Netflix Signaling Game:

1) Minimally Informative Sender:

$D = \{u_1, u_2, \dots, u_m\}$

$u_i \in \mathbb{R}^k$: unknown

$M = A = \{v_1, v_2, \dots, v_n\}$

$v_j \in \mathbb{R}^k$: unknown

} vector Represent^n of u_i and v_j has to be inferred.

$$U_S(u_i, v_j, v_j) = \langle u_i, v_j \rangle \in \mathbb{R}$$

2) Non-Markovian Recommender:

Infer $\hat{U}_S : D \times M \times A \to \mathbb{R}$.

Subsets Watched $\subseteq D \times M \times A$
Labeled $\subseteq$ Watched. } $\to$ Used in Stat. Inf. $\hat{U}_S$

Loss Function: $\sum_{\alpha \in \text{Labeled}} \| U_S(\alpha) - \hat{U}_S(\alpha) \|$ ← L_2-norm.

u_i is recommended an action v_j

iff $\hat{U}_S(u_i, v_j, v_j) \geq \theta$

and $\langle u_i, v_j, v_j \rangle \notin$ Watched.

3) Non-oblivious Verifier:

Function of $\hat{U}_S(\cdot, v_j, v_j)$

or $f(\{U_S(u_i, v_j, v_j) \mid \langle u_i, v_j, v_j \rangle \in \text{labelled}\})$

→ $\alpha(u_i, V_j)$ = Pr. that u_i will reject all $v_j \in V_j$.

4) $U_S(u_i, v_j, v_j) = \langle u_i, v_j \rangle \in \mathbb{R}$

$U_R(u_i, v_j, v_j) = \text{const}.$

$$
\begin{array}{c|ccccc}
 & v_1 & v_2 & \cdots & & v_n \\
\hline
u_1 & \perp & \perp & \perp & \perp & \perp \\
u_2 & \perp & 1 & \perp\ 2 & 1 & 5 \\
\vdots & 6 & 3 & \perp & \perp & 4 \\
 & \vdots & & \cdots & & \vdots \\
u_m & \perp & \perp & 1 & \perp & \perp
\end{array}
\leftarrow \hat{U}_S
$$

$$u_S(u_i, v_j, v_j) = \begin{cases} \in \mathbb{R} & \text{if } (u_i, v_j, v_j) \in \text{Labelled} \\ \perp & \text{if Unlabelled.} \end{cases}$$

Map u_i's and v_j's to k-dim vectors

$$\underset{m \times k}{\begin{pmatrix} U \end{pmatrix}} \quad \underset{k \times n}{\left(V^T \right)} \longrightarrow UV^T = \hat{U}_S$$

$m \times n$ matrix of rank-k

→ Minimizes a loss function.

UDV^T → Diagonal $k \times k$.

SINGULAR VALUE DECOMPOSITION (SVD)

(Chapter 4 Hopcroft & Kannan pp107)

$A = n \times d$ matrix

n points in a d-dimensional space.

Sender View:
- Each point is a user and the vector $\in \mathbb{R}^d$ is his utility function.
- Represented by a row.

Receiver View: There is a different matrix $B = d \times n$ which represents the receiver's view.

$A \neq B^T$ if $U_S \neq U_R$

Item-Distance: $d_i(v_p, v_q) = \| A(\cdot p) - A(\cdot q) \|$

User Distance: $d_u(u_p, u_q) = \| B(\cdot p) - B(\cdot q) \|$

Singular- Value - Decomposition of A

= A factorization of A into product matrices

$A = UDV^T$.

U and V are orthonormal $\langle u_i, u_j \rangle = \begin{cases} 1 & \text{if } i=j \\ 0 & \text{o.w.} \end{cases}$

$\langle v_i, v_j \rangle = \begin{cases} 1 & \text{if } i=j \\ 0 & \text{o.w.} \end{cases}$

D = Diagonal with positive entries.

Data - Science Application:

Privacy Issues:

Data matrix A is close to a matrix of low rank $k \ll d \leq n$.

Goal is to find a low rank approximation $\tilde{A} \approx A$.

$$\tilde{A} = \underset{n\times k}{U} \; \underset{k\times k}{D} \; \underset{k\times d}{V^T} \qquad \|\tilde{A} - A\| < \varepsilon_k.$$

Columns of V in the SVD = Right Singular Vectors of A

= Orthogonal Set

(for any A)

$\rightarrow \in \mathbb{R}^d \rightsquigarrow$ Feature Set for V (independent)

Columns of U in the SVD = Left Singular vectors of A

= Another Orthogonal Set

(for any A, also)

$\rightarrow \in \mathbb{R}^n \rightsquigarrow$ Feature Set for U (also independent).

$$VD^{-1}U^T \times UDV^T = VD^{-1}DV^T = VV^T = I$$

$$UDV^T \times VD^{-1}U^T = UDD^{-1}U^T = UU^T = I.$$

$\therefore$ $VD^{-1}U^T$ = Inverse of A.

