**Assignment 2**

**Machine Learning**

MET CS767

Faculty - Farshid Alizadeh-Shabdiz, PhD, MBA

**Problem 1(80 points)**

Building a simple neural network using Numpy.

Using linear regression (which is a single layer Perceptron neural network) to predict value of succer players (value\_in\_M column) in “SoccerPlayers” data set using “Reaction”, “Balance”, and Strength”.

use (i)batch gradient descent and (ii) Stochastic gradient descent to adjust the weights.

1. Input: data is “SoccerPlayer” data which is available as part of the assignment.
2. Data consists five hundred rows and 14 columns.
3. Write a code and build a linear regression (which is a single layer Perceptron) as follows.

W0

W4

W1

X1

Y

X4

X3

X2

W2

W3

1. Use three features to predict “Value\_in\_M”.
2. Divide data to 80% training, and 20% test set.

from sklearn.model\_selection import train\_test\_split

x\_train, x\_test, y\_train, y\_test = train\_test\_split(x, y, test\_size=0.2, random\_state=10)

1. Assume anything that is needed to solve the problem. Make sure to state your assumptions.
2. Use Batch Gradient Descent to adjust the weights. (Write batch gradient descent code.)
   1. Plot the MSE (Mean Square Error) of the training set as a function of iteration.
   2. Plot the MSE (Mean Square Error) of the testing set as a function of iteration.
3. Use Stochastic Gradient Descent to adjust the weights (Write stochastic gradient descent code).
   1. Plot the MSE (Mean Square Error) of the training set as a function of iteration
   2. Plot the MSE (Mean Square Error) of the testing set as a function of iteration
4. Use mini-batch of size 12 to adjust the weights.
   1. Plot the MSE (Mean Square Error) of the training set as a function of iteration.
   2. Plot the MSE (Mean Square Error) of the testing set as a function of iteration.

**Problem 2 (10 points)**

What are the hyper parameters of ADAM Gradient descent and what are their impact on performance of the ADAM Gradient Descent?

**Problem 3 (10 points)**

RMSProp is Prof Geoffrey Hinton’s favorite Gradient Descent approach. Why? Name two reasons.