Homework EXTRA: Gradient Descent on Data and PCA

Note: These points will be added to your score for homework 4 (or any other homework assignment if you specify clearly). You can earn more than 100/100 points on this assignment.

Instructions: Your answers are due at 11:59pm. 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 one dataset, here: http://www.cs.utah.edu/~jeffp/teaching/FoDA/D4redo.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. [14 points] In the D4redo.csv dataset provided, use the first two columns as explanatory variables $x_1, x_2$, and the third as the dependent variable $y$. Run gradient descent on $\alpha \in \mathbb{R}^3$, using the dataset provided to find a linear model

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

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 parameters you found ($[\alpha_0, \alpha_1, \alpha_2]$) 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.

[Suggestion: initialize $\alpha = (0, 0, 0)$, and use learning rates 0.003 for (a) and 0.04 for (b)].

2. [6 points]

Consider a matrix $A$ in $\mathbb{R}^{10 \times 3}$. Given the eigenvectors $v_1, v_2, v_3$ of $A^T A$. Explain step by step how to recover the following. Specifically, you should write the answers as linear algebraic expressions in terms of $v_1, v_2, v_3$, and $A$; it can involve taking norms, matrix multiply, addition, subtraction, but not something more complex like SVD.

(b) the second singular value of $A$
(c) the first right singular vector of $A$
(d) the third left singular vector of $A$