CS/ECE/ME 532
Homework 4: orthogonality and more least squares

due: Friday October 14, 2016

1. **Range-Nullspace duality.** Suppose $A \in \mathbb{R}^{m \times n}$ is any matrix. Further suppose that $p \in \text{range}(A)$ and $q \in \text{null}(A^\mathsf{T})$. Prove that $p^\mathsf{T} q = 0$.

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

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

   a) By hand, find two orthonormal vectors that span the plane spanned by columns of A.

   b) Make a sketch of these vectors and the columns of A in three dimensions.

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

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

   a) Test your code by applying it to Problem 2 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, etc.).

$$A_1 = \begin{bmatrix} 3 & 1 & 2 \\ 0 & 3 & 3 \\ 0 & 4 & 4 \\ 6 & 1 & 4 \end{bmatrix} \qquad A_2 = \begin{bmatrix} 1 & 1 & 2 \\ 0 & 3 & 3 \\ 0 & 4 & 4 \\ 3 & 1 & 4 \end{bmatrix}$$

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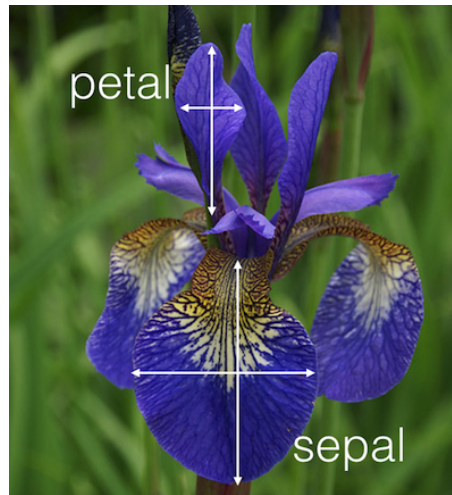
Fisher's dataset is available in Matlab (`fisheriris.mat`) and is widely available on the web (e.g., Wikipedia). The dataset consists of 50 examples of three types of iris flowers. The sepal and petal measurements can be used to classify the examples into the three types of flowers.

a) Formulate the classification task as a least squares problem. For the labels, use setsoa $= -1$, versicolor $= 0$, virginica $= 1$. Then, classify based on which of the labels $\widehat{b}$ is closest to. We'll define the *average classification error* to be:

$$\frac{\text{number of misclassified examples}}{\text{total number of classified examples}}$$

What is the average classification error when your classifier is used on the entire data set?

b) Let's use cross-validation this time. Write code to train a LS classifier based on 40 labeled examples of each of the three flower types, and then test the performance (by computing the average classification error) on the remaining 10 examples from each type. Repeat this with 10,000 different randomly chosen subsets of training and test. What is the average of all classification errors computed?

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 classification 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 classification 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 classification error in this case?