COMP 432 - Machine Learning
Assignment 1

Ali Hanni - 40157164

Gina Cody School of Computer Science and Software Engineering
Concordia University, Montreal, QC, Canada

Fall 2022

Exercise 1 [1 point]:

Cover's Theorem states that casting a non-linear dataset to a higher-dimensional space is more likely to result in a linearly separable set. Here, we have that $M > D$. The resulting data set might be linearly separable.

Exercise 2 [4 points]:

a

$$y_1 = w_1 \cdot x + b_1$$
$$y_1 = (2 \cdot -\tfrac{1}{2}) + (\tfrac{1}{2} \cdot -2) + 2$$
$$y_1 = -1 - 1 + 2 = 0$$

$$y_2 = w_2 \cdot x + b_2$$
$$y_2 = (3 \cdot -\tfrac{1}{2}) + (-3 \cdot -2) + 2$$
$$y_2 = -\tfrac{3}{2} + 6 + \tfrac{9}{2} = 9$$

$$y_3 = w_3 \cdot x + b_3$$
$$y_3 = (-1 \cdot -\tfrac{1}{2}) + (-\tfrac{3}{2} \cdot -2) - 2$$
$$y_3 = \tfrac{1}{2} + 3 - 2 = \tfrac{3}{2}$$

b Given that the linear discrimant for class 2 defined by w_2 and b_2 has the highest value (9), we conclude that the system would predict that input $x = (-\frac{1}{2}, -2)$ is member of the **second class**.

Exercise 3 [4 points]:

a. **TRUE**

b. Bagging reduces the accuracy of the predictions. It reduces the gap between training accuracy and testing accuracy. **Bagging reduces the testing error.**

c. sample_weight=[1, 1, 0, 2]

Exercise 4 [6 points]:

a. Backpropagation computes parameters gradient. It helps tune the neural network's weights to achieve better accuracy.

b. The number of parameters for a layer corresponds to the number of neurons in the current layer times the number of neurons in the previous layer plus the number of neurons in the current layer neurons. We can now compute the number of parameters per layer:
- Hidden layer 1: $(100 \cdot 50) + 100 = 5100$ parameters
- Hidden layer 2: $(20 \cdot 100) + 20 = 2020$ parameters
- Output layer: $(10 \cdot 20) + 10 = 210$ parameters

In total, the model has 7330 parameters.

c. Using the data provided in the graph, it is straightforward to solve a system of linear equations with 2 variables and 2 values. Here are the values found for the weights:
$w_1 = 0.5$
$b_1 = -1$
$w_2 = -0.5$
$b_2 = -1$
$w_3 = 1$
$w_4 = 1$
$b_3 = 0$

Exercise 5 [5 points]:

a. Gaussian Mixture Model is of parametric form.

b. In practice, in the EM algorithm centers can be chosen at random, at random but centered on data points, or using the K-means clustering algorithm (traditional or K-means++).

c. This is not a valid Gaussian Mixture density function. The weights for each components do not sum to 1. The first component would have a variance of 4 ($\sqrt{32/2}$. The variance should also be in the numerator next to $\sqrt{2\pi}$. It is not there, implying that the value of the numerator to cancel it should be 4. If w_1 is 4 then sum of weights can't be equal to 1 making it impossible for this function to be a Gaussian Mixture Density.