1. How many neurons should the hidden layer of a network with a single hidden layer and an output layer have in the context of a classification problem with 25 classes have?
A. Depends on the problem and should be determined by means of validation
B. 3
C. 10
D. 25

2. What is the resulting data after applying L1 normalization to this vector [10, 20, 30]?
A. [10, 20, 30]
B. [0.16, 0.33, 0.5]
C. [1, 2, 3]
D. [0.0, 0.5, 1.0]

3. What advantage does using a bias value bring in the context of the artificial neuron?
A. It significantly improves convergence time
B. It does not bring any advantage
C. It prevents the neuron hyperplanes from being forced to go through the origin
D. It significantly helps in the context of imbalanced data sets by provinding a bias towards the misrepresented class

4. Which of the following neuron activation is the result of the tanh activation function?
A. [0.99, 0.05, 0.99]
B. [-1.2, 0.11, 1.2]
C. [1.01, 0.11, 0.2]
D. [0.9, 0.11, -1.1]

5. What is the output of the perceptron if input=[2.4, 3.0], weights=[-0.5, 0.2], bias=1.0 (activation function - sign)?
A. 0
B. -1
C. 1
D. 2.2
6. What is the value of the loss function of a Rigde regression model if the predicted values y_hat are [-2, -3, -1], the ground-truth values are [-2, -3, -2.5], the wights are W = [1, 0], bias = 5 and alpha = 0.1?
A. 0.85
B. 0.75
C. 0.22
D. 0.95

7. If we have the following probabilities for events P(A)=0.5 P(B)=0.9 P(A|B)=0.3, what is the value of P(B|A)?
A. 0.54
B. 0.75
C. 0.63
D. 0.27

8. What is the label of the test example t = [5, 3, 8] if you apply the k-nearest neighbors classifier with k = 3 and metric = L1 (Manhattan distance) given the training data X = [[1, 4, 2], [5, 4, 8], [2, 6, 5], [1, 1, 1], [2, 9, 6]], Y = [2, 3, 3, 1, 2]?
A. 2
B. 3
C. 1
D. 0

9. In which scenario is measuring the accuracy of the model not enough to evaluate the model properly?
A. When the data set is made out of audio samples
B. When the dataset is imbalanced
C. When there are 3 classes in the dataset
D. When the data set is balanced but the training set and test set come from different sources

10. Can an SVM be used to achieve 100% training accuracy on the following 2D data set [([0, 1], 1), ([1, 0], 1), ([0, 0], 1), ([-2, 2], 0), ([2, 2], 0), ([-2, -2], 0), ([2, -2], 0)]?
A. Yes, but only if the data is normalized
B. No, because the data is not linearly separable
C. Yes, by using the kernel trick
D. No, because the dataset is imbalanced

1. Which of the following neuron activation is the result of the softmax activation function?
A. [0.6, 0.2, 0.2]
B. [0.5, 0.2, 0.2]
C. [0.6, 0.2, 0.3]
D. [0.6, -0.2, 0.2]

2. Given the following vocabulary {0 - dogs, 1 - cats, 2 - candies, 3 - likes, 4 - she, 5 - he}. What is the bag of words (BOW) representation of the sentence "she likes dogs and horses."?
A. [1, 0, 0, 1, 1, 0]
B. [2, 0, 0, 1, 1, 0]
C. [1, 0, 0, 1, 1, 0, 1, 1]
D. [1, 0, 1, 1, 1, 0]

3. How many neighbors should you consider in order to obtain the best result from a KNN classifier on the test set?
A. 1
B. 3
C. It depends on the problem and should be determined by means of validation
D. 7

4. What is the label of the test example t = [1, 2, 6] if you apply the k-nearest neighbors regressor with k = 3 and metric = L1 (Manhattan distance) given the training data X = [[1, 4, 2], [5, 4, 8], [2, 6, 5], [1, 1, 1], [2, 9, 6]], Y = [0.3, 0.6, 0.9, 0.6, 0.5]?
A. 0.6
B. 0.55
C. 0.65
D. 0.1

5. What will be the shape of the activation maps if we apply a 5x5 convolutional filter with stride=1 and no padding to a 16x16 image?
A. 14x14
B. 12x12
C. 18x18
D. 16x16

6. Suppose our model has the following metrics TP (true positives)=30, FP (false positives)=10, FN (false negatives)=30. What is the precision (P) and recall (R)?
A. P=50%, R=75%
B. P=75%, R=50%
C. P=10%, R=50%
D. P=30%, R=75%

7. How many learned parameters (weights + biases) will a network with input size = 2, hidden layer size = 5, output layer size = 1, have?
A. 10

B. 8
C. 21
D. 13

8. What type of metric can achieve 100% training accuracy on the following 2D data set [([1, 1], 1), ([5, 5], 1), ([10, 10], 1), ([5, 4], 0), ([6, 5], 0), ([6, 4], 0)] when considering a 1-NN classifier?
A. Cosine
B. None of the answers
C. L2
D. L1

9. Which of the following is a linear classifier?
A. A 3-NN classifier
B. A neuron with no activation
C. A two layer neural network with ReLU activations
D. An SVM with polynomial kernel

10. What is the value of the Mean Absolute Error function if the ground-truth labels are y = [6, 8, -9, 5] and the predicted labels are y_hat = [6.5, 7.2, 1, 7]?
A. 13.3
B. 3.325
C. 3.5
D. 13.5