Applications of SVD

§ Low Rank Approximation

First, Let us define the Rank of the Matrix. There are many ways one can define the rank of a matrix.
Rank of a Matrix M, rank (M), is the number of linearly independent columns in M. It is equal to the number of linearly independent Rows in M. In addition, it is equal to the number of non-zero eigen values (or singular values) of M.

- In the low Rank approximation, the goal is to approximate a Given Matrix M, with a low Rank matrix $\tilde{M}_k$ such that $\| M - \tilde{M}_k \|$ is approximately 0.
We need to specify the norm.
We have a high dimensional data represented as a Matrix and we want to approximate it with a much lower dimensional object. Note that a rank k approximation of M can be expressed using only k-singular values & the corresponding k-singular vectors. So, in principal, a rank k approximation only needs $O(kn)$ bits of memory. (store m×n Matrix with only $O(kn)$ bits.

Definition (Norm of the Matrix). The operator norm of a matrix M is defined as follows:

$$\|M\|_2 = \max_x \frac{\|Mx\|_2}{\|x\|_2}$$

Claim: $\|M\|_2 = \sigma_{max}(M)$

pf.

$$\max_x \frac{\|Mx\|_2}{\|x\|_2} = \max_x \left[\frac{\|Mx\|_2^2}{\|x\|_2^2}\right]^{1/2} = \max \sqrt{\frac{\langle Mx, Mx\rangle}{\langle x, x\rangle}}$$

$$= \max \sqrt{\frac{x^T M^T M x}{x^T x}} = \sqrt{\lambda_{max}(M^T M)}$$

$$= \sigma_{max}(M)$$

singular value.

[λ is an eigenvalue of $M^T M$, then $\lambda = \sigma^2$, σ is a singu. val]

Theorem: Let $M \in \mathbb{R}^{m\times n}$ with singular values $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_n$. For any integer $k \geq 1$,

$$\min_{\tilde{M}_k} \|M - \tilde{M}_k\|_2 = \sigma_{k+1}.$$

pf. We need to prove two statements:

(i) $\exists$ rank-k matrix $\tilde{M}_k$ such that $\|M - \tilde{M}_k\|_2 = \sigma_{k+1}$

(ii) For any rank-k matrix $\tilde{M}_k$: $\|M - \tilde{M}_k\|_2 \geq \sigma_{k+1}$.

(i) Let $\tilde{M}_k = \sum_{i=1}^{k} \sigma_i u_i v_i^T$

By SVD $M \in M_{m,n}(\mathbb{R})$ ($m \leq n$)

$M = \sum_{i=1}^{k} \sigma_i u_i v_i^T$

$k \leq \min\{m,n\}$ for each i, $\sigma > 0$

By the def. of M,

$$M - \tilde{M}_k = \sum_{i=k+1}^{m} \sigma_i u_i v_i^T$$

$$\Rightarrow \|M - \tilde{M}_k\|_2 = \sigma_{max}\left(\sum_{k+1}^{m} \sigma_i u_i v_i^T\right) = \sigma_{k+1}.$$

Now, we prove part (ii). So, assume $\tilde{M}_K$ is an arbitrary rank-k matrix. For a Matrix M, the null space of M

$$\text{Null}(M) = \{x : Mx = 0\}$$

It is well-known that for any $M \in \mathbb{M}_{m\times n}(\mathbb{R})$,

$$\text{rank}(M) + \text{null}(M) = n.$$

Since, $\tilde{M}_K$ has rank(K), we must have:

$$\text{null}(\tilde{M}_K) = n - k.$$

Now, Rayleigh-Quotient

$$\sigma_{K+1}(M)^2 = \lambda_{K+1}(M^T M) = \min_{S:(n-K)\dim} \max_{x \in S} \frac{x^T M^T M x}{x^T x}.$$

$$\leq \max_{x \in \text{NULL}(\tilde{M}_K)} \frac{x^T M^T M x}{x^T x}.$$

$$= \max_{x \in \text{NULL}(\tilde{M}_K)} \frac{x^T (M - \tilde{M}_K)^T (M - \tilde{M}_K) x}{x^T x}.$$

