Lecture 10 :- Basic Matrix Analysis

Q: Suppose you are given $A \in \mathbb{R}^{m \times n}$ (on a computer, with floating point entries). How do you compute rank(A) ?

– Gaussian Elimination ?

Ex:- Pick your favourite linear algebra package and look up the documentation/implementation of its 'rank' function.

SVD:- Singular Value Decomposition

Given: $A \in \mathbb{R}^{m \times n}$

$$v_1 := \arg\max_{\|v\| = 1} \|Av\| \quad \Big\} \text{ First singular vector}$$

$$\sigma_1 := \|Av_1\| \quad \Big\} \text{ First singular value.}$$

Caution:-

Eigenvalues of: $\begin{pmatrix} 1 & n \\ 0 & 2 \end{pmatrix}$ Diagonalizable, eigenvalues 1 and 2.

$$\begin{pmatrix} 1 & n \\ 0 & 2 \end{pmatrix} \begin{pmatrix} \cos\theta \\ \sin\theta \end{pmatrix} = \begin{pmatrix} \cos\theta + n\sin\theta \\ 2\sin\theta \end{pmatrix},$$

which has norm, $\sqrt{(\cos\theta + n\sin\theta)^2 + 4\sin^2\theta}$

So $\theta = \frac{\pi}{2}$, we already see $\underline{\sigma_1 \geq n}$.

whereas eigenvalues did not depend upon n at all!!

— × — × — × —

$$v_2 := \arg\max_{\substack{v : \|v\| = 1 \\ v \perp v_1}} \|Av\|$$

$$\sigma_2 = \|Av_2\|.$$

$\vdots$

$$v_k := \underset{\substack{v : \|v\| = 1 \\ v \perp v_1, v_2, \ldots v_{k-1}}}{\arg\max} \|Av\|$$

$$\sigma_k := A v_k$$

$1 \leq k \leq \max(m,n)$ but all except the first $\min(m,n)$ would be 0.

- This is not yet a precise definition perhaps because at each step there might be a choice of several v_i. But we will see that the σ_i are uniquely determined and the spaces spanned the vectors with same singular value are uniquely determined.

Already this tells us that we can write

$$A = \sum_{i=1}^{\min(m,n)} \sigma_i u_i v_i^T,$$ where u_i are unit vectors in the direction of Av_i (Check!).

$(u_i)_{i=1}^{k=\min(m,n)}$ Left-Singular-Vectors [Normalized, but are they orthogonal?]

$(v_i)_{i=1}^{k=\min(m,n)}$ Right-Singular-Vectors [Orthonormal by definition]

Lemma: $(u_i)_{i=1}^{k=\min(m,n)}$ are also orthogonal.

Proof: Suppose that they are not. Then let i

be the smallest index with $\sigma_i > 0$ s.t. there exists $j > i$ with $\sigma_j > 0$ s.t. u_i and u_j are not orthogonal. Pick the smallest such j: $u_i^T u_j = \tau \neq 0$.

Consider $v'(\varepsilon) := v_i + \varepsilon v_j$.

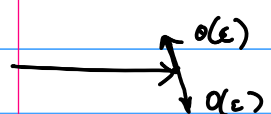
$$A v'(\varepsilon) = A v_i + \varepsilon A v_j$$
$$= \sigma_i u_i + \varepsilon \sigma_j u_j$$

$$\frac{\| A v'(\varepsilon) \|}{\| v'(\varepsilon) \|} = \frac{\| \sigma_i u_i + \varepsilon \sigma_j u_j \|}{\| v_i + \varepsilon v_j \|}$$
$$= \frac{\| \sigma_i u_i + \varepsilon \sigma_j u_j \|}{\sqrt{1 + \varepsilon^2}}$$

