We have seen the magic to solve non-convex problems when the model is inherent sparse, in terms of l_0-norm of a vector.

This week, we want to understand the structure of any "sparse" problems, the low-rank matrix recovery. You will see how low rank is connected to l_0-norm a bit later.

But, Q: Do you have any intuition? ~~How~~ Why low-rank matrix can be regarded as sparse models?

A: The rank of low-rank matrix can be regarded as "sparse"

Hence, the singular value decomposition (SVD) of a matrix X, s.t.

$$X = U \underbrace{\boxed{\Sigma}}_{\text{diagonal matrix.}} V^* = \sum_{i=1}^{r} \sigma_i u_i \cdot v_i^*$$

$\|\Sigma\|_0$ is sparse!

This week, we will see the power of SVD.

Before that, Let's first look at some motivating examples. Some of them ~~are~~ have been introduced in Section I.

* Motivating Example 1: Recommendation Systems.

Imagine that we have n_2 products of interest, and n_1 users. (Taobao, JD, etc)

Users ~~com~~ consume products and rate them based on the quality of their experience.

Our goal is to use the information of all the users' ratings to predict which products will appeal to a given user.

Formally, our objective of interest is a large, unknown matrix

$$X \in \mathbb{R}^{n_1 \times n_2},$$

$$\begin{array}{cc} \text{users} & \begin{bmatrix} 5 & 3 & \cdots & ? \\ ? & 2 & \cdots & 4 \\ \vdots & \vdots & & \\ 5 & ? & \cdots & ? \end{bmatrix} \\ & \text{Products} \end{array}$$

whose (i,j) entry contains user i's rating for product j.

If we let $\Omega \doteq \{(i,j) \mid \text{user } i \text{ has rated product } j\}$,

then, we observe

$$\underset{\substack{\uparrow \\ \text{observed ratings}}}{Y} = P_\Omega [\underset{\substack{\uparrow \\ \text{complete ratings.}}}{X}]$$

$$P_\Omega[X](i,j) = \begin{cases} X_{i,j}, & (i,j) \in \Omega \\ 0, & \text{else.} \end{cases}$$

Our goal is to infer X based on the knowledge of Y.

Q: If you are given this task, what you would do?

A: The most straightforward thought is to look at the "correlation" between "similar customers". If we know the rating patterns of A and B are similar, we can make use of their observed information to complete the matrix.

Q: How to identify similar A and B's?

A: The idea ~~to~~ is to do a "clustering". ~~We~~ If we know the full information, we can group the customers into ~~a~~ a number of ~~clas~~ clusters.

Q: How many clusters do you expect to see?

A: People tend to assume that compared with the # of customers and # of products (n_1 and n_2), # of clusters are small ($\ll n_1$ and n_2).

~~This is the~~ ~~When~~ For customers ~~can~~ ~~are~~ in the same cluster, they share similar rating pattern, the rank of the matrix containing only those rows should be very small.

This is the intuition of the low rank assumption for $\underline{\underline{X}}$.

$$\min \ \operatorname{rank}(X) \quad \text{s.t.} \quad P_\Omega(X) = Y.$$

matrix completion problem.

* Representing Low-Rank Matrix via SVD.

3

Mathematically, our goal is to recover an unknown X whose columns live on an r-dimensional linear subspace of the data space $\mathbb{R}^{n_1}$.

This subspace can be characterized via the singular value decomposition (SVD) of X.

Thm (Compact SVD). Let $X \in \mathbb{R}^{n_1 \times n_2}$ be a matrix, and $r = \text{rank}(X)$.

Then there exist $\Sigma = \text{diag}(\sigma_1, \cdots, \sigma_r)$ with numbers $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r > 0$ and matrices $U \in \mathbb{R}^{n_1 \times r}$, $V \in \mathbb{R}^{n_2 \times r}$, s.t. $U^* U = I$, $V^* V = I$ and

$$X = U \cdot \Sigma \cdot V^* = \sum_{i=1}^{r} \sigma_i u_i v_i^*$$

The proof relies on the following thm, you have learnt in Linear Algebra:

Thm: Every Hermitian (Symmetric) matrix $A \in \mathbb{R}^{n \times n}$ can be diagonalized by a unitary matrix U, s.t.

$$U^* \cdot A \cdot U = \Lambda,$$

where Λ is a diagonal matrix.

