Introduction to Machine Learning -IITKGP

Assignment - 5

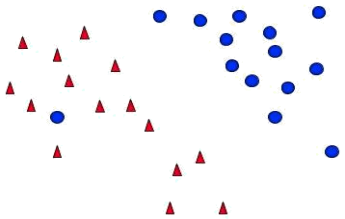
TYPE OF QUESTION: MCQ/MSQ

Number of questions: 15 Total mark: 2 * 15 = 30

---

1. What would be the ideal complexity of the curve which can be used for separating the two classes shown in the image below?

a. Linear
b. Quadratic
c. Cubic
d. insufficient data to draw a conclusion

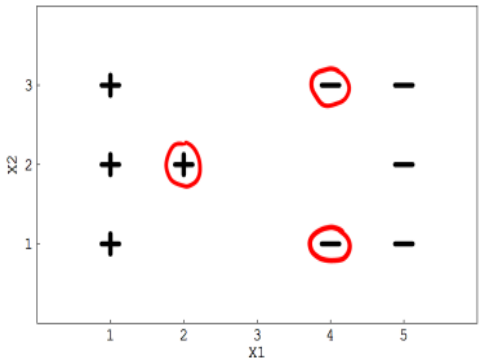


Correct Answer: a

Explanation: The blue point in the red region is an outlier (most likely noise). The rest of the data is linearly separable.

---

2. Suppose you are using a Linear SVM classifier with 2 class classification problem. Now you have been given the following data in which some points are circled red that are representing support vectors.



If you remove the following any one red points from the data. Will the decision boundary change?

a. Yes
b. No

Correct Answer: a

Explanation: These three examples are positioned such that removing any one of them introduces slack in the constraints. So, the decision boundary would completely change.

---

3. What do you mean by a hard margin in SVM Classification?

a. The SVM allows very low error in classification

b. The SVM allows high amount of error in classification

c. Both are True

d. Both are False

Correct Answer: a

Explanation: A hard margin means that an SVM is very rigid in classification and tries to work extremely well in the training set, causing overfitting.

---

4. Which of the following statements accurately compares linear regression and logistic regression?

a. Linear regression is used for classification tasks, while logistic regression is used for regression tasks.
b. Linear regression models the relationship between input features and continuous target variables, while logistic regression models the probability of binary outcomes.
c. Linear regression and logistic regression are identical in their mathematical formulation and can be used interchangeably.
d. Linear regression and logistic regression both handle multi-class classification tasks equally effectively.

Correct Answer: b

Explanation: Linear regression is employed to predict continuous numeric target variables based on input features. It finds the best-fitting linear relationship between features and the target variable. Logistic regression, on the other hand, is designed for binary classification tasks where the goal is to estimate the probability that a given input belongs to a particular class. It employs the logistic (sigmoid) function to map the linear combination of features to a probability value between 0 and 1. Linear regression and logistic regression serve different purposes and are not interchangeable due to their distinct objectives and mathematical formulations.

---

5. After training an SVM, we can discard all examples which are not support vectors and can still classify new examples?

a. True
b. False

Correct Answer: a

Explanation: Since the support vectors are only responsible for the change in decision boundary.

---

6. Suppose you are building a SVM model on data X. The data X can be error prone which means that you should not trust any specific data point too much. Now think that you want to build a SVM model which has quadratic kernel function of polynomial degree 2 that uses Slack variable C as one of it's hyper parameter.

What would happen when you use very large value of C (C->infinity)?
a. We can still classify data correctly for given setting of hyper parameter C.
b. We can not classify data correctly for given setting of hyper parameter C
c. None of the above

Correct Answer: a

Explanation: For large values of C, the penalty for misclassifying points is very high, so the decision boundary will perfectly separate the data if possible.

---

7. Following Question 6, what would happen when you use very small C (C~0)?
a. Data will be correctly classified
b. Misclassification would happen
c. None of these

Correct Answer: b

Explanation: The classifier can maximize the margin between most of the points, while misclassifying a few points, because the penalty is so low.

---

8. If g(z) is the sigmoid function, then its derivative with respect to z may be written in term of g(z) as

a. g(z)(1-g(z))
b. g(z)(1+g(z))

c. -g(z)(1+g(z))
d. g(z)(g(z)-1)

Correct Answer: a

Detailed Solution: $g'(z) = \frac{d}{dz}(\frac{1}{1+e^{-z}}) = \frac{1}{(1+e^{-z})^2}.e^{-z}$

$$= \frac{1}{1+e^{-z}}.(1 - \frac{1}{1+e^{-z}})$$

$$= g(z)(1 - g(z)$$

---

9. In the linearly non-separable case, what effect does the C parameter have on the SVM mode.

a. it determines how many data points lie within the margin
b. it is a count of the number of data points which do not lie on their respective side of the hyperplane
c. it allows us to trade-off the number of misclassified points in the training data and the size of the margin
d. it counts the support vectors

Correct Answer: c

Explanation: A high value of the C parameter results in more emphasis being given to the penalties arising out of points lying on the wrong sides of the margins. This results in reducing the number of such points being considered in deciding the decision boundary by reducing the margin.

