



Course -Introduction to Machine Learning

Assignment- Week 5 (Logistic Regression, SVM, Kernel Function, Kernel SVM)

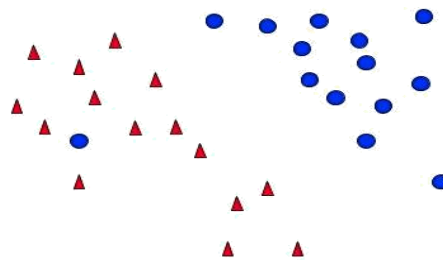
TYPE OF QUESTION: MCQ/MSQ

Number of Question: 10

Total Marks: $10 \times 2 = 20$

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 conclusion



Answer: A

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

2. I. Logistic Regression is used for regression purposes.
II. Logistic Regression is used for classification purposes.

- A) Only I is Correct
- B) Only II is Correct
- C) **Both I and II are Correct**
- D) Both I and II are Incorrect

Answer: C

Logistic Regression is used for both the classification and regression task.

3. Which of the following methods do we use to best fit the data in Logistic Regression?

- A) Least Square Error
- B) Maximum Likelihood
- C) Jaccard distance
- D) **Both A and B**



Answer: B

In logistic regression, both least square error and maximum likelihood are used as estimation methods for fitting the data.

4. Consider a following model for logistic regression: $P(y=1|x, w) = g(w_0 + w_1 x)$ where $g(z)$ is the logistic function.

In the above equation the $P(y=1|x; w)$, viewed as a function of x , that we can get by changing the parameters w .

What would be the range of P in such a case?

- A) $(-\infty, 0)$
- B) $(0, 1)$**
- C) $(-\infty, \infty)$
- D) $(0, \infty)$

Answer: B

For values of x in the range $(-\infty, +\infty)$, logistic function always give a output in the range $(0, 1)$.

5. State whether True or False.

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

- A) TRUE**
- B) FALSE

Answer: A

This is true because the support vectors only affect the boundary.

6. Suppose you are dealing with 3 class classification problem and you want to train a SVM model on the data for that you are using One-vs-all method.

How many times we need to train our SVM model in such case?

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

Answer: C



In a N-class classification problem, we have to train the SVM at least N times in a one vs all method.

7. What 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
- B) 2
- C) 1 and 2**
- D) None of these.

Answer: C

Kernels are used in SVMs to map low dimensional data into high dimensional feature space to classify non-linearly separable data. It is a similarity function between low-dimensional data points and its high dimensional feature space to find out what data points can be mapped into what sort of feature space.

8. Suppose you are using RBF kernel in SVM with high Gamma value. What does this signify?

- A) The model would consider even far away points from hyperplane for modelling.
- B) The model would consider only the points close to the hyperplane for modelling.**
- C) The model would not be affected by distance of points from hyperplane for modelling.
- D) None of the above

Answer: B

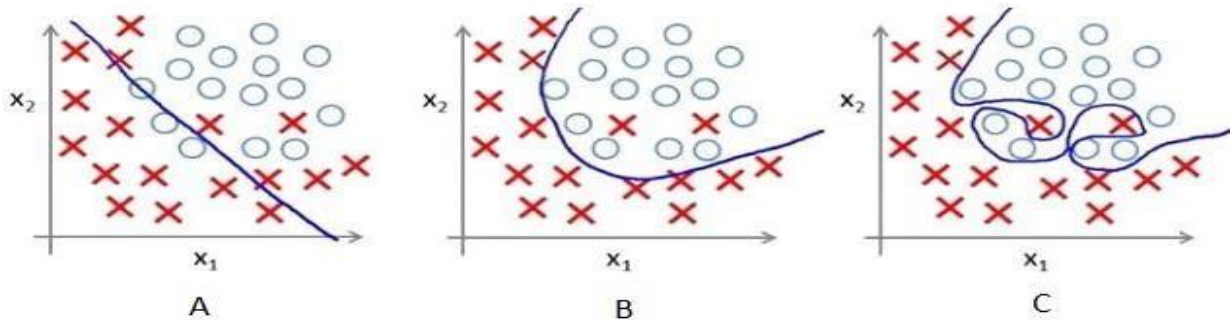
The gamma parameter in SVM tuning signifies the influence of points either near or far away from the hyperplane.

For a low gamma, the model will be too constrained and include all points of the training dataset, without really capturing the shape.

For a higher gamma, the model will capture the shape of the dataset well.

9. Below are the labelled instances of 2 classes and hand drawn decision boundaries for logistic regression. Which of the following figure demonstrates overfitting of the training data?

- A) A
- B) B
- C) C**
- D) None of these



Answer: C

In figure 3, the decision boundary is very complex and unlikely to generalize the data.

10. What do you conclude after seeing the visualization in previous question?

- C1. The training error in first plot is higher as compared to the second and third plot.
- C2. The best model for this regression problem is the last (third) plot because it has minimum training error (zero).
- C3. Out of the 3 models, the second model is expected to perform best on unseen data.
- C4. All will perform similarly because we have not seen the test data.

- A) C1 and C2
- B) C1 and C3**
- C) C2 and C3
- D) C4

Answer: B

From the visualization, it is clear that the misclassified samples are more in the plot A when compared to B. So, C1 is correct. In figure 3, the training error is less due to complex boundary. So, it is unlikely to generalize the data well. Therefore, option C2 is wrong.

The first model is very simple and underfits the training data. The third model is very complex and overfits the training data. The second model compared to these models has less training error and likely to perform well on unseen data. So, C3 is correct.

We can estimate the performance of the model on unseen data by observing the nature of the decision boundary. Therefore, C4 is incorrect

End