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/D4.csv and here: http://www.cs.utah.edu/~jeffp/teaching/FoDA/A.csv

There are many ways to import data in python (see Canvas for a discussion). The pandas package seems to be the most general one.

1. [25 points] In the first D4.csv dataset provided, use the first three columns as explanatory variables $x_1, x_2, x_3$, and the fourth as the dependent variable $y$. Run gradient descent on $\alpha \in \mathbb{R}^4$, using the dataset provided to find a linear model

$$\hat{y} = \alpha_0 + \alpha_1 x_1 + \alpha_2 x_2 + \alpha_3 x_3$$

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 parameters you found $([\alpha_0, \alpha_1, \alpha_2, \alpha_3])$ at that step. (These are the sort of things you would do to check/debug a gradient descent algorithm; you may also want to plot the function value and norm of the gradient.)

(a) First run batch gradient descent.

(b) Second run incremental gradient descent.

2. [10 points] Explain what parts of the above procedures would change if you instead are minimizing the sum of residuals, not the sum of square residuals?

- Is the function still convex?
- Is incremental gradient descent still possible?
- Is the gradient more or less complex for batch gradient descent?

3. [25 points]

Consider two matrices $A_1$ and $A_2$ both in $\mathbb{R}^{4 \times 3}$. $A_1$ has singular values $\sigma_1 = 20$, $\sigma_2 = 2$, and $\sigma_3 = 1.5$. $A_2$ has singular values $\sigma_1 = 8$, $\sigma_2 = 4$, and $\sigma_3 = 0.001$. 


(a) For which matrix will the power method converge faster to the top eigenvector of \(A_1^TA_1\) (or \(A_2^TA_2\), respectively), and why?

Given the eigenvectors \(v_1, v_2, v_3\) of \(A_1^TA_1\). Explain step by step how to recover the following. [You should write the answers as linear algebraic expressions in terms of \(v_1, v_2, v_3\), and \(A_1\).]

(b) the singular values of \(A_1\),
(c) the right singular vectors of \(A_1\), and
(d) the left singular vectors of \(A_1\).

4. [40 points] Read data set \(A\texttt{.csv}\) as a matrix \(A \in \mathbb{R}^{30 \times 6}\). Compute the SVD of \(A\) and report

(a) the third right singular vector,
(b) the second singular value, and
(c) the fourth left singular vector.
(d) What is the rank of \(A\)?

Compute \(A_k\) for \(k = 2\).

(e) What is \(\|A - A_k\|_F^2\) ?
(f) 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

(g) \(\|A - \pi_F(A)\|_F^2\) and
(h) \(\|A - \pi_F(A)\|_2^2\).