ECE/CS 559 - Fall 2018 - Homework #3
Due: 10/04/2018, the end of class.

Erdem Koyuncu

Note: All notes in the beginning of Homework #1 apply. Only a subset of the problems may be graded.

1. **Theorem:** *Let X be a real $m \times q$ matrix. There exists a unique $q \times m$ matrix X^+ with the following properties: (i) $XX^+X = X$, (ii) $X^+XX^+ = X^+$, and (iii) X^+X and XX^+ are symmetric matrices. The matrix X^+ is called the pseudo-inverse, or the Moore-Penrose pseudo-inverse of X.*

   Show that if X has linearly independent columns, then X^TX is invertible, and $X^+ = (X^TX)^{-1}X^T$. You may use the theorem above.

   Show that if X has linearly independent rows, then XX^T is invertible, and $X^+ = X^T(XX^T)^{-1}$. You may use the theorem above.

2. In this computer experiment, we will implement the gradient descent method and Newton's method. Let $f(x, y) = -\log(1 - x - y) - \log x - \log y$ with domain $\mathcal{D} = \{(x, y) : x + y < 1,\, x > 0,\, y > 0\}$.

   (a) Find the gradient and the Hessian of f on paper.

   (b) Begin with an initial point in $w_0 \in \mathcal{D}$ with $\eta = 1$ and estimate the global minimum of f using the Gradient descent method, which will provide you with points $w_1, w_2, \ldots,$. Report your initial point w_0 and η of your choice. Draw a graph that shows the trajectory followed by the points at each iteration. Also, plot the energies $f(w_0), f(w_1), \ldots,$ achieved by the points at each iteration. Note: During the iterations, your point may "jump" out of $\mathcal{D}$ where f is undefined. If that happens, change your initial starting point and/or η.

   (c) Repeat part (b) using Newton's method.

   (d) Compare the speed of convergence of gradient descent and Newton's method, i.e. how fast does each method approach the estimated global minimum?

3. Perform the following steps:

   (a) Let $x_i = i$, $i = 1, \ldots, 50$.

   (b) Let $y_i = i + u_i$, $i = 1, \ldots, 50$, where each u_i should be chosen to be an arbitrary real number between -1 and 1.

   (c) Find the linear least squares fit to (x_i, y_i), $i = 1, \ldots, 50$. Note that the linear least squares fit is the line $y = w_0 + w_1x$, where w_0 and w_1 should be chosen to minimize $\sum_{i=1}^{50}(y_i - (w_0 + w_1x_i))^2$.

   (d) Plot the points (x_i, y_i), $i = 1, \ldots, 50$ together with their linear least squares fit.

   (e) Find (on paper) the gradient of $\sum_{i=1}^{50}(y_i - (w_0 + w_1x_i))^2$ (derivatives with respect to w_0 and w_1).

   (f) (Re)find the linear least squares fit using the gradient descent algorithm. Compare with (c).

   (g) Show (on paper) that a single iteration of Newton's method with $\eta = 1$ provides the globally optimal solution (the solution in (c)) regardless of the initial point.