1. Which of the following is a technique for using an SVM as a multi-class classifier?

A. Split group classification

B. One versus all

C. All versus all

D. N-way split

2. What is the label of the test example t = [2, 3, 5] if you apply the k-nearest neighbors classifier with k = 1 and metric = L1 (Manhattan distance) given the training data X = [[1, 4, 1], [2, 4, 7], [2, 30, 5], [0, 1, 0]], Y = [1, 3, 2, 2]?

A. 2 B. 0 C. 1 D. 3

3. 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. 0.27 B. 0.75 C. 0.63 D. 0.54

4. Which of the following is a linear classifier?

A. A two layer neural network with ReLU activations

B. A 3-NN classifier

C. An SVM with polynomial kernel

D. A neuron with no activation

5. 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. 13 D. 21

6. If the data is split into 9 classes, and we want to train a SVM for classification. How many binary classifiers will be trained in the one-vs-one approach?

A. 18 B. 9 C. 36 D. 81

7. In which scenario is measuring the accuracy of the model not enough to evaluate the model properly?

A. When the dataset is imbalanced

B. When the data set is balanced but the training set and test set come from different sources

C. When there are 3 classes in the dataset

D. When the data set is made out of audio samples

8. What is the recall of the classifier if the ground-truth labels are y = [0, 1, 1, 0, 0, 0, 0, 1] and the predicted labels are y_hat = [1, 0, 0, 0, 0, 1, 1, 1]?

A. 0.33 B. 0.23 C. 0.99 D. 0.45

9. 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, 1, 1]

B. [1, 0, 0, 1, 1, 0]

C. [1, 0, 1, 1, 1, 0]

D. [2, 0, 0, 1, 1, 0]

10. Which of the following does not constitute a valid loss for a neural network trained with gradient descent?

A. L1 Loss B. Cross Entropy C. MSE D. L2 Loss

11. What advantage does using a bias value bring in the context of the artificial neuron?

A. It significantly improves convergence time

B. It prevents the neuron hyperplanes from being forced to go through the origin

C. It significantly helps in the context of imbalanced data sets by provinding a bias towards the misrepresented class

D. It does not bring any advantage

12. The training data set contains the following examples [(3, PASS), (2, PASS), (2, PASS), (4, PASS), (0, FAIL), (1, FAIL), (3, FAIL), (1, FAIL)], the first component being the number of hours of study and the second denoting wether the student passed the exam. What is the probability of passing the exam with 2 hours of study - P(PASS|2)?

A. 25% B. 50% C. 75% D. 100%

13. What is the dimension of the weights from the second layer of a neural network with the following configuration 4-6-2-1 (the first number is the input size, the other numbers represent the amount of neurons in each layer)?

A. 6x2 B. 6x1 C. 4x6 D. 2x1

14. 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. 1 B. 2.2 C. 0 D. -1

15. What is the MSE for the following predicted labels y_pred = [0.1, 0.4, 0.7, 0.3] and truth labels=[1, 0, 1, 0]?

A. 0.3315 B. 0.1430 C. 0.0715 D. 0.2875

16. What is the difference between using an L1 loss and an L2 loss?

A. Using the L1 loss you can avoid getting stuck in a local minima when using stochastic gradient descent in the case of neural networks.

B. The L2 loss generally favors having smaller errors instead of a having fewer but greater errors while the L1 loss does not differentiate between these cases.

C. The L1 loss generally favors having smaller errors instead of a having fewer but greater errors while the L2 loss does not differentiate between these cases.

D. Using the L2 loss you can avoid getting stuck in a local minima when using stochastic gradient descent in the case of neural networks.

17. What is the resulting data after applying L1 normalization to this vector [10, 20, 30]?

A. [0.0, 0.5, 1.0]

B. [10, 20, 30]

C. [0.16, 0.33, 0.5]

D. [1, 2, 3]

18. What is the f1-score of the classifier if the ground-truth labels are y = [0, 1, 1, 0, 0, 0, 1, 1] and the predicted labels are y_hat = [1, 0, 0, 0, 0, 1, 1, 1]?

