CIS526: Homework 4
Assigned: October 7th
Due: October 16th

Homework Policy
All assignments are INDIVIDUAL! You may discuss the problems with your colleagues, but you must solve the homework by yourself. Please acknowledge all sources you use in the homework (papers, code or ideas from someone else). Assignments should be submitted in class on the day when they are due. No credit is given for assignments submitted at a later time, unless you have a medical problem.

Problems: upload hand-written or typed-in pdf file on Canvas

Problem 1. (20 points) Let us suppose we know that label can only be positive and decide to learn prediction function of type $relu(\sum_{j=0}^{M} w_j x_j)$. Remember $relu(x) = x$ if $x \geq 0$ and $relu(x) = 0$ if $x < 0$. Show that MSE of this predictor is not a quadratic form. We decided to use stochastic gradient descent algorithm to find **w** that minimizes MSE of this predictor. Derive the update formula in a vector form. EXTRA POINT: Show that MSE has the global minimum (i.e, show that MSE is a convex function).

Problem 2. (30 points) Let us suppose we want to train on training data with *N* examples and *M* features a feedforward neural network with *H* hidden *relu* nodes in one hidden layer and one sigmoid output neuron. Assume that label *y* is a binary variable ($y = \{0,1\}$). The loss function is cross-entropy loss defined as $Loss = \sum_{i=1}^{N} y_i \log(nn(x_i, w)) + (1 - y_i)\log(1 - nn(x_i, w))$. Use backpropagation to derive gradient descent update for weights in the output neuron and for weights in the hidden layer (weight gradient in the hidden layer should be expressed using gradient of weights in the output neuron). You can derive an update for each separate weight using partial derivatives, but it is preferable if you use gradients to express the updates in the matrix form.

Problem 3. (10 points) Let us suppose we want to learn a linear predictor $\sum_{j=0}^{M} w_j x_j$ on training data with *N* examples and *M* features that minimizes loss function defined as $Loss = MSE + \alpha \sum_{j=0}^{M} {w_j}^2$, where α is a regularization hyperparameter. Show that *Loss* can be written as a quadratic form (so, there is a closed form solution for **w**). Find the expression for **w** that minimizes *Loss*. Write this expression in a matrix form. (Hint: we showed the solution in class, it led to Ridge regression)

Programming Assignment: upload the ipynb files and pdf file with the answers through Canvas

1. (20 points) You are given a function $f(\mathbf{x}) = 3x_1^2 + 2x_2^2 + 4x_1x_2 - 5x_1 + 6$.
    a. Derive analytically $\mathbf{x}^*$ that satisfies $\nabla f(\mathbf{x}) = \mathbf{0}$. Is there a unique solution for $\mathbf{x}^*$?
    b. Is $\mathbf{x}^*$ minimum or maximum of $f(\mathbf{x})$? (Help: a symmetric matrix A is positive definite if all its principal minors have strictly positive determinants)
    c. Derive gradient descent iteration formula for finding $\mathbf{x}^*$.
    d. Implement the gradient descent procedure derived in (1c) in Python. (Hint: You should write a Python function "*xfinal* = ***gd_hw3_1***(*x0*, *alpha*, *iter*)", using numpy library should be helpful). Starting from $\mathbf{x}^0 = [0\ 0]$, and using iter = 1000 iterations, run the gradient descent for several different choices of alpha = {0.0001, 0.001, 0.001, 0.01, 0.1, 1, 10}. Plot the evolution of **x** for all different choices of alpha in the same figure. Discuss the influence of alpha on the convergence. How close were the final solutions to the $\mathbf{x}^*$ obtained in (1a)?

2. (20 points) You are given a function $f(x) = \sin(x) + 0.3x$.
    a. Plot function $f(x)$ in the range x = [-10, 10]. How many points satisfy $f'(x) = 0$?
    b. Derive gradient descent iteration formula for finding a (local) minimum of $f(x)$.
    c. Implement the gradient descent procedure derived in (2b) in Python. (Hint: You should write a Python function "*xfinal* = ***gd_hw3_2***(*x0*, *alpha*, *iter*)"). Explore results of gradient descent procedure for different choices of x0 and alpha (keep the number of iterations to iter = 1000). Summarize your conclusions briefly (you could use one or two figures to illustrate the main points).

3. (40 points) Do the programming assignment in ipynb file "CIS5526 2020 - Homework4.ipynb" provided on Canvas