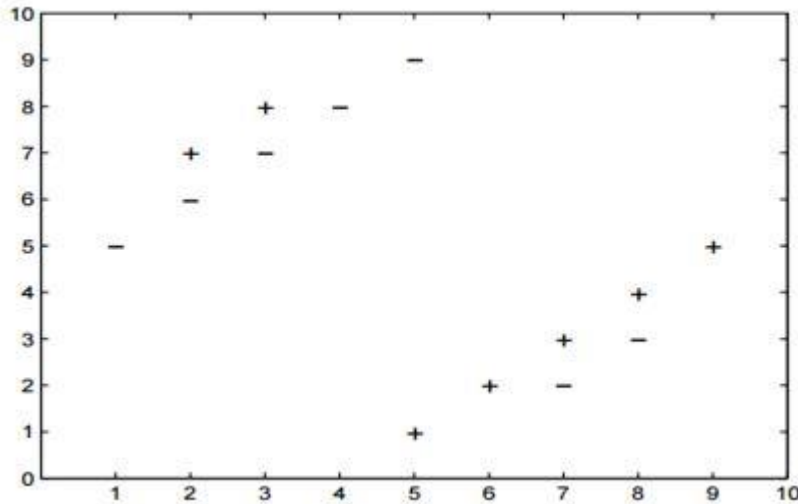
16) Which of the following value of k in k-NN would minimize the leave one out cross validation accuracy?

A) 3
B) 5
C) Both have same
D) None of these

Solution: B

5-NN will have least leave one out cross validation error.

MACHINE LEARNING

17) The following two distances (Euclidean Distance and Manhattan Distance) have given to you which generally we used in K-NN algorithm. These distance are between two points A(x1, y1) and B(x2,Y2).

Your task is to tag the both distance by seeing the following two graphs. Which of the following option is true about below graph ?


A) Left is Manhattan Distance and right is Euclidean Distance
B) Left is Euclidean Distance and right is Manhattan Distance
C) Neither left or right are a Manhattan Distance
D) Neither left or right are a Euclidian Distance
Solution: B

Left is the graphical depiction of how Euclidean distance works, whereas right one is of Manhattan distance.

MACHINE LEARNING

18) Suppose you have given the following images(1 left, 2 middle and 3 right), Now your task is to find out the value of k in k-NN in each image where k1 is for 1st, k2 is for 2nd and k3 is for 3rd figure.
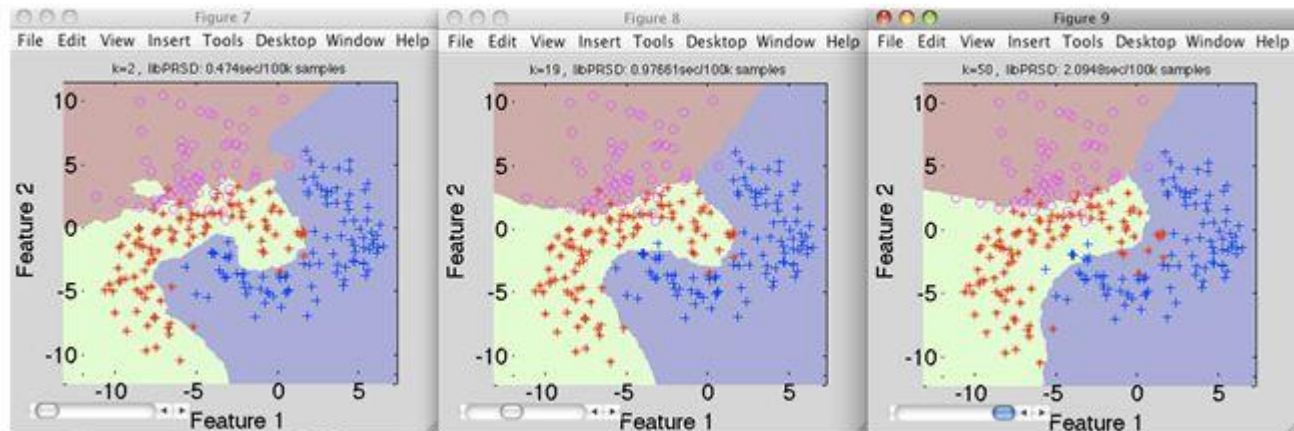
A) k1 > k2> k3
B) k1<k2
C) k1 = k2 = k3
D) None of these
Solution: D

Value of k is highest in k3, whereas in k1 it is lowest