CS698C: Sketching and Sampling for Big Data Analysis

Exercises: SVD 4

Note: All questions pertain to real matrices.

1. Find all the entries in the SVD of a rank 1 matrix.

$$\begin{bmatrix} 1 & 4 \\ 2 & 8 \end{bmatrix} = \begin{bmatrix} u_1 & u_2 \end{bmatrix} \begin{bmatrix} \sigma_1 & \\ & 0 \end{bmatrix} \begin{bmatrix} v_1 & v_2 \end{bmatrix}^T$$

(*Hint:* Find the eigenvalues and eigenvectors of $A^T A$ or AA^T.)

2. Find the SVD of A whose eigenvector decomposition is given below.

$$A = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{bmatrix} \begin{bmatrix} 1 & \\ & -1 \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{bmatrix}$$

3. Let $A = U\Sigma V^T$ be an SVD decomposition. Are U and V unique? Σ? Give reasons/arguments.

4. (a) Give examples of matrices for which eigenvalues equal singular values.
   (b) Give examples of matrices for which the singular values are the absolute values of the corresponding eigen values.
   (c) Give examples of matrices for which the eigenvalues are the squares of the singular values.

   Now show the following.

   (a) Singular values and eigenvalues coincide for positive semi-definite matrices.
   (b) For symmetric matrices, the singular values are the absolute values of eigen values.
   (c) For all general matrices, the singular values are the positive square roots of the eigenvalues of $A^T A$ or AA^T.

5. Prove or disprove. Let A be an n by n symmetric matrix whose eigenvector decomposition is $A = U\Lambda U^T$, where, $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$. The singular values of A are $|\lambda_1|, |\lambda_2|, \ldots, |\lambda_n|$ and the SVD of A is

$$A = \begin{bmatrix} u_1 & u_2 & \cdots & u_n \end{bmatrix} \begin{bmatrix} |\lambda_1| & & \\ & \ddots & \\ & & |\lambda_n| \end{bmatrix} \begin{bmatrix} -v_1^T- \\ -v_2^T- \\ \vdots \\ -v_n^T- \end{bmatrix}$$

where, $U = \begin{bmatrix} u_1 & u_2 & \cdots & u_n \end{bmatrix}$ and

$$v_i = \begin{cases} u_i & \text{if } \lambda_i \geq 0 \\ -u_i & \text{otherwise.} \end{cases}$$

6. Suppose A has columns $\begin{bmatrix} w_1 & w_2 & \cdots & w_n \end{bmatrix}$ where, the w_i's are orthonormal (i.e., $w_i^T w_j = 0$ if $i \neq j$ and $1 \leq i, j \leq n$) and have lengths $\sigma_1, \sigma_2, \ldots, \sigma_n$, that is, $\|w_i\| = \sigma_i$, for $i = 1, 2, \ldots, n$. What is $A^T A$, and what is the SVD $A = U\Sigma V^T$.

7. Suppose an m by n matrix A has SVD $A = U\Sigma V^T$. Find unit vectors w and x such that $\|Ax\|_2 = \max_{\|v\|_2=1}\|Av\|_2$ and $\|Aw\|_2 = \min_{\|v\|_2=1}\|Av\|_2$.

8. Show the following properties of SVD decomposition.

   (a) If A is invertible, then $A^{-1} = A^{-}$ (the inverse of A is the pseudo-inverse).

   (b) If A is square, then $|\det A|$ = product of the singular values of A.

   (c) Suppose A is a positive definite matrix and $A = U\Lambda U^T$ be its eigen decomposition, then it is also its SVD. This holds for positive semi-definite matrix also. (Positive definite and semi-definite matrices are always symmetric).

   (d) Suppose A is a symmetric matrix with eigen decomposition $A = U\Lambda U^T$. Its singular values are the absolute values of its eigenvalues.

   (e) If U is an m by m orthogonal matrix and A is an m by n matrix, then A and UA have the same singular values.

9. Suppose A is symmetric. Then the eigenvalues of $A + I$ are $\lambda_i + 1$, where, $\lambda_1, \ldots, \lambda_n$ are the eigenvalues of A. What can be said about the SVD of $A + I$?

10. Use the rank 1 decomposition of the SVD of $A = U\Sigma V^T = \sigma_1 u_1 v_1^T + \sigma_2 u_2 v_2^T + \cdots + \sigma_r u_r v_r^T$.

