Homework # 9

1. Reading

   - Section 4.3 in Sauer discusses the QR decomposition.
   - Here is some optional reading describing PCA using R. In this hw, I ask you to do the PCA yourself, but typically you would call R's pca function.

     `http://www.r-bloggers.com/computing-and-visualizing-pca-in-r/`

2. Below, let A be an $n \times k$ matrix. Define the span of A, written span(A), as the span of the column vectors of A. In class we discussed Gram-Schmidt (GS) orthogonalization. Here I just want you to go through and finish the arguments I made in class.

   (a) Given a matrix A, write down the GS iteration that will produce an orthogonal matrix Q with span$(Q) =$ span(A).

   (b) Prove that the GS iteration you wrote down in (a) produces orthonormal vectors with the correct span (see Sauer if you get stuck).

   (c) Write an R function **GramSchmidt(A)** which returns the matrix Q in (a). Check that your function works and compare to the result of using R's **qr** function for some non-trivial choice of A.

3. This problem covers some of the technical details we covered in discussing PCA.

   (a) Let v be a vector in $\mathbb{R}^n$ with $\|v\| = 1$ and consider $\Omega =$ span(v). Show that for any $x \in \mathbb{R}^n$, the projection of x onto Ω is given by $(v \cdot x)v$. Explain why computing the projection reduces to determining c in the following minimization:
   $$\min_{c \in \mathbb{R}} \|x - cv^{(1)}\|^2 \tag{1}$$
   (Hint: Define $f(c) = \|x - cv^{(1)}\|^2$. Solve for c by solving $f'(c) = 0$.)

(b) Now repeat (a), but this time let $v^{(1)}$, $v^{(2)}$ be two orthonormal vectors in $\mathbb{R}^n$. Then, $\Omega = \text{span}(v^{(1)}, v^{(2)})$. Show, that the projection of a $x \in \mathbb{R}^n$ onto Ω is given by $(v^{(1)} \cdot x)v^{(1)} + (v^{(2)} \cdot x)v^{(2)}$. In this case you need to consider c_1, c_2 rather than the single c of part (a).

(c) Let M be an $n \times n$ symmetric matrix. Show that the following maximization,

$$\max_{v \in \mathbb{R}^n, \|v\|=1} v^T M v, \tag{2}$$

is solved by setting v equal to the dominant eigenvector of M. (Hint: Expand v in the eigenvector basis of M. Plug the expansion into the v in $v^T M v$ and simplify using the orthonormality of the eigenvectors. Also plug the expansion into $\|v\| = 1$ and see what the constraint implies for the coefficients of the eigenvectors in the expansion.)

(d) Now you will use (a)-(c) to derive the PCA results we discussed in class, but this time with the details filled in. Let $x^{(i)} \in \mathbb{R}^n$ for $i = 1, 2, \ldots, N$. In applying a 1-d PCA, we project the $x^{(i)}$ onto a 1-d linear subspace given by $\text{span}(v)$ for $v \in \mathbb{R}^n$. Write down the loss/error function that we use to find the "best" v and explain intuitively why this loss function makes sense. Then, go through the details of optimizing the loss function and determining the "best" v. Finally, explain how we can use the projection of the $x^{(i)}$ onto $\text{span}(v)$ to transform the dataset into a 1-d dataset.

(e) Now repeat (d), but in the case of a 2-d PCA.

4. Attached are two files, `senators_formatted.txt`, which provides the names of all Senators in the 109th Senate of the U.S. along with their state and party affiliations (D=democrate, R=republican) and, `votes_formatted.txt`, which provides all votes for each senator over all bills considered (542). The 2nd through 101st column of `votes_formatted.txt` give the votes for each senator in the same order of senators as given in `senators_formatted.txt`, an entry of 1, 0, -1 corresponds to a YES, ABSENT, and NO votes. The first column gives the name of the bill voted on.

(a) Let $x^{(i)}$ be the votes for senator i. Typically in PCA, the data is first centered. To do this set

$$\mu = \frac{1}{100} \sum_{i=1}^{100} x^{(i)} \tag{3}$$

and then replace each $x^{(i)}$ by $x^{(i)} - \mu$. (We are just subtracting off the mean.) With these centered $x^{(i)}$, define

$$\Theta = \sum_{i=1}^{100} x^{(i)} (x^{(i)})^T, \tag{4}$$

where we think of the $x^{(i)}$ as column vectors. Use **eigen** (or equivalent) to compute the first two dominant eigenvectors of Θ.

(b) Perform a 1-d PCA on the centered senator data. To do this, project each $x^{(i)}$ onto the dominant eigenvector and produce a 1-d plot of the senators. Color the senators according to party affiliation. What fraction of the total variance is captured by the PCA?

(c) Now repeat for a 2-d PCA and produce a 2-d plot.