Recitation 13

Tuesday October 25, 2022

Directions Spend at most 15 minutes on the TLDR and then break up into groups to solve problems.

1 TLDR

1.1 Operator Norm

The *operator norm*, denoted $\|A\|$, is

$$\|A\| = \max_{x \text{ s.t. } \|x\| \leq 1} \|Ax\|.$$

In words, it is the maximum amount the matrix A can scale a unit vector x. It is also equal to the largest singular value of A.

1.2 Low Rank Approximation

We can find the best low rank approximation to B through its truncated singular value decomposition. Suppose $B = \sum_{i=1}^{r} \sigma_i u_i v_i^\top$ and we want to find, among all rank at most k matrices, the matrix C that minimizes $\|B - C\|$. The minimum is achieved by setting

$$C = \sum_{i=1}^{k} \sigma_i u_i v_i^\top$$

in which case we have $\|B - C\| = \sigma_{k+1}$.

1.3 Principal Component Analysis (PCA)

The goal of PCA is to find a k dimensional subspace that maximizes the projected variance. Suppose we are given p data points $a_1, \cdots, a_p$ and assume they are centered so that $\sum_{i=1}^{p} a_i = 0$. Then we want to find a matrix C with k orthonormal columns so as to maximize

$$\frac{1}{p} \sum_{i=1}^{p} \|C^T a_i\|^2$$

If we collect the data points into a matrix $A = [a_1, \cdots, a_p]$ then the maximum is obtained by setting $c_1, \cdots, c_k = u_1, \cdots, u_k$ where $u_1, \cdots, u_k$ are the first k left singular vectors from the SVD of A. This same choice also minimizes the reconstruction error

$$\frac{1}{p} \sum_{i=1}^{p} \|a_i - CC^T a_i\|^2$$

2 Exercises

1. Show that for any $n \times m$ and $m \times p$ matrices A and B we have

$$\|AB\| \leq \|A\|\|B\|$$

   Can you give an example where the inequality is strict?

2. Let A be a matrix with $\|A\| = 10$. For each of the following matrices, either determine the operator norm or argue that it cannot be determined from $\|A\|$ alone.

   (a) $AA^\top$

   (b) A^+

3. Let A be a 4×5 matrix with singular values $\sigma_1 = 5, \sigma_2 = 3, \sigma_3 = 2$ and the rest are zero.

   (a) Find the dimension of the nullspace of A.

   (b) Let B be the best rank two approximation to A. What is the operator norm of the *noise* $A - B$? What is the smallest non-zero singular value of B?

4. Suppose your data $a_1, \ldots, a_p$ is centered and spans a r dimensional space. If you perform PCA to find the best k dimensional subspace that minimizes reconstruction error. If $k < r$ can the reconstruction error be zero? Why or why not?

3 Solutions

1. By definition, the operator norm of AB is $\max_{x \text{ s.t. } \|x\|\leq 1} \|ABx\|$. Suppose that x^* achieves the maximum. Now let $y^* = Bx^*$. Then we have

$$\|ABx^*\| = \|Ay^*\| \leq \|A\|\|y^*\|$$

where we have again invoked the definition of the operator norm. Similarly

$$\|y^*\| = \|Bx^*\| \leq \|B\|\|x^*\|$$

and putting it all together, combined with the fact that $\|x^*\| \leq 1$, we have

$$\|ABx^*\| \leq \|A\|\|B\|$$

An example where the inequality is strict is

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

in which case $\|A\| = \|B\| = 1$ but $\|AB\| = 0$.

2. $\|A\| = 10$ means A's largest singular value is 10. The relationship between singular values can be found in recitation 12 note.

   (a) The singular values of $AA^\top$ are the square of those of A. This means $\|AA^\top\|$ is the largest singular value of $AA^\top$ which is $10^2 = 100$.

   (b) The singular values of A^+ are the reciprocal of those of A. This means $\|A^+\|$ is the largest singular value of A^+ which is the reciprocal of the *smallest* singular values of A. Therefore, $\|A^+\|$ cannot be determined given $\|A\|$ alone.

3. (a) The rank of A is the number of non-zero singular values which is 3. Rank-Nullity Theorem then tells us that rank $N(A) = 5 - 3 = 2$.

   (b) The best rank-2 approximation is given by $B = \sigma_1 u_1 v_1^\top + \sigma_2 u_2 v_2^\top$. The noise is $A - B = \sigma_3 u_3 v_3^\top$ which has operator norm $\sigma_3 = 2$. Finally, B has two non-zero singular values $\sigma_1 = 5$ and $\sigma_2 = 3$.

4. The reconstruction error must be nonzero. For the sake of contradiction, suppose not. Then each individual reconstruction error

$$\|a_i - CC^T a_i\|^2$$

must be zero. But this can only happen if each a_i is in the subspace spanned by the columns of C. Since the a_i's span an r dimensional subspace and $r > k$, there is no k dimensional subspace that contains them all.