CS/ECE/ME 532
Homework 3: Subspaces and orthogonality

1. **Orthogonal columns**. Consider the matrix and vector

$$\boldsymbol{A} = \begin{bmatrix} 3 & 1 \\ 0 & 3 \\ 0 & 4 \end{bmatrix} \quad \text{and} \quad \boldsymbol{b} = \begin{bmatrix} 1 \\ 3 \\ 1 \end{bmatrix}.$$

   a) By hand, find two orthonormal vectors that span the plane spanned by columns of $\boldsymbol{A}$.

   b) Make a sketch of these vectors and the columns of $\boldsymbol{A}$ in three dimensions.

   c) Use these vectors to compute the LS estimate $\widehat{\boldsymbol{b}} = \boldsymbol{A}(\boldsymbol{A}^T\boldsymbol{A})^{-1}\boldsymbol{A}^T\boldsymbol{b}$.

2. **Tikhonov regularization**. Sometimes we have competing objectives. For example, we want to find an x that minimizes $\|\boldsymbol{b} - \boldsymbol{A}\boldsymbol{x}\|_2$ (least-squares), but we also want the weights $\boldsymbol{x}$ to be small. One way to achieve a compromise is to solve the following problem:

$$\text{minimize} \quad \|\boldsymbol{b} - \boldsymbol{A}\boldsymbol{x}\|_2^2 + \lambda \|\boldsymbol{x}\|_2^2 \tag{1}$$

   where $\lambda > 0$ is a parameter we choose that determines the relative weight we want to assign to each objective. This is called *Tikhonov regularization* (also known as L_2 regularization).

   a) Solve the optimization problem (1) by finding an expression for the minimizer $\widehat{\boldsymbol{x}}$.
   Hint: one approach is to reformulate (1) as a modified least-squares problem with different "$\boldsymbol{A}$" and "$\boldsymbol{b}$" matrices. Another approach is to use the vector derivative method we saw in class.

   b) Suppose that $\boldsymbol{A} \in \mathbb{R}^{m \times n}$, with $m < n$. Is there a unique least squares solution? Is there a unique solution to (1) ? Explain your answers.

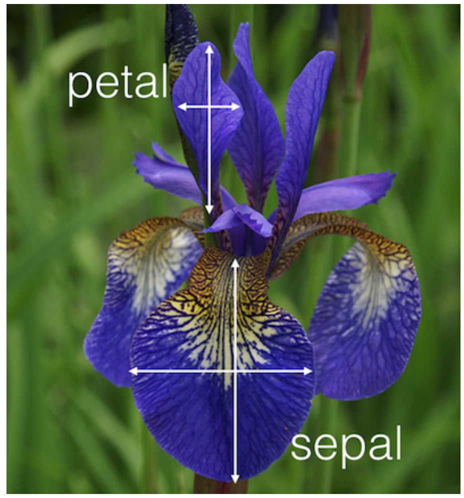
3. **Gram-Schmidt**. Write your own code to perform Gram-Schmidt orthogonalization. Your code should take as input a matrix $\boldsymbol{A} \in \mathbb{R}^{m \times n}$ and return as output a matrix $\boldsymbol{U} \in \mathbb{R}^{m \times r}$ where $\boldsymbol{U}$ is orthogonal and has the same range as $\boldsymbol{A}$. Note that r will indicate the rank of $\boldsymbol{A}$, so your code can also be used to find the rank of a matrix!

   a) Test your code by applying it to Problem 1 above.

   b) Use your code to determine the rank of the following matrices and compare the result to Matlab's `rank` function (or Python's `numpy.linalg.matrix_rank` function).

$$\boldsymbol{A}_1 = \begin{bmatrix} 3 & 1 & 2 \\ 0 & 3 & 3 \\ 0 & 4 & 4 \\ 6 & 1 & 4 \end{bmatrix} \qquad \boldsymbol{A}_2 = \begin{bmatrix} 1 & 1 & 2 \\ 0 & 3 & 3 \\ 0 & 4 & 4 \\ 3 & 1 & 4 \end{bmatrix}$$

4. In 1936 Ronald Fisher published a famous paper on classification titled "The use of multiple measurements in taxonomic problems." In the paper, Fisher study the problem of classifying iris flowers based on measurements of the sepal and petal widths and lengths, depicted in the image below.



Fisher's dataset is available in Matlab (`fisheriris.mat`) and is widely available on the web (e.g., Wikipedia). The dataset consists of 50 examples of three types of iris flowers. The sepal and petal measurements can be used to classify the examples into the three types of flowers.

a) Formulate the classification task as a least squares problem. Least squares will produce real-valued predictions, not discrete labels or categories. What might you do to address this issue? round

b) Write a Matlab or Python program to "train" a classifier using LS based on 40 labeled examples of each of the three flower types, and then test the performance of your classifier using the remaining 10 examples from each type. Repeat this with many different randomly chosent subsets of training and test. What is the average test error (number of mistakes divided by 30)?

c) Experiment with even smaller sized training sets. Clearly we need at least one training example from each type of flower. Make a plot of average test error as a function of training set size.

d) Now design a classifier using only the first three measurements (sepal length, sepal width, and petal length). What is the average test error in this case?

e) Use a 3d scatter plot to visualize the measurements in (d). Can you find a 2-dimensional subspace that the data approximately lie in? You can do this by rotating the plot and looking for plane that approximately contains the data points.

f) Use this subspace to find a 2-dimensional classification rule. What is the average test error in this case?