CS 559: Machine Learning: Fundamentals and Applications

Assignment 3 Due: 10/14/2024 Tuesday 11:59 p.m.

- The assignment must be individual work and must not be copied or shared. Any tendency to cheat/copy evidence will lead to a 0 mark for the assignment.
- Students must only use Pandas, NumPy, Matplotlib, and Scipy if the problem does not specify libraries/packages.
- All codes will be tested in grading. Any codes with an error will be marked 0. Make sure to restart the kernel and run it all before the submission. Delete any codes that do not want to be graded.
- Results must be displayed.
- All problems must be submitted in a single notebook file. Do not use a text editor to write codes.

1 Nerual Networks [60 pts]

Consider a data point, $\mathbf{x} = [0.7, 0.1, 0.3, 0.5]$ and $y = 1.5$. In this experiment, students will implement a simple neural network algorithm. In the lecture, the data point was used to observe the computational process of a neural network with a single hidden layer consisting of two neurons.

a. [5 pts] Perform a forward propagation when the network with the one hidden layer of two neurons where $\mathbf{W}^{(1)} = \begin{bmatrix} 0.16 & 0.02 & 0.63 & 0.36 \\ 0.16 & 0.25 & 0.22 & 0.29 \end{bmatrix}$ and $\mathbf{W}^{(2)} = \begin{bmatrix} 0.05 \\ 0.33 \end{bmatrix}$. The predicted value $\hat{h}$ should be the same as shown in the lecture if the learning rate $\eta = 0.1$ is used. However, it can be changed if needed.

b. [5 pts] Perform the backpropagation. Report the updated $\mathbf{W}^{(1)}$ and $\mathbf{W}^{(2)}$.

c. [10 pts] Repeat the forward and backpropagation to optimize $\mathbf{W}^{(1)}$ and $\mathbf{W}^{(2)}$. Predict y using the optimized $\mathbf{W}^{(1)}$ and $\mathbf{W}^{(2)}$.

d. [10 pts] Consider the same network as (c) except that a non-linear activation is applied. Derive analytical solution of the error matrices, $\boldsymbol{\delta}^{(2)}$ and $\boldsymbol{\delta}^{(1)}$, when a sigmoid function is applied as a non-linear activation function between the input and hidden layers. The derivation can be submitted separately is it is done on a paper.

e. [15 pts] Predict y for (d). Visualize the convergence of error and compare it to the result in c. Explain which one converged faster.

f. [15 pts] Extend the network further by increasing the number of hidden layers and neurons on hidden layers. The forward propagation can be summarized as follows:

$$h(\mathbf{x}) = a^{(3)} = \mathbf{W}^{(3)}\mathbf{z}^{(2)} = \mathbf{W}^{(3)}\Big(\sigma\big(\mathbf{W}^{(2)}\mathbf{a}^{(2)}\big)\Big) = \mathbf{W}^{(3)}\Big(\sigma\big(\mathbf{W}^{(2)}\big(\sigma\big(\mathbf{W}^{(1)}\mathbf{x}\big)\big)\big)\Big)$$

The dimension of $\mathbf{W}^{(1)}, \mathbf{W}^{(2)}$, and $\mathbf{W}^{(3)}$ must not exceed 7×7. Predict y.

2 Trainin Data Process [40 pts]

In this assignment, students will practice preparing the training data using the data **Hitters.csv**. There are no specific directions except the following.

a. Split the data into training and test sets by 8 to 2 using sklearn.model_selection.

b. [15 pts] Perform EDA using the training data and explain the findings with visualizations.

c. [15 pts] Perform feature extraction, engineering, and selection using the training data.

d. [10 pts] Repeat (c) on the test set so it has the same structure as the training set.

e. Save the data sets. They will be used in the next assignment.