pset11

September 7, 2017

1 18.06 pset 11

Due **Friday**, May 12 at 11am.

1.1 Problem 1

A key fact leading to the SVD is that for any $m \times n$ real matrix A, the positive semidefinite (eigenvalues ≥ 0) matrices $A^T A$ and AA^T have the same nonzero eigenvalues $\sigma_k^2 > 0$, where the σ_k are called the **singular values** of A, for $k = 1, 2, \ldots, r$, for $r = \text{rank}(A)$. So, the SVD simultaneously diagonalizes $A^T A$ and AA^T.

In this problem, you will derive the "reduced" form of the SVD based only on what you know about eigenvectors.

(a) Suppose $\lambda = \sigma^2 > 0$ is one of the r nonzero eigenvalues (if any) of AA^T (they cannot be negative because AA^T is positive semidefinite for any A). That is, $AA^T u = \sigma^2 u$ for some eigenvector u, normalized to $u^T u = 1$. Find an eigenvector v of $A^T A$ with the *same* eigenvalue, normalized to $v^T v = 1$. (Hint: show that $A^T A(A^T u) = \cdots$. Check your $v^T v$ to make sure that it is 1!)

(b) Why aren't the eigenvectors for λ=0 eigenvalues related in the same way, i.e. why isn't there a 1-to-1 correspondence between the λ=0 eigenvectors of AA^T and $A^T A$, just as in the previous part? (Hint: long ago, in class, we showed $N(A^T) = N(AA^T)$ for any A... this was a key point in least-squares problems.)

(c) How do your eigenvectors u and v from (a) relate to the solution of problem 2 from pset 9?

(d) Since $N(A) = N(A^T A)$ (we derived this long ago in class, for least-squares and projection problems), explain why $Ax = A\hat{V}\hat{V}^T x$ for any x, where $\hat{V}$ is the $n \times r$ matrix whose columns are the *orthonormal eigenvectors* $v_1, \ldots, v_r$ of $A^T A$ with positive eigenvalues $\sigma_1^2, \ldots, \sigma_r^2 > 0$. (Recall that $\hat{V}\hat{V}^T$ is the projection operation onto $C(\hat{V})$. Hint: $C(\hat{V})$ is the orthogonal complement of the nullspaces of what matrices?)

(e) Take the $r = \text{rank}(A)$ nonzero eigenvectors σ_k^2 of AA^T (or $A^T A$) and the corresponding orthonormal eigenvectors u_k and v_k from part (a). Form the $m \times r$ matrix $\hat{U}$ whose columns are $u_1, \ldots, u_r$, along with the corresponding $\hat{V}$ matrix from above. Form the $r \times r$ *diagonal* matrix $\hat{\Sigma}$ whose diagonal entries are $\sigma_1, \ldots, \sigma_r$.

- Show that $A\hat{V} = \hat{U}\hat{\Sigma}$.
- Explain why it follows from (d) that $A = \hat{U}\hat{\Sigma}\hat{V}^T$. This is the reduced SVD: in the ordinary SVD you have *square* unitary matrices U and V and a non-square $m \times n$ diagonal matrix Σ.

1.2 Problem 2

Execute the following code cells in the Julia notebook, reading along, and answer the question at the end.

The following matrix represents the Iris flower dataset. Each row is a different flower (150 flowers), and the columns are the measurements (in cm) of the lengths of four different flower parts.

```
In [ ]: X = [5.1 3.5 1.4 0.2; 4.9 3.0 1.4 0.2; 4.7 3.2 1.3 0.2; 4.6 3.1 1.5 0.2; 5.0 3.6 1.4 0.2; 5.4 3
```

This data actually includes 3 different species of flower, and the goal is to figure out how to differentiate between the species based on the above data. If we number the species 0, 1, and 2 (Iris setosa, Iris virginica and Iris versicolor), then the following array contains the species of each row in X:

```
In [ ]: species = [0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,
```

Before we do any data analysis, we should subtract the mean of all the samples (all the flowers), which can be done efficiently in Julia via the "broadcasting" `.-` operations combined with the function `mean(X,1)` that returns a row vector of the means of each column:

```
In [ ]: A = X .- mean(X,1)
```

Now, each flower could be viewed as a point in a four-dimensional space (i.e. the 4 measurements) called feature space. We'd like to find out what are the important directions in this 4d space in order to explain the variations among the individual flowers, and hopefully to identify the fact that there are three different species.

If we just slice this 4d parameter space in a "random" direction, e.g. we plot the 4th coordinate (4th column of A) vs the 1st coordinate (1st column of A), we just see a blob of points. The different colors are the different species, and they are all mixed up together:

```
In [ ]: using PyPlot
        scatter(A[:,1], A[:,4], c=species, cmap="viridis")
        xlabel("sepal length (cm)")
        ylabel("petal width (cm)")
```

Now, let's perform some principal components analysis (PCA) to try to figure out the "right" way to look at this data.

First, compute the SVD of this matrix A, which is equivalent to diagonalizing the covariance matrix AA^T as discussed in class (and in the book):

```
In [ ]: U, σ, V = svd(A)
        σ
```

The singular values σ indicate how much of the variance is explained by each singular vector.

Given this information, you should be able to pick out **two directions** in 4d (two directions in feature space, i.e. two combinations of measurements) that **explain most of the variation between**.

Fix the code below to compute two 150-component vectors `x1` and `x2`, corresponding two these two combinations of measurements for each flower, and plot `x2` vs `x1`.

You should see that the flowers separate into three groups corresponding to the three different species: PCA identifies the key measurements (or rather, the key combination of measurements) that explain the variations.

```
In [ ]: x1 = ???
        x2 = ???
        scatter(x1, x2, c=species, cmap="viridis")
```

Print out this plot (or print to PDF or make a screenshot and attach it electronically on Stellar) and your ??? code with your solutions.