CSE 6363 - *Machine Learning*

Homework 2- Spring 2021

Due Date: Mar. 24, 2021

Feed-Forward Neural Networks

1. Consider a perceptron (a single unit with a sign function as activation function) with 3 binary inputs, x_i, and one bias input, b, that is constantly 1, and the following dataset (corresponding to a 3 input OR):

x_1	x_2	x_3	b	y
1	1	1	1	1
1	1	-1	1	1
1	-1	1	1	1
1	-1	-1	1	1
-1	1	1	1	1
-1	1	-1	1	1
-1	-1	1	1	1
-1	-1	-1	1	-1

a) Starting with weights of 0 for all connections, show 10 iterations of the Perceptron Learning rule. Each iteration here corresponds to randomly picking a data point and updating the weights accordingly. For each iteration show the sample picked, the update values (Δw resulting rom the learning rule) and the updated weights.

b) Implement the Perceptron and the perceptron learning rule for 3-dimensional binary inputs (plus a constant bias input)

c) Train your perceptron implementation from part $b)$ using the 3-dimensional OR dataset and with the following dataset representing a 3-input XOR:

x_1	x_2	x_3	b	y
1	1	1	1	-1
1	1	-1	1	1
1	-1	1	1	1
1	-1	-1	1	1
-1	1	1	1	1
-1	1	-1	1	1
-1	-1	1	1	1
-1	-1	-1	1	-1

In both cases show how the perceptron converges by showing how the number of data items that are incorrectly classified changes over time. Discuss the results and the differences between the performance on the two datasets.

2. Consider a fully-connected 2-layer feed-forward neural network with sigmoid functions as activation functions (fully connected means that every output of a unit in the first layer serves as an input to every unit in the second layer).

   a) Implement a network with 3 inputs corresponding to the features, 4 units in the first layer and one unit in the second layer (do not forget the bias weight in each of the units) as well as the error backpropagation algorithm for this network[1].

   b) Starting from randomly assigned weights in the range [-0.1 .. 0.1][2], train your network with the two datasets from Question 1 and again show how the network converges by showing how the number of data items that are incorrectly classified changes over time. Discuss the results and the differences between the performance of your network and the perceptron from Question 1.

Support Vector Machines

3. Consider the following linearly separable training data set:

$$
\begin{array}{lll}
D = \{ & ((3,4), & -1), \\
 & ((2,3), & -1), \\
 & ((2,1), & 1), \\
 & ((1,2), & 1), \\
 & ((1,3), & 1), \\
 & ((4,4), & -1) \quad \}
\end{array}
$$

[1] You only need to implement it for this specific network, not for general networks.

[2] You need to assign each weight randomly to make sure that the weits of the units are not identical.

a) Formulate the optimization function as well as the constraints for the corresponding linear maximum margin optimization problem without a regularization term. Also show the corresponding Lagrangian as well as the Lagrangian Dual for this problem.

b) Manually perform 2 iterations of the SMO algorithm on this data. You do not have to use any specific heuristic to pick the two α parameters in each iteration.

c) Use a SVM solver (e.g. MatLab's *fitcsvm* function) to learn the linear SVM parameters for this problem. Show the resulting decision boundary and identify the support vectors in this problem.