CMSC 25300 / 35300
Extra Credit

1. Answer the following questions. Make sure to explain your reasoning.

a) Are the columns of the following matrix linearly independent? Are they orthogonal?

$$\boldsymbol{A} = \begin{bmatrix} +0.5 & +0.5 \\ -0.5 & +0.5 \\ +0.5 & -0.5 \\ -0.5 & -0.5 \end{bmatrix}$$

b) Are the columns of the following matrix linearly independent?

$$\boldsymbol{A} = \begin{bmatrix} +1 & +1 & +1 \\ -1 & +1 & -1 \\ +1 & -1 & -1 \end{bmatrix}$$

c) What is the rank of the following matrix?

$$\boldsymbol{A} = \begin{bmatrix} +2 & +1 \\ -2 & +1 \\ +2 & -1 \end{bmatrix}$$

d) Suppose the matrix in part c is used in the least squares optimization $\min_{\boldsymbol{x}} \|\boldsymbol{b} - \boldsymbol{A}\boldsymbol{x}\|_2$. Does a unique solution exist?

2. A biological experiment measures the abundance of two different proteins in three different conditions. The goal of the problem is to predict the presence/absence of certain *enzyme* based on the abundances of the two proteins in each of the experimental conditions. The data for enzyme whose presence we are trying to predict are

$$\boldsymbol{b} = \begin{bmatrix} +1 \\ -1 \\ -1 \end{bmatrix},$$

where $+1$ indicates the presence of the enzyme and -1 its absence, in each of the three experimental conditions. The data for the proteins whose abundances we will use to predict $\boldsymbol{b}$ are

$$\boldsymbol{A} = \begin{bmatrix} +0.5 & +0.5 \\ -0.5 & 0 \\ 0 & -0.5 \end{bmatrix}.$$

Each column of $\boldsymbol{A}$ corresponds to one of the two proteins. The protein abundances are centered so that 0 is the average level, positive values indicate increased levels, and negative values indicate decreased values.

a) Find vector $\boldsymbol{x}^\star \in \mathbb{R}^2$ so that $\text{sign}(\boldsymbol{A}\boldsymbol{x}^\star) = \boldsymbol{b}$? Hint: You should be able to do this by inspection, without resorting to any complicated calculations.

b) Let $\widehat{\boldsymbol{x}}$ be the solution to $\min_{\boldsymbol{x}\in\mathbb{R}^2} \|\boldsymbol{b} - \boldsymbol{A}\boldsymbol{x}\|_2$.

i. Compute $\widehat{\boldsymbol{x}}$ and $\widehat{\boldsymbol{b}}$.

ii. Sketch a picture in $\mathbb{R}^3$ that depicts $\boldsymbol{b}$, the columns of $\boldsymbol{A}$, and $\widehat{\boldsymbol{b}}$.

iii. How would you use $\widehat{\boldsymbol{b}}$ to predict the presence/absence of the enzyme?

3. Reconsider the same biological experiment in Problem 1. Suppose the data for the proteins used for prediction are different:

$$\boldsymbol{A} = \begin{bmatrix} +0.5 & +0.5 \\ -0.5 & +0.5 \\ +0.5 & -0.5 \end{bmatrix}.$$

but the data for enzyme are the same:

$$\boldsymbol{b} = \begin{bmatrix} +1 \\ -1 \\ -1 \end{bmatrix},$$

a) Can you find a vector $\boldsymbol{x}^\star \in \mathbb{R}^2$ so that $\text{sign}(\boldsymbol{A}\boldsymbol{x}^\star) = \boldsymbol{b}$? Explain and fully justify your answer.

b) In this case, it appears that the presence/absence of the enzyme may be influenced by a nonlinear interaction between the two protein levels.

i. Can you suggest a new, third feature to add to the model that can capture such nonlinear interactions between the two proteins? This will give you a new matrix, $\widetilde{\boldsymbol{A}}$, with an additional column. What are the values in the $\widetilde{\boldsymbol{A}}$ matrix?

