Problem Set – Linear Algebra II

1. Show that the sum of the eigenvalues of a matrix is equal to the trace, and that the product of the eigenvalues is equal to the determinant. You can show this just for a 2x2 matrix, but this property holds for larger matrices.

2. Prove that if X is a matrix (of size nxn), where every column is an eigenvector (normalized to be have unit length) of a symmetric matrix Y, then the transpose of X is equal to the inverse of X ($X^T = X^{-1}$).

3. Practice PCA on this data set (from this website: http://setosa.io/ev/principal-component-analysis/).

	England	N Ireland	Scotland	Wales
Alcoholic drinks	375	135	458	475
Beverages	57	47	53	73
Carcase meat	245	267	242	227
Cereals	1472	1494	1462	1582
Cheese	105	66	103	103
Confectionery	54	41	62	64
Fats and oils	193	209	184	235
Fish	147	93	122	160
Fresh fruit	1102	674	957	1137
Fresh potatoes	720	1033	566	874
Fresh Veg	253	143	171	265
Other meat	685	586	750	803
Other Veg	488	355	418	570
Processed potatoes	198	187	220	203
Processed Veg	360	334	337	365
Soft drinks	1374	1506	1572	1256
Sugars	156	139	147	175

Note that this data forms a 17 x 4 matrix – meaning that we are plotting 4 points in a 17-dimensional space. Think about how to interpret the transpose of this matrix.

For those less familiar with MATLAB – I have written up most of the code for you in the MATLAB script posted with this problem set (titled 'HW2_prob3_notComplete.m) – but if you are comfortable with MATLAB, try coding this up for yourself! Also, for this problem, don't normalize the data set – just make sure that you subtract the mean off of each row.

Plot the first two rows of the re-expressed data set against each other (matrix Y in the lecture slides; this is the data projected onto the first two principal components). Which country is different from the others? Interpret the differences using the principal components (the eigenvectors of the covariance matrix of the data).

A CHALLENGE PROBLEM THAT IS ALSO FUN:

4. Practice the SVD by 'solving' the Netflix matrix completion problem. In this problem, Netflix wants to recommend movies to a customer. In order to do this, Netflix would like to consider other movies that the customer has liked, and what other similar customers have liked. But there is a problem: Netflix does not have the ratings for every movie for every customer, making this recommendation difficult. However, if we assume that this matrix is low-rank, we can complete this matrix.

   In order to complete a matrix, we want to fill in the missing values while retaining the known ones. In order to do this, we follow this algorithm:
   1. Let's call the original matrix X. Replace the missing elements of X with the mean of all other elements in X.
   2. Take a low-rank approximation of X using the method described in class. By low, I mean the first couple (5-10) of the few dimensions. Call this matrix X^1.
   3. Replace the known values (the ones that weren't missing in the original matrix) in matrix X^1.
   4. Repeat steps 2 and 3 (taking a low-rank approximation, and then replacing the known values) until the matrix stops changing, or for maybe 50 iterations. To compute how fast the matrix changes, you can compute the sum of squared differences for each values in the matrix (described in the accompanying MATLAB file).

   Download the 2 matlab files associated with this problem (HW2_prob4_notComplete.m and movie_person_matrix.m). I mostly coded the above algorithm up for you, but you have to fill in a couple remaining commands. Do this, and then, find the highest-ranking movie for the 100th customer.