INSIGHT.

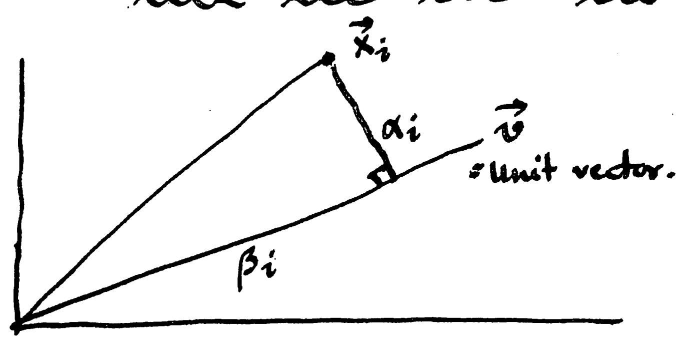
Rows of an $n \times d$ matrix A
$\equiv$ n-points in a d-dim space $\mathbb{R}^d$

Find the "best" k-dim subspace with respect to the set of points.
↳ Represented by U.

"BEST" : Minimize the Sum-of-Squares (SOS) of the perpendicular distances of the points to the subspace.

Start with $k = 1$. → 1-dim subspace = A line through origin.

→ Best Least-Square Fit.



minimize $\sum_i \alpha_i^2$
$\equiv$ maximize $\sum_i \beta_i^2$
$(\because \|\vec{x}_i\|_2 = \text{const.})$.

$\vec{a}_i$ = i^{th} row of A $\in \mathbb{R}^d$

$\beta_i = |\vec{a}_i \cdot \vec{v}|$

$\sum_i \beta_i^2 = |A\vec{v}|^2$ = Sum of length squared of the projections.

First Singular Vector v_1 of A

$$\vec{v}_1 = \underset{|v| = 1}{\arg\max} |A\vec{v}|$$

First Singular Value σ_1 of A

SOS of the projections of the points to line determined by v_1

$$\sigma_1 = |Av_1| \longrightarrow \sigma_1^2 = \sum_{i=1}^{n} (a_i \cdot v)^2 = \sum_i \beta_i^2$$

A GREEDY APPROACH:

$$\vec{v}_2 = \underset{\substack{\vec{v} \perp \vec{v}_1 \\ |\vec{v}| = 1}}{\arg\max} |Av| \; ; \qquad \sigma_2 = |A\vec{v}_2|$$

Second Singular Vector   Second Singular Value.

The 2-dimensional subspace spanned by the unit vectors $\vec{v}_1$ and $\vec{v}_2$

= Maximizes SOS of the projections of the n-points to the 2-dim subspace = span$(\vec{v}_1, \vec{v}_2)$.

$$\vec{v}_p = \underset{\substack{\vec{v} \perp \vec{v}_1, \vec{v}_2, \ldots \vec{v}_{p-1} \\ |\vec{v}| = 1}}{\arg\max} |Av| \; ; \qquad \sigma_p = |A\vec{v}_p|$$

p^{th} Singular Vector p^{th} Singular Value.

The process stops with

$\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_r$

as singular vectors and

$\sigma_1, \sigma_2, \ldots, \sigma_r$

as singular values and

$$\underset{\substack{\vec{v} \perp \vec{v}_1, \vec{v}_2, \ldots \vec{v}_r \\ |\vec{v}| = 1}}{\arg\max} |Av| = 0$$

Forbenius Norm of A.

$$\|A\|_F = \sqrt{\sum_{j,k} a_{jk}^2}$$

<u>Theorem</u>:

Let A be an $n \times d$ matrix with singular vectors $\vec{v}_1, \vec{v}_2, \cdots, \vec{v}_r$

a) For $1 \leq k \leq r$, let V_k be the subspace spanned by $\vec{v}_1, \vec{v}_2, \cdots, \vec{v}_k$

$$V_k = \text{span}(\vec{v}_1, \cdots, \vec{v}_k)$$

Then for each k, V_k is the best-fit k-dim subspace of A.

b) Let the corresponding singular values be

$$\sigma_1, \sigma_2, \cdots, \sigma_r$$

Then

$$\|A\|_F^2 = \sum_{i=1}^{r} \sigma_i^2.$$

$\square$.

