Homework #6: Machine Learning

Kyler Little

April 20, 2018

Problem #1

Let $A = U\Sigma V^T$ be the SVD of A, where $A \in R^{m\times n}$, $U \in R^{m\times m}$ and $V \in R^{n\times n}$ are orthogonal matrices, $\Sigma = \text{diag}(\sigma_1, \cdots, \sigma_r, 0, \cdots, 0)$, and $r = \text{rank}(A)$. Show that:
(a) The first r columns of U are eigenvectors of AA^T corresponding to nonzero eigenvalues.

$$\begin{aligned} AA^T &= U\Sigma V^T(U\Sigma V^T)^T \\ &= U\Sigma V^T(V^T)^T\Sigma^T U^T \\ &= U\Sigma\Sigma^T U^T \\ &= U\begin{bmatrix} \Sigma^{*2} & 0 \\ 0 & 0 \end{bmatrix} U^T \end{aligned}$$

Note that $\Sigma^{*2} = \text{diag}(\sigma_1^2, \cdots, \sigma_r^2)$, and the transition from the second step to the third occurs because V is an orthonormal matrix. Because AA^T is symmetric (true for any matrix A), we see that the final equation is actually the eigen decomposition of AA^T. Therefore, the columns of U are the eigenvectors of AA^T. Clearly, only the first r columns of U can possibly correspond to nonzero eigenvalues since the remaining diagonal entries of Σ^2 are zeros. But to prove this, we must show $(AA^T)_r$ is positive definite. This is relatively easy, using the knowledge of past homeworks. Given any $x \in R^r$ s.t. $x \neq 0$, we have that $x^T(AA^T)_r x = (A^T x)^T A^T x = ||A^T x||^2 > 0$ for nonzero x. Therefore, we have that the eigenvalues of $(AA^T)_r$ are all positive and nonzero, so we are done.

(b) The first r columns of V are eigenvectors of A^TA corresponding to nonzero eigenvalues.

$$\begin{aligned} A^TA &= (U\Sigma V^T)^T U\Sigma V^T \\ &= (V^T)^T\Sigma^T U^T U\Sigma V^T \\ &= V\Sigma^T\Sigma V^T \\ &= V\Sigma^{*2}V^T \end{aligned}$$

As we can see, a nearly identical argument can be made for this case. Clearly, the last equation is an eigen decomposition for A^TA. It only needs to be proved that the diagonal entries of Σ^{*2} are strictly positive. Similar to the part (a), we have that given any $x \in R^r$ s.t. $x \neq 0$, $x^T(A^TA)_r x = (Ax)^T Ax = ||Ax||^2 > 0$ for nonzero x. Therefore, we have that the eigenvalues of $(A^TA)_r$ are all positive and nonzero, so we are done once more.

Problem #2

Given a symmetric matrix $A \in R^{3\times3}$, suppose its eigen-decomposition can be written as:

$$A = \begin{pmatrix} u_{11} & u_{12} & u_{13} \\ u_{21} & u_{22} & u_{23} \\ u_{31} & u_{32} & u_{33} \end{pmatrix} \begin{pmatrix} 3 & 0 & 0 \\ 0 & -2 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} u_{11} & u_{21} & u_{31} \\ u_{12} & u_{22} & u_{32} \\ u_{13} & u_{23} & u_{33} \end{pmatrix}$$

What is the singular value decomposition of this matrix?
Firstly, note that the decomposition is already quite close to the SVD. The only issue is that -2 should be positive, but otherwise, the magnitudes of the eigenvalues are nonincreasing and all greater than zero. Thus, a nice trick we can apply is to note that $AA^T = U\Lambda U^T U\Lambda U^T = U\Lambda^2 U^T$ (recall $U^T U = I$ and Λ is a diagonal matrix), which is a singular value decomposition of AA^T. Because A is symmetric, we have that $A^T A = AA^T$. Knowing this, we can see that the only thing that changes between the eigen-decomposition for A and the SVD for AA^T is actually the central matrix (Λ versus Λ^2). Thus, to obtain the SVD for A, we must ensure that the sign of the singular values is kept positive. We can do this by putting the sign of each eigenvalue on to the corresponding column vector in U and then making all eigenvalues positive by taking their absolute values. In other words, let $A = \sum_{i=1}^{3} \text{sign}(\lambda_i) u_i |\lambda_i| u^T$. Therefore, the SVD for A is:

$$A = \begin{pmatrix} u_{11} & -u_{12} & u_{13} \\ u_{21} & -u_{22} & u_{23} \\ u_{31} & -u_{32} & u_{33} \end{pmatrix} \begin{pmatrix} 3 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} u_{11} & u_{21} & u_{31} \\ u_{12} & u_{22} & u_{32} \\ u_{13} & u_{23} & u_{33} \end{pmatrix}$$

Problem #3

Given a data matrix $X = \begin{bmatrix} x_1, & x_2, & \cdots & x_n \end{bmatrix} \in R^{p\times n}$ consisting of n data points, and each data point is p-dimensional,

- Outline the procedure for computing the PCA of X.
- State what is the minimum reconstruction error property of PCA.
- Prove the minimum reconstruction error property of PCA by using the best low-rank approximation property of SVD.

Problem #4

Use the similarity matrix in Table 1 to perform single (MIN) and complete (MAX) link hierarchical clustering. Show your results by drawing a dendrogram. The dendrogram should clearly show the order in which the points are merged.

Table 1: Similarity Matrix

	p1	**p2**	**p3**	**p4**	**p5**
p1	1.00	0.10	0.41	0.55	0.35
p2	0.10	1.00	0.64	0.47	0.98
p3	0.41	0.64	1.00	0.44	0.85
p4	0.55	0.47	0.44	1.00	0.76
p5	0.35	0.98	0.85	0.76	1.00

Problem #5

Summarize results of PCA implementation and attach images.