Utilizing this theorem, we would like to examine the relationship between σ_i and the ~~sin~~ eigenvalues of matrix $\underline{\overset{X^*X}{\cancel{A^*A}} \ \& \ \overset{XX^*}{\cancel{AA^*}}}$.

We start from $\overset{X^*X}{\cancel{A^*A}}$. There exists V, s.t. $\underbrace{V^* \cdot \overset{X^*X}{\cancel{A^*A}} \cdot V = \Lambda}$

$$\Rightarrow \overset{X^*X}{\cancel{A^*A}} \cdot V = V \cdot \Lambda = V \cdot \text{diag}\{\lambda_1, \cdots, \lambda_r, \lambda_{r+1}, \cdots, \lambda_{n_2}\}.$$

all diagonal elements in Λ is non-negative!

$$\cancel{\lambda_1 \geq}\ \lambda_1, \cdots, \lambda_r \cancel{\geq} > 0$$

$$\lambda_{r+1} = \cdots = \lambda_{n_2} = 0.$$

Next, we construct U from V. $V = [v_1 | v_2 | \cdots | v_n]$.

For $i \leq r$, we construct.

$$u_i = \frac{\cancel{A}\, X \cdot v_i}{\sqrt{\lambda_i}}$$

$$\langle u_i, u_j \rangle = \frac{\cancel{v_i^*}\, v_j^* \cdot X^* \cdot X \cdot v_i}{\sqrt{\lambda_i \lambda_j}} = \frac{\lambda_i v_j^* v_i}{\sqrt{\lambda_i \lambda_j}} \leftarrow \text{orthogonal!}$$

Also, $\{u_i\}$ are eigen~~value~~ vectors of XX^*.

$$XX^* u_i = X \cdot X^* \cdot X \cdot \frac{v_i}{\sqrt{\lambda_i}} = X \cdot \sqrt{\lambda_i} \cdot v_i = \lambda_i \cdot u_i \, !$$

The set $\{u_i : i = 1, \cdots, r\}$ can be extended using the Gram-~~Schmidt~~ Schmidt procedure to form an orthonormal basis for $\mathbb{R}^{n_1}$. Let

$$U = [\,u_1 | \cdots | u_{n_1}\,].$$

For $i \le r$, we know

$$u_i^* X \cdot V = \frac{1}{\sqrt{\lambda_i}} \cdot v_i^* \cdot X^* \cdot X \cdot V = \frac{1}{\sqrt{\lambda_i}} \cdot v_i^* \cdot V \cdot \Lambda = \sqrt{\lambda_i} \cdot e_i^*$$

the i^{th} element is 1, all other elements are 0.

For $i > r$, $u_i^* \cdot X \cdot V = 0$.

$$\Rightarrow U^* \cdot X \cdot V = \mathrm{diag}(\sqrt{\lambda_1}, \cdots, \sqrt{\lambda_{n_1}}) = \Sigma.$$

$\boxed{\text{In fact, I directly prove a more general version}}$

Thm: Let $X \in \mathbb{R}^{n_1 \times n_2}$ be a matrix. Then there exist orthogonal matrices $U \in O(n_1)$ and $V \in O(n_2)$, and numbers

$$\sigma_1 \ge \sigma_2 \ge \cdots \ge \sigma_{\min\{n_1, n_2\}}.$$

s.t. if we ~~let~~ let $\Sigma \in \mathbb{R}^{n_1 \times n_2}$ with $\Sigma_{ii} = \sigma_i$, and $\Sigma_{ij} = 0$ for $i \ne j$

$$X = U \cdot \Sigma \cdot V.$$

Given our proof, it should be fairly easy to verify that

* The left singular vectors u_i are the eigenvectors of XX^*.

$$X \cdot X^* = (U \cdot \Sigma \cdot V)(V^* \cdot \Sigma^* \cdot U^*) = U \cdot \Sigma \cdot \Sigma^* \cdot U^*$$

* The right singular vectors v_i are the eigenvectors of $X^* X$.
* The nonzero singular values σ_i are the positive square roots of the positive eigenvalues λ_i of $X^* X$.