Course -Introduction to Machine Learning

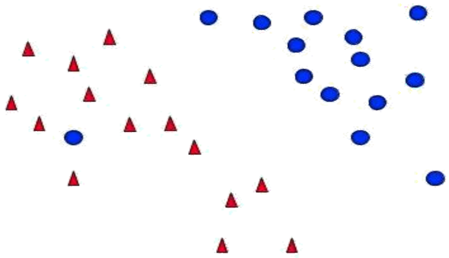
Assignment- Week 5 (Logistic Regression, SVM, Kernel Function, Kernel SVM)

TYPE OF QUESTION: MCQ/MSQ

Number of Question: 10 **Total Marks:10x2 =20**

Question 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

Correct Answer: A

Detailed Solution: The blue point in the red region is an outlier. The rest of the data is linearly separable.

Question 2:

Suppose you have a dataset with n=10 features and m=1000 examples. After training a logistic regression classifier with gradient descent, you find that it has high training error and does not achieve the desired performance on training and validation sets. Which of the following might be promising steps to take?

1. Use SVM with a non-linear kernel function
2. Reduce the number of training examples
3. Create or add new polynomial features

A) 1, 2
B) 1, 3
C) 1, 2, 3
D) None

Correct Answer: B
Detailed Solution: As logistic regression did not perform well, it is highly likely that the dataset is not linearly separable. SVM with a non-linear kernel works well for non-linearly separable datasets. Creating new polynomial features will also help in capturing the non-linearity in the dataset.

Question 3:

In logistic regression, we learn the conditional distribution p(y|x), where y is the class label and x is a data point. If h(x) is the output of the logistic regression classifier for an input x, then p(y|x) equals:

A. $h(x)^y(1 - h(x))^{(1-y)}$
B. $h(x)^y(1 + h(x))^{(1-y)}$
C. $h(x)^{1-y}(1 - h(x))^y$
D. $h(x)^y(1 + h(x))^{(1+y)}$

Correct Answer: A
Detailed Solution: Refer to the lecture.

Question 4:

The output of binary class logistic regression lies in the range:
A. [-1,0]
B. [0,1]
C. [-1,-2]
D. [1,10]

Correct Answer: B
Detailed Solution: The output of binary class logistic regression lies in the range: [0,1].

Question 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

Correct Answer: A
Detailed Solution : Using only the support vector points, it is possible to classify new

examples.

Question 6:

Suppose you are dealing with a 3-class classification problem and you want to train a SVM model on the data. For that you are using the One-vs-all method. How many times do we need to train our SVM model in such a case?

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

Correct Answer: C
Detailed Solution: In a N-class classification problem, we have to train the SVM N times in the one vs all method.

Question 7:

What is/are true about kernels in SVM?

1. Kernel function can 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.

Correct Answer: **C**
Detailed Solution: Kernels are used in SVMs to map low dimensional data into high dimensional feature space to classify non-linearly separable data. It also acts as a similarity function.

Question 8:

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

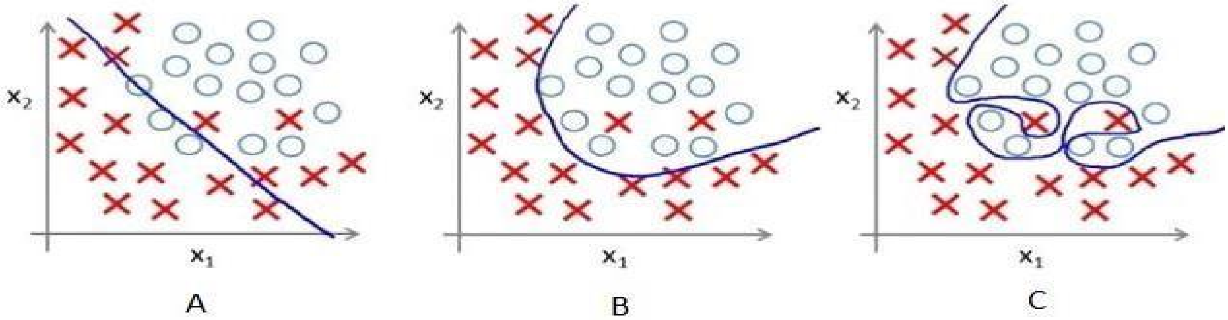
A) g(z)(g(z)-1)
B) g(z)(1+g(z))
C) -g(z)(1+g(z))
D) g(z)(1-g(z))

Correct Answer: D
Detailed Answer:

$$g'(z) = \frac{d}{dz}\left(\frac{1}{1+e^{-z}}\right) = \frac{e^{-z}}{(1+e^{-z})^2} = \frac{1}{1+e^{-z}}\left(1 - \frac{1}{1+e^{-z}}\right) = g(z)(1 - g(z))$$

Question 9:

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



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

Correct Answer: **C**
Detailed Solution: In figure 3, the decision boundary is very complex and unlikely to generalize the data.

Question 10:

What do you conclude after seeing the visualization in the previous question (Question 9)?

C1. The training error in the 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

Correct Answer: B
Detailed Solution: From the visualization, it is clear that the misclassified samples are more in the plot A when compared to B and C. So, C1 is correct. In figure 3, the training error is less due to complex boundaries. 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 is 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