Recitation 14

Thursday October 27, 2022

Directions Spend at most 10 minutes on the TLDR and then break up into groups to solve problems. Even if you have read the midterm solutions, test your understanding by trying to explain the solution to someone else.

1 TLDR

1.1 Word Embeddings

Yet another application of the SVD is computing word embeddings. Given a collection of m documents, we form the word-by-word co-occurrence matrix A and let $A = U\Sigma V^T$ be its SVD. Let $U_{1:k}$ be the first k columns of U. Then the rows of $U_{1:k}$ represents each word as a k-dimensional vector. Now we can solve analogies like

Man:Woman :: King:?

by forming the vector difference $v_{king} + (v_{woman} - v_{man})$ and searching for the word whose vector is closest. As we saw in lecture, word embeddings can reveal hidden biases in your data.

2 Midterm Revisit and Exercises

Q4 Suppose T is an invertible matrix. Let $B = AT$. Then $N(B) = N(A)$. True or False?

Q6 Consider two $n \times n$ projection matrices

$$P = I - v_1 v_1^\top \quad \text{and} \quad Q = I - v_2 v_2^\top$$

where v_1 and v_2 have unit norm and are orthogonal to each other. Let $A = PQ$

(a) What is the dimension of $N(A)$? Find an orthonormal basis for $N(A)$.
(b) What is the rank of A?
(c) Is A a projection matrix?

Q8 We want to measure the mass m of a bunny. Since in practice measurements always have errors, we weigh the bunny 4 times, obtaining slightly different results each time. This procedure gives rise to the system of equations

$$m = w_1, \quad m = w_2, \quad m = w_3, \quad m = w_4,$$

where w_i is the result of the i-th measurement.

(a) Write this as a linear system in matrix form $Ax = b$. What are the sizes of your matrices A and b?

Hint: How many variables are there? How many linear equations? Does this tell you what the dimensions of your linear system should be?

(b) Is this system solvable for every right-hand side? When the system is solvable, is the solution necessarily unique?

(c) Give bases for the subspaces $C(A)$ and $C(A)^{\perp}$. What are their dimensions?

(d) Compute the projection of b onto $C(A)$, and (optionally) interpret the results.

1. If you understand it deeply, the SVD gives a unified way to understand a lot of linear algebraic statements. Let's revisit some assertions we've made in class (particularly in the context of least squares) and give a direct argument via the SVD:

(a) If A has full column rank then $A^T A$ is invertible.

(b) If A has full row rank then it has right-inverse.

(c) Let A be an $n \times m$ matrix and $\lambda > 0$. Then, $A^{\top}A + \lambda I_n$ is invertible. Is this true when λ can be negative? Show a counter example. *Hint: What is an SVD of $A^{\top}A$? What would happen if we try to add λI_n?*

3 Solutions

Q4 **False.** Consider the following matrices

$$A = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \text{ and } T = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix} \text{ which implies } B = \begin{bmatrix} 0 & 1 & 1 \\ 0 & 1 & 1 \\ 1 & 0 & 0 \end{bmatrix}$$

It is easy to see that

$$N(A) = \operatorname{span}\left(\begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix} \right) \text{ but } N(B) = \operatorname{span}\left(\begin{bmatrix} 0 \\ 1 \\ -1 \end{bmatrix} \right)$$

Q6 (a) The dimension of $N(A)$ is two and $\{v_1, v_2\}$ forms an orthonormal basis. It is easy to see that $Av_i = 0$. Moreover for any vector v we can form an orthogonal decomposition $v = u + w$ where u is in the span of v_1 and v_2 and w is in the orthogonal complement. Then $Av = u$ and so if v is not in the span of v_1 and v_2 it is not in the nullspace.

(b) By the rank-nullity theorem, we have that

$$\operatorname{rank}(A) + \dim N(A) = n.$$

By the previous item, $\dim N(A) = 2$, and thus the rank of A is $n - 2$.

(c) Yes. We can write out

$$A = (I - v_1 v_1^\top)(I - v_2 v_2^\top) = I - v_1 v_1^\top - v_2 v_2^\top,$$

so this is the projection onto the orthogonal complement of $\operatorname{span}\{v_1, v_2\}$. We can also verify that $A^2 = A$.

Q8 (a) In this problem the measurements w_i are given, and we are trying to find the value of the mass m (i.e., m is the variable to solve for). The given equations can then be written as the linear system $Am = b$, where

$$A = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}, \qquad b = \begin{bmatrix} w_1 \\ w_2 \\ w_3 \\ w_4 \end{bmatrix}.$$

Both matrices A and b have size 4×1.

(b) The system is only solvable for some particular right-hand sides, namely when all the measurements w_i are equal. If that's the case, the solution is unique: $m = w_1 = \cdots = w_4$.

(c) We have

$$C(A) = \operatorname{span}\left\{ \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} \right\}, \qquad C(A)^\perp = N(A^T) = \operatorname{span}\left\{ \begin{bmatrix} 1 \\ -1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ -1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \\ -1 \end{bmatrix} \right\},$$

which have dimensions 1 and 3, respectively.

(d) The orthogonal projection of $[w_1, w_2, w_3, w_4]$ onto $C(A)$ is $[\bar{w}, \bar{w}, \bar{w}, \bar{w}]$, where $\bar{w} := (w_1 + w_2 + w_3 + w_4)/4$ is the *average* of the measurements.

1. (a) Let $A = U\Sigma V^T$ be the SVD. Suppose A is $n \times m$. If A has full column rank then it has m nonzero singular values, because the dimension of $N(A)$ is m minus the rank. Now we compute

$$A^T A = V\Sigma^T \Sigma V^T$$

It is easy to see that $\Sigma^T\Sigma$ is an $m \times m$ matrix whose singular values are σ_i^2. Since all its m singular values are nonzero, we know that $A^T A$ is invertible.

(b) If A is $n \times m$ and has full row rank, its SVD and pseudoinverse must have the form

$$A = U \begin{bmatrix} \Sigma & 0 \end{bmatrix} V^T, \qquad A^+ = V \begin{bmatrix} \Sigma^{-1} \\ 0 \end{bmatrix} U^T,$$

where Σ is $n \times n$, diagonal, and nonsingular. Then, we can easily verify that $AA^+ = I_n$, and thus A^+ is a right-inverse of A.

(For comparison, here's an "old style" proof: since A is full row rank, the matrix AA^T is invertible (why?). Then, an explicit right inverse is $R = A^T(AA^T)^{-1}$, since $AR = I_n$. The SVD approach seems simpler – no theorems or formulas to remember!)

(c) Consider the SVD $A = U\Sigma V^\top$. As derived earlier, we have $A^\top A = V\Sigma^\top\Sigma V^\top$. Notice that $I_n = V(\lambda I_n)V^\top$; therefore

$$A^\top A + \lambda I_n = V(\Sigma^\top\Sigma + \lambda I_n)V^\top$$

The above expression is the SVD and thus the singular values of $A^\top A + \lambda I_n$ are $\sigma_i^2 + \lambda$. Since $\lambda > 0$ all these singular values are positive which implies that $A^\top A + \lambda I_n$ has rank m, or, equivalently, invertible.