ii. Can you find a vector $\widetilde{\boldsymbol{x}}^\star \in \mathbb{R}^3$ so that $\text{sign}(\widetilde{\boldsymbol{A}}\,\widetilde{\boldsymbol{x}}^\star) = \boldsymbol{b}$?

iii. Let $\widetilde{\boldsymbol{x}}$ be the solution to $\min_{\boldsymbol{x}\in\mathbb{R}^3} \|\boldsymbol{b} - \widetilde{\boldsymbol{A}}\,\boldsymbol{x}\|_2$. Compute $\widetilde{\boldsymbol{x}}$ and $\widetilde{\boldsymbol{b}} = \widetilde{\boldsymbol{A}}\,\widetilde{\boldsymbol{x}}$.

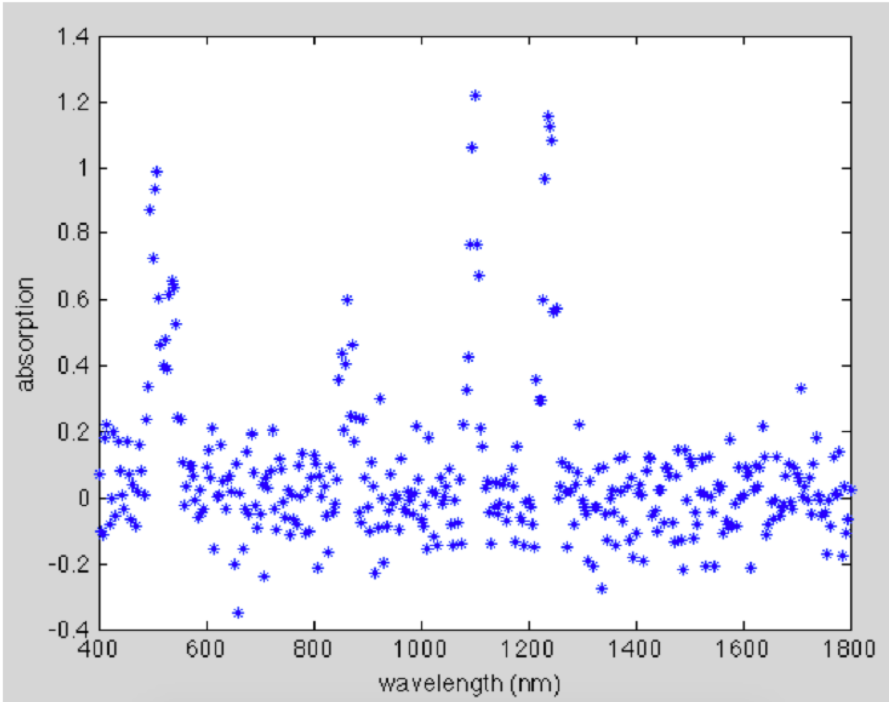
iv. Sketch a picture of this least squares problem, showing $\boldsymbol{b}$, the columns of the new $\widetilde{\boldsymbol{A}}$, and $\widetilde{\boldsymbol{b}}$.

4. Spectroscopy is a measurement technique based on the interaction of light and matter. It can be used to determine the molecules present in a sample of material or gas. The figure below depicts a simulation of the absorption data. Each point denotes measured absorption $a(\omega)$ at the corresponding wavelength ω. The peaks indicate the presence of particular molecules.

There are 5 possible molecules present in the sample. The (normalized) ideal absorption profiles are as follows:

molecule 1: $a_1(\omega) = \exp(-0.1(\omega - 504)^2)$

molecule 2: $a_2(\omega) = \exp(-0.1(\omega - 536)^2)$



molecule 3: $a_3(\omega) = \exp(-0.1(\omega - 676)^2)$
molecule 4: $a_4(\omega) = \exp(-0.1(\omega - 864)^2)$
molecule 5: $a_5(\omega) = \exp(-0.1(\omega - 1100)^2)$

The data are noisy, and don't perfectly agree with these ideals.

a) Explain how you would use least squares to estimate the absorption levels of each molecule.

b) How would you decide if a particular molecule was present or absent from the sample?

c) How would you form of the orthogonal projection on to the space spanned by the ideal absorption profiles?

