CIS526: Homework 3

Assigned: September 25th, 2020

Due: October 1st at 5pm

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 (each 10 points): submit as a pdf file through canvas (you can take a photo of your handwriting but save it as pdf)

Problem 1. For function $f(x_1,x_2) = x_1^2 + 2x_2^2 + 3x_1x_2 + 2x_1 + 6$ show that you can write it as the quadratic form $\mathbf{x}'Q\mathbf{x} + \mathbf{bx} + c$, where $\mathbf{x}$ is a 2x1 vector, Q is symmetric 2x2 matrix Q = [1 3/2; 3/2 2], $\mathbf{b}$ is 2x1 vector $\mathbf{b}$ = (2, 0), and c is a scalar $c = 6$.

Problem 2. Starting from $f(\mathbf{x}) = \mathbf{x}'Q\mathbf{x} + \mathbf{bx} + c$, where $\mathbf{x}$ is an nx1 vector, Q is an nxn symmetric matrix, $\mathbf{b}$ is an nx1 vector, and c is a scalar, show that $\nabla f(\mathbf{x}) = 2Q\mathbf{x} + \mathbf{b}$. **Hint**: Find partial derivatives of the quadratic form representation that uses sums:

$$f(\mathbf{x}) = \sum_{i=1}^{n} \sum_{j=1}^{n} Q_{ij} x_i x_j + \sum_{i=1}^{n} b_i x_i + c$$

Problem 3. Using the gradient $\nabla f(\mathbf{x}) = 2Q\mathbf{x} + \mathbf{b}$ from problem 2, and assuming $f(\mathbf{x})$ has the minimum, find $\mathbf{x}$ that minimizes $f(\mathbf{x})$ as a matrix formula.

Problem 4. Starting from the definition for Mean Squared Error (MSE) for linear regression (N is the number of examples and M is number of features in the training data),

$$MSE(w) = \frac{1}{N} \sum_{i=1}^{N} (y_i - \sum_{j=0}^{M} w_j x_{ij})^2$$

a) Show that MSE can be written as a quadratic form $MSE(\mathbf{w}) = \mathbf{w}'Q\mathbf{w} + \mathbf{bw} + c$, where $\mathbf{w}$ is an (M+1)-dimensional vector $\mathbf{w} = (w_0, w_1, \dots, w_M)$. Hint: treat x and y as constants, and treat w_j's as variables.

b) What are Q, $\mathbf{b}$, c?

c) Using the answers from problem 3 and problem 4.b quickly derive $\mathbf{w}$ that minimizes MSE of linear regression.

d) We decided to use gradient descent algorithm to find $\mathbf{w}$ that minimizes MSE. Derive the update formula in a vector form (**hint**: we showed it in class).

Problem 5. Assuming that X ~ Uni(0,1), a uniform random variable in range between 0 and 1, find mean and variance of random variable Y = 3X + 2.

Problem 6. The following problem is related to Bayes Theorem that is stated as P(A|B) = P(B|A)P(A)/P(B), where A and B are random variables. Let us assume A is a binary random variable stating whether a woman has breast cancer and B is a binary random variable stating whether a woman tested positive on a mammogram. Let us a assume the following background knowledge: 0.1% of women have breast cancer; 90% of women who have breast cancer test positive on mammograms; 8% of healthy women test positive on mammograms. What is the probability that a woman has cancer if she has a positive mammogram result? What is the probability that a woman has cancer if she has a negative mammogram result?

Programming Assignment (each 20 points): hand the ipynb files and pdf file with the answers through Canvas

1. 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 $\mathbf{x}$ 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. 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).