The point now is that if $u_i^T u_j \neq 0$ then $\| \sigma_i u_i + \varepsilon \sigma_j u_j \| \approx \sigma_i \pm O(\varepsilon)$. So by choosing ε small sign and with appropriate sign, we will get $v' = v_i + \varepsilon v_j \perp (v_1 \ldots v_{i-1})$ s.t. $\| A v' \| > \sigma_i \| v' \|$. That is a contradiction to the definition of σ_i.

$$\| \sigma_i u_i + \varepsilon \sigma_j u_j \|^2 = \sigma_i^2 + \varepsilon^2 \sigma_j^2 + 2 \varepsilon \sigma_i \sigma_j u_i^T u_j$$
$$= \sigma_i^2 + \varepsilon^2 \sigma_j^2 + 2 \varepsilon \tau \sigma_i \sigma_j$$

$\tau \neq 0$ so not zero.



$$= \sigma_i^2 \left(1 + 2 \varepsilon \tau \left(\frac{\sigma_j}{\sigma_i} \right) + \varepsilon^2 \left(\frac{\sigma_j}{\sigma_i} \right)^2 \right)$$

Thus there exist $\varepsilon_0 > 0$ (depending upon σ_i, σ_j, τ) s.t. for all $\|\varepsilon\| \leq \varepsilon_0$,

$$\|\sigma_i u_i + \varepsilon \sigma_j u_j\| \in \sigma_i \left(1 + 2\varepsilon\tau\left(\frac{\sigma_j}{\sigma_i}\right)\left(1 \pm \frac{1}{2}\right)\right)$$

and
$$\frac{1}{\sqrt{1+\varepsilon^2}} \geq 1 - 0.1\,\varepsilon\,\tau\left(\frac{\sigma_j}{\sigma_i}\right).$$

Together these will show that we have

$$\frac{\|Av'(\varepsilon)\|}{\|v'(\varepsilon)\|} > \sigma_i \quad \text{for a specific non-zero } \varepsilon,$$

which is a contradiction as argued above.

[Note: the u_i corresponding to the zero singular values can anyway be chosen to be orthogonal to the rest of the u_i]. $\blacksquare$

Now
$$A = \sum_{k=1}^{\min(m,n)} \sigma_i u_i v_i^T \quad \text{has}$$

u_i are orthonomal

v_i are orthormal.

$A = U\mathrm{diag}(s)V^T$ where u_i are the columns of U, s is the vector of the σ_i and v_i are the columns of V.

This also implies the uniqueness statements above: What one can show is that (inductively) $T_\sigma = \{ v \mid \|Av\| = \sigma\|v\| \}$ where σ is a singular value is actually a vector space.

(Suppose v_1, v_2 are right singular vectors corresponding to a singular value σ.

Then
$$A v_1 = \sigma u_1$$
$$A v_2 = \sigma u_2$$

$$A(v_1 + v_2) = \sigma(u_1 + u_2)$$

Now we know $u_1 \perp u_2$ so
$$\frac{\|A(v_1+v_2)\|}{\|v_1+v_2\|} = \frac{\sigma \|(u_1+u_2)\|}{\|v_1+v_2\|} = \sigma. \quad]$$

<u>Using this</u> :- The sequence $(\sigma_i)_{i=1}^{\min(m,n)}$ is uniquely determined by A. For each singular value σ in the sequence, the space spanned by all singular vectors with singular value σ is uniquely determined by A.

<u>Q</u>: Given a matrix $A \in \mathbb{R}^{m\times n}$, what is the matrix $B \in \mathbb{R}^{m\times n}$ of rank at most r which has the smallest possible value of
$$\|A - B\|_{2\to 2} \; ?$$

$$\|M\| = \|M\|_{2\to 2} := \max_{v \neq 0} \frac{\|Mv\|_2}{\|v\|_2} = \sigma_1(M).$$

<u>Theorem</u> :- Let $r \leq \min(m,n)$. Then for any matrix B of rank at most r, we have,
$$\sigma_{r+1} = \|A - C(A,r)\| \leq \|A - B\| \quad \text{where}$$
$$C(A,r) := \sum_{i=1}^{r} \sigma_i u_i v_i^T.$$

<u>Proof</u> :- Let us look at $D = \text{span}(v_1, v_2, \ldots v_r, v_{r+1})$
(We expand v_i to a full orthonormal

basis of $\mathbb{R}^n$ if it is not already so).