5. Ridge regression

a) Write the equation for gradient descent updates for the Tikhinov / ridge regression problem

$$\hat{\boldsymbol{w}} = \arg\min_{\boldsymbol{w}} \|\boldsymbol{y} - \boldsymbol{X}\boldsymbol{w}\|_2^2 + \lambda\|\boldsymbol{w}\|_2^2.$$

b) Sometimes we use a penalty of the form $\|\boldsymbol{D}\boldsymbol{w}\|_2^2$ where $\boldsymbol{D}$ is a matrix defined such that the i-th element is $(\boldsymbol{D}\boldsymbol{w})_i = \boldsymbol{w}_i - 2\boldsymbol{w}_{i-1} + \boldsymbol{w}_{i-2}$. What is the matrix $\boldsymbol{D}$?

c) Give an expression for $\hat{\boldsymbol{w}}$ where

$$\hat{\boldsymbol{w}} = \arg\min_{\boldsymbol{w}} \|\boldsymbol{y} - \boldsymbol{X}\boldsymbol{w}\|_2^2 + \lambda\|\boldsymbol{D}\boldsymbol{w}\|_2^2$$

in terms of D.

d) Implement this new kind of Tikhinov regularization in the deblurring problem considered in Homework 5. How do the results compare with what you got last week?

e) Write the equation for gradient descent updates for the Tikhinov / ridge regression problem with $\boldsymbol{D}$.

f) Now implement your estimator on the blurring problem using gradient descent and show that you get almost the same solution as with the closed-form solution.

6. Gradient descent convergence. Consider the iterations for solving a standard least-squares problem with $\boldsymbol{X} \in \mathbb{R}^{m \times n}$, $\boldsymbol{y} \in \mathbb{R}^m$, and $\boldsymbol{X}$ has full column rank. Recall that this iteration begins with some initial $\boldsymbol{w}^{(1)}$ and then:

$$\boldsymbol{w}^{(k+1)} = \boldsymbol{w}^{(k)} - \tau \boldsymbol{X}^\top (\boldsymbol{X}\boldsymbol{w}^{(k)} - \boldsymbol{y}) \qquad \text{for } k = 0, 1, \ldots \tag{1}$$

a) We expect the algorithm to converge to $\boldsymbol{w}^\star = (\boldsymbol{X}^\top \boldsymbol{X})^{-1} \boldsymbol{X}^\top \boldsymbol{y}$. Define the *error* as $\boldsymbol{e}^{(k)} := \boldsymbol{w}^{(k)} - \boldsymbol{w}^\star$. Show how to rewrite (1) in the form $\boldsymbol{e}^{(k+1)} = \boldsymbol{P}\boldsymbol{e}^{(k)}$. What is the matrix $\boldsymbol{P}$?

b) Define the *residual* $\boldsymbol{r}^{(k)} := \boldsymbol{X}\boldsymbol{w}^{(k)} - \boldsymbol{y}$. Show how to rewrite (1) in the form $\boldsymbol{r}^{(k+1)} = \boldsymbol{Q}\boldsymbol{r}^{(k)}$. What is the matrix $\boldsymbol{Q}$?

c) Let $\{\sigma_i\}$ be the singular values of $\boldsymbol{X}$. Prove that when $0 < \tau < \frac{2}{\sigma_1^2}$, we have $\lim_{k \to \infty} \boldsymbol{e}^{(k)} = \boldsymbol{0}$. **Hint:** substitute the SVD of $\boldsymbol{X}$ into your expression for $\boldsymbol{P}$.

d) **Extra extra credit:** Prove that if $\boldsymbol{X}$ is rank-deficient and $\boldsymbol{w}^{(1)} = \boldsymbol{0}$, then the gradient descent iterations converge to the minimum norm solution. **Hint:** redo part **a)** using $\boldsymbol{w}^\star = \boldsymbol{X}^\dagger \boldsymbol{y}$ and see how this affects part **c)**.