Subject Code: MDS 6304

Subject Name: Deep Learning Principles and Applications

Segment 2: Linear Classifiers

1. Suppose we have 104 samples corresponding to 10 output labels and that each sample is a 32 x 32 grayscale image. What will be the shape of the weights matrix used for calculating the sample’s raw scores as z = Wx + b?
2. 11 x 1024
3. 1024 x 11
4. 1024 x 10
5. 10 x 1024

Correct Answer: d. 10 x 1024

1. Suppose we have a sample that can correspond to 5 possible output classes. If the raw scores for that sample using some weights and biases for all the classes are equal, what is the multiclass SVM hinge loss for this sample?
2. 4
3. 5
4. 0
5. 6

Correct Answer: b. 0

1. Consider a dataset in which a sample has 2 output classes (survived/not survived) and 5 features:
2. heart rate (2) blood pressure (3) temperature (4) age (5) gender.

The 3rd column of the weights matrix corresponds to what?

1. Weights for the “survived” output class
2. Weights for the “not survived” output class
3. Weights for the feature “Blood Pressure”
4. Weights for the feature “Temperature”

Correct Answer: d. Weights for the feature “Temperature”

1. The sensitivity of the loss with respect to some weight parameter evaluated at its current value is -10. This means that decreasing this weight (by a tiny amount) would \_\_\_\_\_\_\_\_ the loss
2. increase
3. decrease
4. not change
5. negate

Correct Answer: b. Increase

1. Suppose . What is ?
2. 
3. 
4. 
5. 

Correct Answer: b. 2w

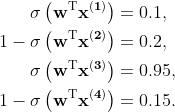
1. When performing the bias trick, the bias feature takes the value
2. 1
3. 0
4. -1
5. any fixed positive value

Correct Answer: a. 1

1. Suppose we have 103 samples corresponding to 10 output classes and that each sample is a 32 x 32 grayscale image. Suppose the samples are arranged columnwise in the data matrix and that the bias trick has been applied. What will be the shape of the data matrix?
2. 1024 x 10
3. 1025 x 1000
4. 1000 x 1025
5. 1024 x 1000

Correct Answer: 1025 x 1000

1. Consider a dataset in which a sample x can belong to either the “survived” class (labeled as 1) or the “not survived” class (labeled as 0). Suppose there are 4 patients in the dataset in which the first two survived and the last two did not. Using a particular weight vector w (bias trick applied), we get:



Without calculation, we can say that the loss is the highest for patient

1. 1
2. 2
3. 3
4. 4

Correct Answer: a. 1

1. Which if the following is not true about regularization?
2. It shrinks the weight values closer to zero
3. It helps prevent model overfitting
4. It is applied to both weights and biases
5. It helps the model to generalize well for unseen test data

Correct Answer: c. It is applied to both weights and biases

1. Suppose we apply batch processing for a dataset with 1024 samples using a batch size 16. How many times will the weights be updated in one epoch?
2. 64
3. 16
4. 32
5. 8

Correct Answer: a. 64