Eckart-Young The Closest Rank k Matrix to A

So this is a pretty key lecture this lecture is about principle component analysis PCA which is a major tool in understand a matrix of data.

So what is PCA about. First of all, let's review the last lecture SVD.

$$A = U\Sigma V^T = \sigma_1 u_1 v_1^T + \cdots + \sigma_r u_r v_r^T$$

When we have a big matrix, we want to get the important information out of it not all the information.

Actually, the important facts about the matrix are in its largest k singular values.

If matrix B has rank k, then

$$||A - B|| \geq ||A - A_k||$$

$A_k = U_K \Sigma_k V_k^T$. It is often called **The Eckert Young Theorem**. Let's first talk about norm, before we prove the inequality.

A norm is a value to measure of vector or a matrix.

$$||v||_1 = |v_1| + \cdots + |v_n|$$

$$||v||_2 = \sqrt{v^T v}$$

$$||v||_\infty = \max_{1 \leq i \leq n} |v_i|$$

and we call them ℓ-1 norm, ℓ-2 norm and ℓ-infinity norm. The ℓ-1 norm is very important. When you use it to minimizing some function, the minimizing vector turns out to be **sparse**. And the minimizing process is called **basis pursuit**.

There are three property of norm which we have already learned.

There are three main form of matrix norm.

$$||A||_2 = \sigma_1$$

$$||A||_F = \sum_i \sum_j \sqrt{{a_{ij}}^2}$$

$$||A||_N = \sigma_1 + \sigma_2 + \cdots \sigma_{rank}$$

σ_1 is the greatest singular value of A. ''F' is for Frobenius. ''N'' for Nuclear. (Netflix use the Nuclear to win zillion dollars).

Let's see a example about the Eckart-Young Theorem.

Suppose the $k = 2$

$$A = \begin{bmatrix} 4 & 0 & 0 & 0 \\ 0 & 3 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

What is A_2 ?

$$A_2 = \begin{bmatrix} 4 & 0 & 0 & 0 \\ 0 & 3 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

Actually, if we multiply a matrix by an orthogonal matrix U, then they have the same norm. Same property to a vector.

$$QA = (QU)\Sigma V^T$$