CS 3190: Foundations of Data Analysis **Due: Wednesday 11.20 at 2:45 pm**

Homework 4: Gradient Descent on Data and PCA

Instructions: Your answers are due at 2:45, before the beginning of class on the due date. You must turn in a pdf through canvas. I recommend using latex (`http://www.cs.utah.edu/~jeffp/teaching/latex/`) 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. **[40 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. 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. Run for as many steps as you feel necessary. On each step of your 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.

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

   (b) Second run incremental gradient descent.

2. **[20 points]**

   Consider a matrix ~~B~~ $A \in \mathbb{R}^{100 \times 8}$ and its SVD $[U, S, V^T] = \mathsf{svd}(A)$ ~~$\mathsf{svd}(B)$~~. Answer the following questions.

   (a) True or False, the *second* right singular vector of ~~B~~ A is the direction in $\mathbb{R}^8$ with the *second* most variance.

   (b) Derive the third eigenvalue of AA^T from $S_{3,3}$, the third singular value of A.

   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 $B = s_1 u_1 v_1^T + s_2 u_2 v_2^T$.

(c) What is the rank of B?

(d) What is dimension of B?

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

3. **[40 points]** Read data set `A.csv` as a matrix $A \in \mathbb{R}^{30 \times 6}$. Compute the SVD of A and report

   (a) the third singular value, and

   (b) the rank of A?

   Compute the eigendecomposition of $A^T A$.

   (c) Report all of the eigenvectors and eigenvalues.

   Compute A_k for $k = 2$.

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

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

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

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

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