$\dim(D) = r+1$

$\ker(B) = \{v \mid Bv = 0\}$.

$\dim(\ker(B)) \geq n - r$.

$\dim(D) + \dim(\ker(B)) \geq n+1$

and D, $\ker(B)$ are both subspaces of $\mathbb{R}^n$

So there must exist a vector $w \in D \cap \ker(B)$, $w \neq 0$. Assume that $\|w\| = 1$ wlog.

$\|(A-B)w\| = \|Aw\|$ $(\because w \in \ker(B))$.

But $w \in \mathrm{span}(v_1 \ldots v_{r+1})$

$\Rightarrow w = \sum_{i=1}^{r+1} c_i v_i$; $\sum c_i^2 = 1$.

$$\|(A-B)w\| = \left\|\sum_{i=1}^{r+1} c_i A v_i\right\| = \left\|\sum_{i=1}^{r+1} c_i \sigma_i u_i\right\|$$

$$= \sqrt{\sum_{i=1}^{r+1} c_i^2 \sigma_i^2} \geq \sigma_{r+1} \sqrt{\sum_{i=1}^{r+1} c_i^2}$$

$$= \sigma_{r+1}.$$

So, $\|A-B\| \geq \sigma_{r+1}$. ■

$$\|A\|_F := \sqrt{\sum_{i,j} |A_{ij}|^2} = \sqrt{\sum_i \|A_i\|_2^2} = \sqrt{\sum_j \|A^j\|^2}$$

$$= \sqrt{\mathrm{Tr}(AA^T)}.$$

$$\|A\|_F = \sqrt{Tr\left(\left(\sum_{i=1}^{k} \sigma_i u_i v_i^T\right)\left(\sum_{i=1}^{k} \sigma_i v_i u_i^T\right)\right)}$$

$$= \sqrt{Tr\left(\sum_{i=1}^{k} \sigma_i^2 u_i u_i^T\right)}$$

$$= \sqrt{\sum_{i=1}^{k} \sigma_i^2 Tr(u_i u_i^T)} = \sqrt{\sum_{i=1}^{k} \sigma_i^2}$$

<u>Theorem</u>:- Let $r \leq \min(m,n)$. Then for any matrix B of rank at most r, we have,

$$\sqrt{\sum_{i \geq r+1} \sigma_i^2} = \|A - C(A,r)\|_F \leq \|A - B\|_F \text{ where}$$

$$C(A,r) := \sum_{i=1}^{r} \sigma_i u_i v_i^T .$$

(Proof in next class).

<u>Note on low-rank approximation</u>:-

Think of M: data matrix with Each column corresponding to a sample, and the rows corresponding to features.

Suppose this is modeled as the following.

(1) Each sample has some small number k of intrinsic attributes / 'independent variables'

(2) Each of the m features for each of the n samples is generated by a linear combination of these variables. [Simplifying assumption].

Then

$$\underbrace{M}_{m\times n} = \underset{m\times k}{A} \times \underbrace{B_{k\times n}}$$

(How to combine the k attributes, for each of the m features)

(The values of the k attributes for each sample.

Ideally $k \ll \min(m,n)$.

$\text{rank}(M) = k$.

— But we will see $\tilde{M}$, a noisy version of M. From this we would like to recover M.

<u>Proof</u> (of rank approximation in Frobenius norm).