Proof: a) By Induction. $k \geq 2$.

By the IH. V_{k-1} is a best-fit $(k-1)$-dim subspace

Let $W \neq V_k$ is a best-fit k-dim subspace.

Choose a basis $\vec{w}_1, \vec{w}_2, \cdots, \vec{w}_k$ of $W = \text{span}(\vec{w}_1, \cdots, \vec{w}_k)$

such that $\vec{w}_k \perp V_{k-1}$.

$\therefore \quad |A\vec{w}_k|^2 \leq |A\vec{v}_k|^2$ (by construction)

$$|A\vec{w}_1|^2 + \cdots + |A\vec{w}_k|^2 \leq |A\vec{v}_1|^2 + \cdots + |A\vec{v}_{k-1}|^2 + |A\vec{w}_k|^2$$
$$\leq |A\vec{v}_1|^2 + \cdots + |A\vec{v}_k|^2$$

If $W \neq V_k$ then the inequality must be strict, contradicting best-fitness of W.

b)

$$\sum_{i=1}^{r} (\vec{a}_j \cdot \vec{v}_i)^2 = |\vec{a}_j|^2$$

$\because$ $\text{span}(v_1, \cdots, v_r) = \text{span}(\vec{a}_1, \cdots, \vec{a}_n)$

$\forall v \perp \text{span}(v_1, \cdots, v_r)$ $|A \cdot v| = 0$

$$\sum_{j=1}^{n} |\vec{a}_j|^2 = \sum_{j=1}^{n} \sum_{i=1}^{r} (\vec{a}_j \cdot \vec{v}_i)^2$$
$$= \sum_{i=1}^{r} \sum_{j=1}^{n} (\vec{a}_j \cdot \vec{v}_i)^2 = \sum_{i=1}^{r} |Av_i|^2 = \sum_{i=1}^{n} \sigma_i^2$$

$\square$

$$\vec{u}_i = \frac{1}{\sigma_i} A \vec{v}_i$$

$\vec{v}_1, \vec{v}_2, \dots, \vec{v}_r$ = Right Singular Vectors (orthonormal by construction)
$\vec{v}_i \in \mathbb{R}^d$

$\vec{u}_1, \vec{u}_2, \dots, \vec{u}_r$ = Left Singular Vectors
$\vec{u}_i \in \mathbb{R}^n$

$\sigma_1, \sigma_2, \dots, \sigma_r$ = Singular Values.

$$\underset{n \times r}{U = (\vec{u}_1, \vec{u}_2, \dots, \vec{u}_r)} \qquad \underset{d \times r}{V = (\vec{v}_1, \vec{v}_2 \dots \vec{v}_r)}$$

$$\underset{r \times r}{D = \begin{pmatrix} \sigma_1 & & 0 \\ & \sigma_2 & \\ 0 & & \ddots \;\; \sigma_r \end{pmatrix}}$$

$$A = \sum_{i=1}^{r} \underbrace{\sigma_i \vec{u}_i \vec{v}_i^T} = U D V^T$$

↳ Each term is an nxd rank 1 matrix.

Note (Duality)

$$\vec{u}_1 = \underset{|\vec{u}|=1}{\operatorname{argmax}} |u^T A| \qquad \sigma_1 = |u_1^T A|$$

$$\vec{u}_p = \underset{\substack{\vec{u} \perp \vec{u}_1, \dots, \vec{u}_{p-1} \\ |\vec{u}|=1}}{\operatorname{argmax}} |u^T A| \qquad \sigma_p = |u_p^T A| .$$

$$\vdots$$

$$\underset{\substack{\vec{u} \perp \vec{u}_1, \dots, \vec{u}_r \\ |\vec{u}|=1}}{\operatorname{argmax}} |u^T A| = 0$$

A = rank-r matrix.

Choose $\tilde{A} = A_k = \sum_{i=1}^{k} \sigma_i \vec{u}_i \vec{v}_i^T = \tilde{U}_k \tilde{D}_k \tilde{V}_k^T$
= Truncated sum.

$\tilde{A}$ = rank-k matrix.
= Best rank-k approximation to A w.r.t. $\mathcal{L}_2$-norm or Forbenius-norm.

Lemma:

a) $\|A - A_k\|_2^2 = \sigma_{k+1}^2$

b) For any matrix B of rank at most k

$$\|A - A_k\|_F \le \|A - B\|_F$$

and $\|A - A_k\|_2 \le \|A - B\|_2$

$\square$.

