Homework Solutions

Exercise 2, p. 65 (SVD) and Exercises 1-3 in the best basis notes.

2, SVD Here's the code:

```
X=[2 1 -1 3
  -1 0 1 -2
   7 2 -5 12
   -3 -2 0 -4
   4 1 -3 7];
y=[5;1;0;-2;6];

[U,S,V]=svd(X);
diag(S)

xhat=V(:,1:3)*diag(1./diag(S(1:3,1:3)))*U(:,1:3)'*y
yhat=X*xhat

% Is xhat in row space?  Is yhat in col space? (Check projections)
xhat-V(:,1:3)*V(:,1:3)'*xhat
yhat-U(:,1:3)*U(:,1:3)'*yhat
```

(a) Because there are 3 non-zero entries in S, the rank of X is 3.

- The dimension of the column space is 3
- The dimension of the row space is 3
- The dimension of the null space is 1
- The dimension of the null space of X^T is 2

(b) The values of $\hat{x}$, $\hat{y}$ were approximately:

$$\hat{x} \approx \begin{bmatrix} 1.4667 \\ 7.0667 \\ -5.7000 \\ -4.1333 \end{bmatrix} \qquad \hat{y} = \begin{bmatrix} 3.3 \\ 1.1 \\ 3.3 \\ -2.0 \\ 1.1 \end{bmatrix}$$

(c) One way to see if the vector are in their respective spaces is to see if the projection into those spaces changes the vector. That is, if a vector is already in a given space, then the projection doesn't change it. In the code we do that, and see that the differences are approximately zero (on the order of 10^{-14}).

1, Best Basis If we project the data to $\vec{e}_1$, that gives us the first coordinate of each vector. Therefore, subsequently finding the variance will mean just finding the variance in dimension 1. That is,

$$\text{Proj}_{\vec{e}_1}(\mathbf{x}^{(i)}) = x_1^{(i)}$$

Therefore,

$$S^2 = \frac{1}{p-1}\sum_{i=1}^{p}(x_1^{(i)})^2$$

Similarly, projecting to $\vec{e}_k$ gives us the variance in the k^{th} dimension.

2, Best Basis This is reproducing the argument given in the text.

3, Best Basis You should be able to simplify your solution to zero. Notice that this says that the low dimensional representation of the data is uncorrelated! For example, if when you find the best basis and project it to the best two dimensional subspace, then the new coordinates have a correlation of zero.