Homework 3

Due March 15, 2011

1. a) Find the SVM discriminant vector and decision boundary for the following two-class discrimination problem. Solve graphically by hand – no need for MATLAB or numerical calculations.
   b) In the convex hull formulation of SVM, the solution for the discriminant vector is expressed in terms of λ coefficients (in this example there will be 6 of them, corresponding to the 6 data examples). What will be the 6 values in this case? (You can reason it out, you do not need to calculate anything).

Example	x_1	x_2	y
1	1	3	-1
2	1	2.5	-1
3	1.5	3	-1
4	2	3.5	1
5	2	2.5	1
6	2.5	4	1

2. Let A be represented by its SVD, i.e., $A = UDV^T$. Show that U is the eigenvector matrix of AA^T . What are the eigenvalues of AA^T ?

3. **Computer problem: PCA**.
   Load the matlab data file `hmwk3.mat.` The file contains a variable `x`, which is a $3\cdot 100$ data matrix (arranged EE style: 3 features, 100 examples). Use the following command to make a 3D scatter plot of the data:
   ```
   plot3(x(1,:),x(2,:),x(3,:),'.');grid on;
   ```
   If you use the rotate tool on the matlab figure window, you will see that these 3D data all lie in a tilted 2D plane; therefore, their intrinsic dimension is 2.
   a) Before you start programming, think about what the PCA basis vectors should be for this data, and write down why. Along which direction in this space does the data have the least variance?
   b) Compute the PCA basis vectors and eigenvalues using the eigenvector method.
   c) Do the same using the SVD method. (You should get the same answer). Don't forget to mean-correct the data before applying SVD. Note that MATLAB computes SVD two different ways. To get the definition we used in class, call it as follows:
   ```
   [u,d,v] = svd(a, 0);
   ```
   where `a` is the matrix you want to decompose.
   d) Using the variances of the original features, graph the % variance accounted for (%VAF) in descending order (largest %VAF first).
   e) Do the same for the PC scores (coordinates in the new basis). (Remember: the variances of the PC scores are the eigenvalues).

f) Compute the PC scores and verify that their variances are equal to the eigenvalues from the PCA.

g) I wrote a MATLAB program for you that makes plots of 2D projections of multidimensional data (the upper-triangular plotting format I described in class). The program is called as follows:

```
trellisplot(a, 3);
```

where a is the matrix of values you want to plot. Use this to plot the original data, and again to plot the PC scores. In words compare the results and describe what PCA has done.

h) One of your PCA eigenvalues should be zero. What is the significance of this?

i) Recall from linear algebra that the number of non-zero eigenvalues of a matrix is equal to the rank of the matrix (i.e., the number of linearly independent columns). Look at the covariance matrix of the original data. Is its rank consistent with our finding that one eigenvalues is zero? Explain.