A. 0.7 B. 0.5 C. 0.6 D. 0.4

19. Which machine learning model can achieve the best performance in the context of an audio classification problem?

A. Depends on problem details and should be determined by means of validation

B. An SVM classifer

C. A Neural Network with five layers

D. A Neural Network with two layers

20. 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

21. 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]

22. 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

23. 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

24. 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

25. 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]

26. 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. 7

D. It depends on the problem and should be determined by means of validation

27. 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

28. 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

29. 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%

30. 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

31. 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

32. What is the output of neuron having sign activation for the input x = [1, -1], if the weights are W = [-1, 2], B = [1]?

A. 1 B. -1 C. 2 D. -2

33. What is the label of the test example x = [1, -1] with a 1-NN model based on the Euclidean distance having the training set S = {([2, -1], 1), ([1, 1], 2), ([-1, -1], 3)}?

A. 4 B. 3 C. 2S D. 1

34. What is the resulting data after applying min-max scaling to this data [[0.1, 0.4], [0.2, 0.5], [0.3, 0.6]] (3 examples, 2 features)?

A. [[0.0, 0.5], [0.25, 0.75], [0.5, 1.0]]

B. [[0.1, 0.4], [0.2, 0.5], [0.3, 0.6]]

C. [[0.0, 0.4], [0.25, 0.5], [0.5, 0.6]]

D. [[0.0, 0.0], [0.5, 0.5], [1.0, 1.0]]

35. Which classifier can achieve the best performance on a e-mail spam classification task?

A. A Neural Network with three layers

B. Depends on problem details and should be determined by means of validation

C. An SVM with RBF kernel

D. An SVM with linear kernel

36. What will be the shape of the activation maps if we apply a 2x2 max pooling with stride=2 to a 32x32 activation map?

A. 16x16 B. 32x32 C. 14x14 D. 28x28

37. Calculate the cost for the Ridge Regression having weights=[3, 2], alpha=0.1, y_true=[10, 1, 9, 4], y_pred=[9, 3, 6, 7].

A. 36.23 B. 23.36 C. 23.00 D. 0.10

38. Which of the following is equivalent to a single artificial neuron without activation?

A. A KNN classifier with 3 neighbors

B. A Naive Bayes classifier

C. A neural network with no activations

D. An SVM with polynomial kernel

39. After training for 5 epochs, we have the following training losses for each epoch [0.60, 0.48, 0.30, 0.28, 0.26], and the following validation losses for each epoch [0.55, 0.43, 0.27, 0.27, 0.25]. Is the model overfitted, underfitted, both, or neither?

A. Neither B. Overffiting C. Both D. Underfitting

40. What is the output of neural network with 3 hidden units and 1 output unit having ReLU activations for the input x = [1, -2], if the weights are W1 = [-0.5, 3, -2; 2, -1, 0], B1 = [0, 1, -1], W2 = [-1; -1; 2], B2 = [2]?

A. 1 B. 4.5 C. 0 D. 8

41. What is the value of PReLU(x) - parametric ReLU, where alpha=0.1 and x=-0.2?

A. -1 B. 0 C. 0.002 D. -0.02

42. If the current weights of a perceptron are [0.2, 0.4], their gradients are=[-2.4, -1.2], and the learning rate is 0.1. What are the weights after the weights update operation?

A. [0.52, 0.44]

B. [0.44, 0.52]

C. [0.44, 0.44]

D. [0.52, 0.52]

43. What is the output of SVM classifier for the input X = [0.1, -2, -5], if the weights are W = [-2, -1.2, -3] and the bias is b = 0.5?

A. 2 B. 0 C. 1 D. -1

44. What is the value of the loss function of a Ridge 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

45. 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. 1 B. 2.2 C. 0 D. -1