Comp 642: Assignment #2 (80 points)

Rice University — Due Date: Thursday, 03/09/2023

Submission Instructions: For coding questions, please submit python notebook along with all the plots and 2-3 paragraphs explaining what you observe and what are your conclusions. Please do familiarize yourself with python notebooks (like Jupyter or Google Collab), they are very convenient, you can run them in your browser without any installations.

1 Forward and Back-propagation in Neural Network 20 Points

You are given a data set of three samples with 2-dimensional feature vectors. $D = \{1, -2, TargetValue = 0.1\}, \{5, -9, TargetValue = 1\}, \{-3, 2, TargetValue = 0.8\}$, here $TargetValue$ is a real valued number.

You are given a neural network with one single hidden layer consisting of only 3 neurons, and each neurons has ReLU activation. The output neuron uses Sigmoid as activations $f(z) = \frac{1}{1+e^{-z}}$. Assume no bias term to any neuron. Initialize the network with a given weight matrix $W_1 = \begin{pmatrix} 1 & 0 & 1 \\ -1 & 0 & 1 \end{pmatrix}$ and $W_2 = \begin{pmatrix} 1 & 0 & -1 \end{pmatrix}$, Write down the pseudocode for forward propagation. Compute the error, and compute the gradient updates (just one step for gradient descent). The loss is mean square error. You should be able to write all these as matrix multiplications or matrix operations (like applying a function to all elements in matrix). (Some comment: Matrix Operations can be made faster with specialized instructions like vectorization and other HPC (High Performance Computing) tricks)

2 (Coding) Decision Tree and Pruning 40 Points

In zip file HW2.zip, you are provided a jupyter notebook named "DecisionTree.ipynb", along with two python files and a data folder. These files include a complete implementation of Decision Tree algorithm from scratch and some code for tree pruning. For this problem, you need to

(1) Complete the implementation of the function called "prune_algorithm" in "DecisionTree.ipynb". This function performs the tree pruning process, and it should only need a few lines of code, as most of the code and helper functions are provided in the notebook. The instructions on how to complete this function are provided in the notebook as well.

(2) After you implement "prune_algorithm", run the cells below it and get a plot that compares the prediction accuracy of the tree before pruning vs after pruning. Report what you observe and briefly explain why you think this happens.

Some reference on tree pruning: https://www.displayr.com/machine-learning-pruning-decision-trees/

We expect full working codes and plots. Please submit the notebook and all the plots. Write a short conclusion about what you observe. For this problem, you don't need to submit other provided python files or the data folder, unless you made changes to them.

3 (Coding) Compare 2-Layer Neural-Network, Logistic Regression, and XGBoost 20 Points

In zip file HW2.zip, you are provided a jupyter notebook named "NeuralNetwork.ipynb".

1. (15 Points) For this problem, you need to develop the following models on a 2D toy dataset and compare the test accuracy of each model: (1) Neural Network with two hidden layers (2) Logistic Regression (3) XGBoost. Moreover, the visualisation of decision boundary should also be plotted. You may use any ML libraries to complete this assignment.
2. (5 points) Find the model that achieves at least 95% accuracy on the toy dataset.

We expect full working codes and plots/output. Please submit the notebook and all the plots/output. Write a short conclusion about what you observe.