---

10. What type of kernel function is commonly used for non-linear classification tasks in SVM?

a. Linear kernel
b. Polynomial kernel
c. Sigmoid kernel
d. Radial Basis Function (RBF) kernel

Correct Answer: d

Explanation: The Radial Basis Function (RBF) kernel is commonly used for non-linear classification tasks in SVM. It introduces non-linearity by mapping data points into a high-dimensional space, where a linear decision boundary corresponds to a non-linear decision boundary in the original feature space. The RBF kernel is suitable for capturing complex relationships and is widely used due to its effectiveness.

---

11. Which of the following statements is/are true about kernel in SVM?

1. Kernel function map low dimensional data to high dimensional space
2. It's a similarity function

a. 1 is True but 2 is False
b. 1 is False but 2 is True
c. Both are True
d. Both are False

Correct Answer: c
Explanation: Follow lecture notes

---

12. The soft-margin SVM is prefered over the hard-margin SVM when:

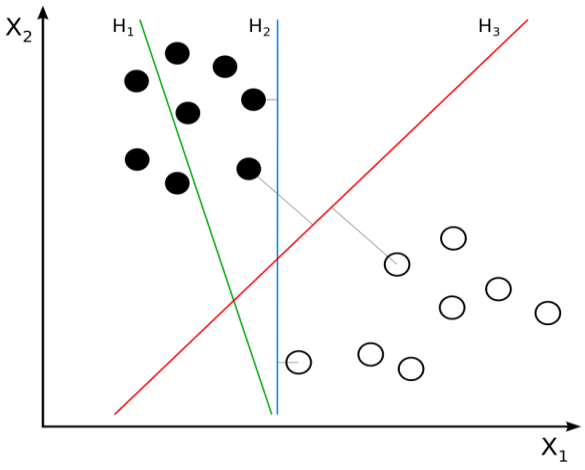
a. The data is linearly separable
b. The data is noisy
c. The data contains overlapping point

Correct Answer: b, c

Explanation: When the data has noise and overlapping points, there is a problem in drawing a clear hyperplane without misclassifying.

---

13. Consider the data-points in the figure below.



Let us assume that the black-colored circles represent positive class whereas the white-colored circles represent negative class. Which of the following among H1, H2 and H3 is the maximum-margin hyperplane?

a. H1
b. H2
c. H3
d. None of the above.

Correct Answer: c

Explanation: In a Support Vector Machine (SVM), the maximum-margin hyperplane is the one that has the largest distance between itself and the nearest data point of either class. This hyperplane ensures the best generalization to unseen data. The SVM aims to maximize this margin while still correctly classifying the training data.

To determine the maximum-margin hyperplane, you need to look for the hyperplane that has the largest "margin" between it and the nearest data point. The margin is the perpendicular distance between the hyperplane and the closest data point from either class.

H3 has the largest gap between itself and the nearest data point. That hyperplane would be the maximum-margin hyperplane.

---

14. What is the primary advantage of Kernel SVM compared to traditional SVM with a linear kernel?

a. Kernel SVM requires less computational resources.
b. Kernel SVM does not require tuning of hyperparameters.
c. Kernel SVM can capture complex non-linear relationships between data points.
d. Kernel SVM is more robust to noisy data.

Correct Answer: c

Explanation: The primary advantage of Kernel SVM is its ability to capture complex non-linear relationships between data points through the use of kernel functions. While traditional SVM with a linear kernel is limited to finding linear decision boundaries, Kernel SVM can transform the data into higher-dimensional spaces where non-linear decision boundaries can be effectively learned. This makes Kernel SVM suitable for a wide range of classification tasks where linear separation is not sufficient.

---

15. What is the sigmoid function's role in logistic regression?

a. The sigmoid function transforms the input features to a higher-dimensional space.
b. The sigmoid function calculates the dot product of input features and weights.
c. The sigmoid function defines the learning rate for gradient descent.
d. The sigmoid function maps the linear combination of features to a probability value.

Correct Answer: d

Explanation: The sigmoid function, also known as the logistic function, plays a crucial role in logistic regression. It transforms the linear combination of input features and corresponding weights into a value between 0 and 1. This value represents the estimated probability that the input belongs to a particular class. The sigmoid function's curve ensures that the output remains within the probability range, making it suitable for binary classification.

**********END**********