$$\leq \max_{x} \frac{x^T (M - \tilde{M}_K)^T (M - \tilde{M}_K) x}{x^T x}.$$

$$= \lambda_{max} \left((M - \tilde{M}_K)^T (M - \tilde{M}_K) \right)$$

$$= \sigma_{max} (M - \tilde{M}_K)^2.$$

$$\sigma_{K+1}(M)^2 \leq \sigma_{max}(M - \tilde{M}_K)^2$$

$$= \|M - \tilde{M}_K\|_2$$

$$\Rightarrow \boxed{\|M - \tilde{M}_K\|_2 \geq \sigma_{K+1}}$$

Approximation of Frobenius Norm

Defn: For any matrix M, the Frobenius norm of M is defined as follows:

$$\|M\|_F = \left(\sum_{i,j} M_{ij}^2\right)^{1/2}$$

Thm: Given a matrix $M \in \mathbb{M}_{m\times n}(\mathbb{R})$ with singular values $\sigma_1, \sigma_2, \ldots, \sigma_n$. For any integer $k \geq 1$,

$$\min_{\tilde{M}_k} \|M - \tilde{M}_k\|_2 = \sum_{i=k+1}^{m} \sigma_i^2$$

pf.

Claim: For any matrix $M \in \mathbb{M}_{m\times n}(\mathbb{R})$.

$$\|M\|_F^2 = \sum \sigma_i^2$$

$$(M^T M)_{i,i} = \sum_j M_{j,i}^2$$

$$\underbrace{\sum_{i=1}^{n} (M^T M)_{i,i}}_{= \operatorname{Tr}(M^T M) = \text{sum of eigenvalues}} = \sum_{i,j} M_{ij}^2 = \|M\|_F^2$$

$$\sum \lambda_i = \|M\|_F^2 \Rightarrow \sum \sigma_i^2 = \|M\|_F^2$$

$$\tilde{M}_k = \sum_{i=1}^{k} \sigma_i u_i v_i^T$$

$$\|M - \tilde{M}_k\|_F^2 = \left\| \sum_{k+1}^{n} \sigma_i u_i v_i^T \right\| = \sum_{i=k+1}^{m} \sigma_i^2$$

Real-World Scenario:

1. A Consumer Matrix ($M \in \mathbb{M}_{m \times n}(\mathbb{R})$) where each row corresponds to a consumer and each column corresponds to a product. Entry i, j of the matrix represents the probability that consumer i purchases product j. We can hypothesize that there are of k hidden features in consumers such as age, gender, and annual income, etc and the decision of each consumer is only a function of hidden features.

With this hypothesis, we can rewrite:

$$M = AB.$$

as product of two matrices: a factor weigh matrix, $A \in \mathbb{M}_{m \times k}(\mathbb{R})$, and a purchase probability matrix $B \in \mathbb{M}_{k \times n}(\mathbb{R})$. In particular, each row of A represents a consumer as a weighted sum of the k- underlying features, and each column of B represents the purchase probability of consumer with only one feature.

- Ideally, all elements of consumer-product matrix are available and we can use low-Rank approximation of M to find these k hidden features for a small value of k. In the Real-world however, we usually have only some of the elements of the matrix. So, our goal is to predict the unknown elements.

In the Netflix challenge, we were given a partial ratings of the Netflix users and our goal was to predict the rating of each users for the rest of the movies.

2. As each image is represented by a matrix, it is possible to compress an image by approximating it by a lower-rank-matrix

$A \in M_{m\times n}(\mathbb{R})$

$\mathrm{im}(A) = \{Ax : x \in \mathbb{R}^n\}$

$\mathrm{Col}(A) = \mathrm{span}\{\vec{a} : \bar{a} \text{ is a column of } A\}$

Lemma:

For any $m\times n$, matrix A, $\mathrm{im}(A) = \mathrm{Col}(A)$

Pf. Let $A = [a_1\ a_2, \ldots, a_n]$. Let $x \in \mathrm{im}(A) \Rightarrow x = Ay$, $y \in \mathbb{R}^n$.

If $y = [y_1, y_2, \ldots, y_n]^T$, then $Ay = y_1\vec{a}_1 + y_2\vec{a}_2 + \cdots + y_n\vec{a}_n \in \mathrm{Col}(A)$

$\Rightarrow \mathrm{im}(A) \subseteq \mathrm{Col}(A)$. each $\vec{a}_k = Ae_k \Rightarrow \mathrm{Col}(A) \subseteq \mathrm{im}(A)$.

The Standard algorithm for computing the singular value decompositon

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

is due to Golub- Reinsch (1970) & is built on ideas of Golub and Kahan (1965).

—.

Reference Book: James E. Gentle : Matrix Algebra. (Theory, Computations and applications in Statistics) (Page: 318)

—.