    (a) Let c be an arbitrary unit vector in $\mathbb{R}^m$. Suppose $c = Vd$ or $d = V^T c$. Show that

    $$Ac = \sigma_1 d_1 u_1 + \sigma_2 d_2 u_2 + \cdots + \sigma_r d_r u_r$$

    and

    $$\|Ac\|_2^2 = (\sigma_1 d_1)^2 + \cdots + (\sigma_r d_r)^2 = (\sigma_1(v_1^T c))^2 + \cdots + (\sigma_r(v_r^T c))^2 \ .$$

Questions on pseudo-inverse. Let $A = U\Sigma V^T$ be the SVD of A. The pseudo inverse of A is defined as $V\Sigma^{-}U^T$, where, Σ^{-} is an n by m matrix with all off-diagonal entries zeros and

$$\Sigma_{ii}^{-} = \begin{cases} \Sigma_{ii}^{-1} = \sigma_i^{-1} & \text{if } \sigma_i > 0 \\ 0 & \text{Otherwise.} \end{cases}$$

In terms of the thin SVD, if A has rank r then $A = U_r \Sigma_r V_r^T$ is called the thin SVD, where, it is assumed that $U_r = \begin{bmatrix} u_1 & u_2 & \cdots & u_r \end{bmatrix}$ and $V_r = \begin{bmatrix} v_1 & \cdots & v_r \end{bmatrix}$ are the first r columns of U and V respectively. These columns directly correspond to the singular values $\sigma_1, \ldots, \sigma_r$ which are all positive. All singular values $\sigma_{r+1}, \ldots$ are zeros. Then,

$$A = U\Sigma V^T = U_r \Sigma_r V_r^T$$

The pseudo-inverse can be written as $A^{-} = V_r \Sigma_r^{-1} U_r^T$.

1. Show that $AA^{-} = U_r U_r^T$ is the projection matrix on the column space of A. (i.e., for any $x \in \mathbb{R}^m$, $AA^{-}x$ is the projection of x onto the column space of A.

2. Show that $A^{-}A = V_r V_r^T$ is the projection matrix on the row space of A.

3. If $r = m$ and $m \leq n$, show that A^{-} is a right inverse.

4. If $r = n$ and $m \geq n$, show that A^{-} is a left inverse.

5. If $r = m = n$, then $A^- = A^{-1}$.

6. Show that $(A^T)^- = (A^-)^T$.

7. Going back to full SVD $A = U\Sigma V^T$, can you find the projection matrix on to the nullspace of A?

1 Linear Regression

We are given an $m \times n$ matrix A and an m dimensional vector b; generally with $m \gg n$. The linear regression problem is $\min_{x \in \mathbb{R}^n} \|Ax - b\|_2^2$. The solution is obtained by projecting b onto the column space of A. Let x^* be a vector that minimizes $\|b - Ax\|_2^2$. Then, $b - Ax^*$ is orthogonal to the column space of A, that is, $A^T(b - Ax^*) = 0$, or

$$A^T A x^* = A^T b$$

This is called normal equations. Writing $A = U_r \Sigma_r V_r^T$ (thin SVD), we have,

$$V_r \Sigma_r U_r^T U_r \Sigma_r V^T x^* = V_r \Sigma_r U_r^T b$$

Since, $U_r^T U_r = I_r$, the equation is equivalent to

$$V_r \Sigma_r^2 V_r^T x^* = V_r \Sigma_r U_r^T b$$

Multiplying from the left by V_r^T, we get $V_r^T V_r = I_r$ on both sides, this gives

$$\Sigma_r^2 V_r^T x^* = \Sigma_r U_r^T b$$

Multiplying on the left by Σ_r^{-2} on both sides, we get

$$V_r^T x^* = \Sigma_r^{-1} U_r^T b \ .$$

Let W_{n-r} denote an orthogonal basis for the nullspace of A, that is,

$$V = \begin{bmatrix} V_r & W_{n-r} \end{bmatrix}$$

Any vector x can be uniquely written as a sum of two orthogonal vectors, $x = x_r + x_n$, where, $x_r = V_r y$ and lies in the rowspace of A and $x_n = W_{n-r} z$ that lies in the nullspace of A, for some $y \in \mathbb{R}^r$ and $z \in \mathbb{R}^{n-r}$. In this notation, we So we write $x^* = x_r^* + x_n^*$. Let

$$x^* = V w^* = \begin{bmatrix} V_r & W_{n-r} \end{bmatrix} \begin{bmatrix} y^* \\ z^* \end{bmatrix}, \qquad y^* \in \mathbb{R}^r, z^* \in \mathbb{R}^{n-r}$$