How to compute SVD efficiency?

$$A = \sum_i \sigma_i u_i v_i^T$$

Hence,

$$B = A^T A = \left(\sum_i \sigma_i v_i u_i^T\right)\left(\sum_j \sigma_j u_j v_j^T\right)$$

$$= \sum_{i,j} \sigma_i \sigma_j v_i (u_i^T u_j) v_j^T$$

$$= \sum_i \sigma_i^2 v_i v_i^T$$

Taking power

$$B^2 = B^T B = \left(\sum_i \sigma_i^2 v_i v_i^T\right)\left(\sum_j \sigma_j^2 v_j v_j^T\right)$$

$$= \sum_{ij} \sigma_i^2 \sigma_j^2 v_i (v_i^T v_j) v_j^T$$

$$= \sum_i \sigma_i^4 v_i v_i^T$$

$$\therefore \quad B^k = \sum_i \sigma_i^{2k} v_i v_i^T$$

Let $\min_{i<j} \log\left(\frac{\sigma_i}{\sigma_j}\right) \geq \lambda > 0$

$$\Rightarrow \min_{i<j} \frac{\sigma_i}{\sigma_j} \geq 2^{\lambda} \quad \Rightarrow \sigma_1 \geq 2^{\lambda}\sigma_2 \geq 2^{\lambda}\sigma_3 \cdots$$

$$\therefore B^k = \sigma_1^{2k} v_1 v_1^T + \sum_{i=2}^{r} \sigma_i^{2k} v_i v_i^T$$

$$= \sigma_1^{2k} v_1 v_1^T \left(1 + O\left[\frac{\sigma_1^{2k}}{2^{2\lambda k}} \sum_{i=2}^{r} v_i v_i^T\right]\right)$$

$$\approx \sigma_1^{2k} v_1 v_1^T$$

Compute $B^k / \|B^k\|_F \longrightarrow$ Converges to a rank-1 matrix $v_1 v_1^T$

Recover $v_1 \longrightarrow$ Normalizing the first colum to be a unit vector

$$\sigma_1 = |A v_1|$$

$$u_1 = \frac{1}{\sigma_1} A v_1 \qquad \tilde{A} = A_1 = \sigma_1 u_1 v_1^T$$

Repeat with $A^{(1)} = A - A_1 = A - \sigma_1 u_1 v_1^T$.

For r steps. [or k-steps to stop with rank-k approximation $\tilde{A} = A_k$.

Problem: In our applications, $A =$ sparse.
But $B = A^T A = \boxed{\text{Dense.}}$

New idea:

Select a random vector $x_0 \in \mathcal{N}(0, I)$
$[\text{ i.e. } x(i) \in \mathcal{N}(0,1)]$

Iterate for s steps, s = Large
$$= \frac{\log\left(4 \log(2n/\delta)/\epsilon\delta\right)}{2\lambda}$$

$$\left.\begin{array}{l} \text{For } i = 1..s \text{ loop} \\ \quad x_i := A^T A x_{i-1} \\ v_1 := x_i / \|x_i\| \\ \sigma_1 = |A v_1| \\ u_1 = A v_1 / \sigma_1 \end{array}\right\}$$

$$x_i = B x_{i-1}$$
$$\downarrow$$
$$x_s = B^s x_0$$
$$\approx \left(\sigma_1^{2s} v_1 v_1^T\right) x_0$$

Note $x_0 = \sum_{i=1}^{r} c_i v_i$

$$x_s = \left(\sigma_1^{2s} v_1 v_1^T\right) \sum_{i=1}^{r} c_i v_i$$

$$\approx c_1 \sigma_1^{2s} v_1 \longrightarrow \frac{x_s}{\|x_s\|} = v_1$$

More detailed analysis:

$$\frac{|\langle x_s, v_1\rangle|}{|\langle x_s, v_i\rangle|} = \frac{|c_1|}{|c_i|}\left(\frac{\sigma_1}{\sigma_i}\right)^{2s}$$

$$\left.\begin{array}{l} \Pr\left[|c_1| > \delta/4\right] > 1 - \frac{\delta}{2} \\ \Pr\left[|c_i| \le \sqrt{\log(2n/\delta)}\right] \ge 1 - \frac{\delta}{2} \end{array}\right\} \Rightarrow \Pr\left[\frac{|\langle x_s, v_1\rangle|}{|\langle x_s, v_i\rangle|} \ge \frac{n}{\epsilon}\right] \ge 1 - \delta$$

□