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

Let B be any matrix of rank r, and suppose that B has the following representation in the $\{v_i\}$ basis. [assume that $\{v_i\}$ has been extended to a full basis].

$$B = \sum_i w_i v_i^T \qquad \left(Bv_i = w_i \ \forall i \right.$$

$$A - B = \sum_i (\sigma_i u_i - w_i) v_i^T.$$

So,

$$\|A-B\|_F^2 = Tr\left((A-B)(A-B)^T\right)$$

$$= Tr\left(\sum_i \sum_j (\sigma_i u_i - w_i) v_i^T v_j (\sigma_j u_j - w_j)^T\right)$$

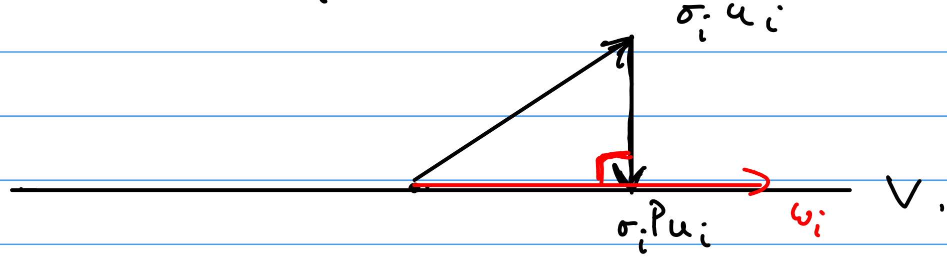
$$= Tr\left(\sum_i (\sigma_i u_i - w_i)(\sigma_i u_i - w_i)^T\right)$$

$$= \sum_i \|\sigma_i u_i - w_i\|^2$$

Let $V = \text{span}(\{w_i\})$. Then $\dim(V) = \text{rank}(B) = r$.

Let P be the orthogonal projection onto V. Then $\forall i$

$$\|\sigma_i u_i - w_i\|^2 \geq \|\sigma_i u_i - \sigma_i P u_i\|^2.$$



$$= \sigma_i^2 \|(I - P) u_i\|^2$$

$$\|A-B\|_F^2 \geq \sum_i \sigma_i^2 \|(I-P) u_i\|^2$$ where P is some orthogonal projection of rank $= r$.

Now consider the orthogonal projection $I-P$, and let $s_1 \ldots\ldots s_k$ be an orthonormal basis for i. (where $k = n - r$, where n is the dimension of the space spanned by the $\{u_i\}$).

Then, $$\|A-B\|_F^2 \geq \sum_{i=1}^n \sigma_i^2 \sum_{j=1}^k |u_i^T s_j|^2. \qquad \left(\because (I-P)u_i = \sum_{j=1}^k (u_i^T s_j) s_j\right)$$

$$= \sum_{i=1}^n \sigma_i^2 \sum_{j=1}^k c_{ij}, \qquad c_{ij} := |u_i^T s_j|^2 \geq 0.$$

where $\sum_{j=1}^k c_{ij} \leq 1 \ \forall i$ $\left(\because u_i \text{ are unit vectors}\right)$.

$$\sum_{i=1}^n c_{ij} = \sum_{i=1}^n |s_j^T u_i|^2 = \|s_j\|^2 = 1.$$ ($\{u_i\}$ are a complete basis.)

So, $$\sum_{j=1}^k \sum_{i=1}^n c_{ij} = k$$

So, we have

$$\|A-B\|_F^2 \geq \sum_{i=1}^{n} \sigma_i^2 t_i$$

where, $t_i \in [0,1] \quad \forall i \quad \left(t_i := \sum_{j=1}^{n} c_{ij} \right.$

$$\sum_{i=1}^{n} t_i = k$$

$$\geq \sum_{i=n-k+1}^{n} \sigma_i^2 .$$

(Under these constraints the smallest value is attained when the weights are all put on the k-smallest σ_i).

$$= \sum_{i=r+1}^{n} \sigma_i^2$$

$$= \|A - C(A,r)\|_F^2$$

where $C(A,r) := \sum_{i=1}^{r} \sigma_i u_i v_i^T$. ■

——— × ——— × ——— × ——— × ——— × ———