Then,

$$V_r^T x^* = V_r^T (V_r y^* + W_{n-r} z^*) = y^* + V_r^T W_{n-r} z^* = y^*$$

since, V_r and W_{n-r} columns are orthogonal to one another (row space is orthogonal to nullspace). Going back to the equation we need to solve, $V_r^T x^* = \Sigma_r^{-1} U_r^T b$, and writing $x^* = V y^*$ as above, we have,

$$y^* = \Sigma_r^{-1} U_r^T b \ .$$

The space of all solutions is then

$$x^* = V_r y^* + W_{n-r} z, \qquad z \in \mathbb{R}^{n-r}$$

Further, since, the column spaces of V_r and W_{n-r} correspond to the rowspaces and nullspaces, respectively, of A, they are orthogonal and by Pythagoras theorem,

$$\|x^*\|_2^2 = \|V_r y_r^* + W_{n-r} z\|_2^2 = \|V y^*\|_2^2 + \|W_{n-r} z\|_2^2 = \|y_r^*\|_2^2 + \|z\|_2^2 \ .$$

Thus among all the possible solutions that minimizes $\|Ax - b\|_2^2$, the value of x^* that has the smallest ℓ_2 norm is when $z = 0$. This gives the least norm solution among all vectors that minimizes $\|Ax - b\|_2^2$ to be

$$x^* = V_r y_r^* = V_r \Sigma_r^{-1} U_r^T b = A^- b$$

as per the definition of $A^- = V_r \Sigma_r^{-1} U_r^T$. This proves the property that $x^* = A^- b$ minimizes $\|Ax - b\|_2^2$ and among all such vectors, has the smallest norm. In general, the space of solutions to

$$\min_{x \in \mathbb{R}^n} \|Ax - b\|_2^2$$

is given by

$$A^- b + W_{n-r} z, \qquad z \in \mathbb{R}^{n-r} \ .$$

When do we have a unique solution? Iff A has nullspace $\{0\}$ or that $r = n$. This means that $A^T A$ is an $n \times n$ matrix and of rank n and hence, invertible. Going back to normal equations, we have,

$$A^T A x^* = A^T b \ .$$

In this case (and only this case), $A^T A$ is invertible, and

$$x^* = (A^T A)^{-1} A^T b \ .$$

Is this solution the same as $x^* = A^- b$? Let us check this. Writing as SVD

$$(A^T A)^{-1} A^T b = ((V_r \Sigma_r U_r^T) U_r \Sigma_r V_r^T)^{-1} V_r \Sigma_r U_r^T b = (V_r \Sigma_r^2 V_r^T)^{-1} V_r \Sigma_r U_r^T b \ .$$

Since $r = n$, $V_r = V$ is an orthogonal matrix and its inverse is V^T. This simplifies our calculation. Since, $(AB)^{-1} = B^{-1} A^{-1}$, when inverses do exist, we have,

$$(V_r^T)^{-1} (\Sigma_r^2)^{-1} (V_r)^{-1} = V_r \Sigma_r^{-2} V_r^T$$

Thus

$$(A^T A)^{-1} A^T b = V_r \Sigma_r^{-2} V_r^T V_r \Sigma_r U_r^T b = V_r \Sigma_r^{-1} U_r^T b = A^- b \ .$$

Another way to view this is to recognize that the closest vector to b on the columnspace of A is $U_r U_r^T b$. Thus, any solution to $Ax^* = U_r U_r^T b$ is the one that minimizes $\|Ax - b\|_2^2$.

$$\|Ax - b\|_2^2 = \|Ax - U_r U_r^T b + (U_r U_r^T - I) b\|_2^2 = \|Ax - U_r U_r^T b\|_2^2 + \|(I - U_r U_r^T) b\|_2^2$$

where, the last step follows from the fact that $Ax - U_r U_r^T b = Ax - Ax^*$ is in the column space of A and $I - U_r U_r^T$ is the projection matrix in the space orthogonal to the colspace of A. Hence,

$$\min_x \|Ax - b\|_2^2 = \min_x \|A(x - x^*)\|_2^2 + \|(I - U_r U_r^T) b\|_2^2 \ .$$

The second term in the *RHS* is a constant and independent of x. Thus, the minimum occurs when $x = x^* + z$, where, z is in the null space of A. As discussed earlier, among all such solutions, the solution with the minimum norm occurs when $z = 0$. Thus,

$$\min_x \|Ax - b\|_2^2 = \|(I - U_r U_r^T)b\|_2^2$$

where, $Ax^* = U_r U_r^T b$. Calculating as before,

$$Ax^* = U_r \Sigma_r V_r^T x^* = U_r U_r^T b$$

and assuming x^* has no nullspace component, this gives $x^* = V y^*$, and so,

$$y^* = \Sigma^{-1} U_r^T b, \text{ or, } x^* = V_r \Sigma_r^{-1} U_r^T b = A^- b \ .$$