CS 3190: Foundations of Data Analysis **Due: Tuesday 11.16 at 11:50 pm**

Homework 4: Gradient Descent on Data and PCA

Instructions: Your answers are due at 11:50pm. You must turn in a pdf through canvas. I recommend using latex (`http://www.cs.utah.edu/~jeffp/teaching/latex/`, see also `http://overleaf.com`) for producing the assignment answers. If the answers are too hard to read you will lose points, entire questions may be given a 0 (e.g. **sloppy pictures with your phone's camera are not ok, but very careful ones are**)

Please make sure your name appears at the top of the page.

You may discuss the concepts with your classmates, but write up the answers entirely on your own. **Be sure to show all the work involved in deriving your answers! If you just give a final answer without explanation, you may not receive credit for that question.**

We will use two datasets, here: `http://www.cs.utah.edu/~jeffp/teaching/FoDA/X4.csv`, here `http://www.cs.utah.edu/~jeffp/teaching/FoDA/Y4.csv`, and here:
`http://www.cs.utah.edu/~jeffp/teaching/FoDA/A.csv`
There are many ways to import data in python, the `genfromtext` command in numpy provides an easy solution.

1. **[50 points]** Using data set `X4.csv` use these $n(= 30)$ rows as the explanatory variables $x \in \mathbb{R}^3$ in a linear regression problem. Note the first column is always 1, so you do not need to deal specially with the offset. Then use data set `Y4.csv` as the corresponding dependent y value. On parts (c) and (f) of this problem, you will run gradient descent on $\alpha \in \mathbb{R}^3$, using the dataset provided to find a linear model

   $$\hat{y} = \alpha_0 x_0 + \alpha_1 x_1 + \alpha_2 x_2 = \alpha_0 + \alpha_1 x_1 + \alpha_2 x_2$$

   minimizing the sum of squared errors. You will run for as many steps as you feel necessary. On parts (c) and (f) of this problem, on each step of your gradient descent run, print on a single line: (i) the value of a function f, estimating the sum of squared errors, and (ii) the norm of the gradient of f, and (iii) the parameters you found ($[\alpha_0, \alpha_1, \alpha_2]$) at that step. For notation purposes, let y_i refer to the ith entry of y, and for $j \in \{0, 1, 2\}$, let x_{ji} refer to the ith entry of the jth explanatory variable x_j
   [An earlier posted version had indexes of α posted below starting at 1 as $(\alpha_1, \alpha_2, \alpha_3)$ instead of $(\alpha_0, \alpha_1, \alpha_2)$. No numerical values were updated.

   (a) For the batch gradient descent method, write down a function $f(\alpha_0, \alpha_1, \alpha_2) = ?$ that evaluates the loss function. After filling in ?, report the value of the loss function at $(\alpha_0 = 1, \alpha_1 = 1, \alpha_2 = 1)$ and also at $(\alpha_0 = 3, \alpha_1 = 2, \alpha_2 = 4)$.

   (b) For the function you wrote down in part (a), write down the gradient function. i.e $\nabla f(\alpha_0, \alpha_1, \alpha_2) = ?$ that evaluates the gradient of the loss function. After filling in ?, report the value of ∇f at $(\alpha_0 = 1, \alpha_1 = 1, \alpha_2 = 1)$ and $(\alpha_0 = 3, \alpha_1 = 2, \alpha_2 = 4)$

   (c) Now run batch gradient descent (a batch size of all n points).

(d) For the incremental gradient descent method, write down a function $f_i(\alpha_0, \alpha_1, \alpha_2) = ?$ that evaluates the loss function for the ith data point. After filling in ?, report the value of the loss function for $i = 1$ at $(\alpha_0 = 1, \alpha_1 = 1, \alpha_2 = 1)$ and also at $(\alpha_0 = 3, \alpha_1 = 2, \alpha_2 = 4)$.

(e) For the function you wrote in part (d), write down the gradient function, i.e $\nabla f_i(\alpha_0, \alpha_1, \alpha_2) = ?$ that evaluates the gradient of the loss function for the ith data point. After filling in ?, report the value of ∇f_i for $i = 1$ at $(\alpha_0 = 1, \alpha_1 = 1, \alpha_2 = 1)$ and $(\alpha_0 = 3, \alpha_1 = 2, \alpha_2 = 4)$

(f) Now run incremental gradient descent.

Choose one method which you preferred (either is ok to choose), and explain why you preferred it to the other method.

2. **[20 points]**

Consider a matrix $A \in \mathbb{R}^{200\times 10}$ and its SVD $[U, S, V^T] = \mathsf{svd}(A)$. Answer the following questions.

(a) True or False, the *first* right singular vector v_1 of A is the direction in $\mathbb{R}^{10}$ satisfying $v_1 = \arg\max_{x\in\mathbb{R}^{10}, \|x\|=1} \|Ax\|_2$

(b) True or False, a left singular vector of A is a direction in $\mathbb{R}^{10}$

(c) True or False, Suppose s_3 is the third singular value of A. Then s_3^2 is the third largest eigenvalue of AA^T and A^TA.

Let u_1, u_2 be the first two left singular vectors; let v_1, v_2 be the first two right singular vectors; and let s_1, s_2 be the first two singular values. Consider $B = s_1u_1v_1^T + s_2u_2v_2^T$.

(d) What is the rank of B?

(e) What is dimension of B?

(f) Let v_3 be the third right singular vector. What is $\|Bv_3\|$?

3. **[30 points]** Read data set `A.csv` as a matrix $A \in \mathbb{R}^{35\times 8}$. Compute the SVD of A and report

(a) the fourth singular value, and

(b) the rank of A?

Compute the eigendecomposition of AA^T.

(c) For every non-zero eigenvalue, report it and its associated eigenvector. How many non-zero eigenvalues are there?

Compute A_k for $k = 3$.

(d) What is $\|A - A_k\|_F^2$?

(e) What is $\|A - A_k\|_2^2$?

Center A. Run PCA to find the best 3-dimensional subspace F to minimize $\|A - \pi_F(A)\|_F^2$. Report

(f) $\|A - \pi_F(A)\|_F^2$ and

(g) $\|A - \pi_F(A)\|_2^2$.