SINGULAR VALUE DECOMPOSITION

DR. MATTHIEU R BLOCH

Wednesday, November 3, 2021

LOGISTICS

General announcements

- Let me know if you have concerns about your grades (disagreements, etc.)
- Be careful with honor code

Office hours on Friday November 05, 2021

- 8am-9:30am on BlueJeans (BlueJeans)
- Focus on Midterm 2 preparation

Midterm 2:

- Moved to **Monday November 8, 2021** (gives you weekend to prepare)
- Coverage: everything since Midterm 1 (dont' forget the fundamentals though), emphasis on **regression**

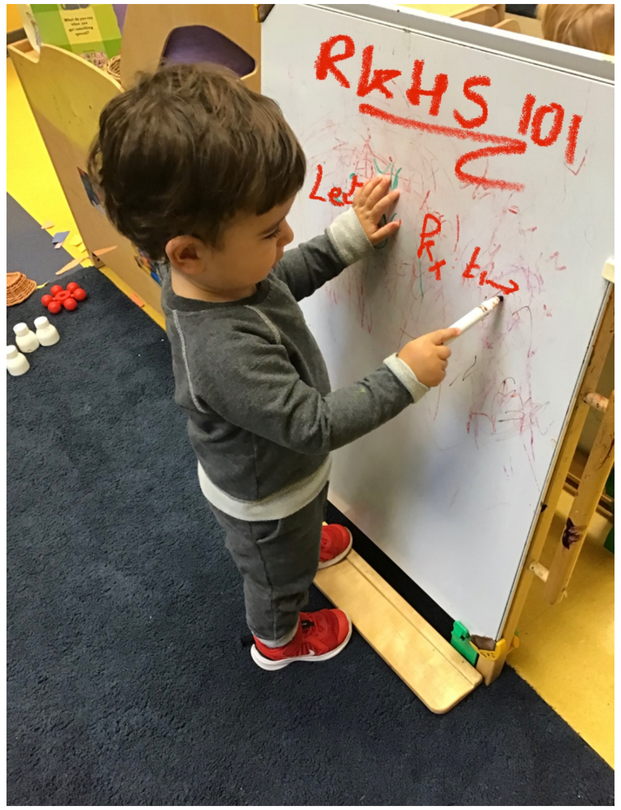
WHAT'S ON THE AGENDA FOR TODAY?

Last time:

- Symmetric matrices and spectral theorem
- *Objective:* further understand least-square problems

Today: singular value decomposition

Reading: lecture notes 12/13



Toddlers can do it!

SPECTRAL THEOREM

Lemma (Existence of eigenvector)

Every complex matrix $\mathbf{A}$ has at least one complex eigenvector and every real symmetrix matrix has real eigenvalues and at least one real eigenvector.

Lemma (Schur triangularization lemma) Every matrix $\mathbf{A} \in \mathbb{C}^{n\times n}$ is unitarily similar to an upper triangular matrix, i.e.,

$$\mathbf{A} = \mathbf{V}\boldsymbol{\Delta}\mathbf{V}^\dagger$$

with $\boldsymbol{\Delta}$ upper triangular and $\mathbf{V}^\dagger = \mathbf{V}^{-1}$.

Theorem (Spectral theorem) Every hermitian matrix is unitarily similar to a real-valued diagonal matrix.

Note that if $\mathbf{A} = \mathbf{V}\mathbf{D}\mathbf{V}^\dagger$ then

$$\mathbf{A} = \sum_{i=1}^{n} \lambda_i \underbrace{\mathbf{v}_i \mathbf{v}_i^\dagger}_{\text{rank } 1}$$

How about real-valued matrices $\mathbf{A} \in \mathbb{R}^{n\times n}$?

Every symmetric matrix can be diagonalized in an orthonormal basis of eigenvectors ($A = VDV^T$ w/ $V^TV = VV^T = I$)

SYMMETRIC POSITIVE DEFINITE MATRICES

Definition.

A symmetric matric~~e~~ $\mathbf{A}$ is positive definite if it has positive eigenvalues, i.e., $\forall i \in \{1, \cdots, n\} \lambda_i > 0$.

A symmetric matric~~e~~ $\mathbf{A}$ is positive semidefinite if it has nonnegative eigenvalues, i.e., $\forall i \in \{1, \cdots, n\} \lambda_i \geq 0$.

SYMMETRIC POSITIVE DEFINITE MATRICES

Definition.

A symmetric matrice $\mathbf{A}$ is positive definite if it has positive eigenvalues, i.e., $\forall i \in \{1, \cdots, n\} \lambda_i > 0$.

A symmetric matrice $\mathbf{A}$ is positive semidefinite if it has nonnegative eigenvalues, i.e., $\forall i \in \{1, \cdots, n\} \lambda_i \geq 0$.

Convention: $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$

Variational form of extreme eigenvalues for symmetric positive definite matrices $\mathbf{A}$

$$\lambda_1 = \max_{\mathbf{x} \in \mathbb{R}^n : \|\mathbf{x}\|_2 = 1} \mathbf{x}^\top \mathbf{A} \mathbf{x} = \max_{\mathbf{x} \in \mathbb{R}^n} \frac{\mathbf{x}^\top \mathbf{A} \mathbf{x}}{\|\mathbf{x}\|_2^2}$$

$$\lambda_n = \min_{\mathbf{x} \in \mathbb{R}^n : \|\mathbf{x}\|_2 = 1} \mathbf{x}^\top \mathbf{A} \mathbf{x} = \min_{\mathbf{x} \in \mathbb{R}^n} \frac{\mathbf{x}^\top \mathbf{A} \mathbf{x}}{\|\mathbf{x}\|_2^2}$$

Note:

$$\max_{x \in \mathbb{R}^n} \frac{x^T A x}{\|x\|_2^2} = \max_{x \in \mathbb{R}^n} \underbrace{\left(\frac{x}{\|x\|_2}\right)}_{\triangleq\, v \text{ and } \|v\|_2 = 1}^T A \left(\frac{x}{\|x\|_2}\right) = \max_{\substack{x \in \mathbb{R}^n \\ \|x\|_2 = 1}} x^T A x$$

Recall that $A = VDV^T$ so that we look for $\displaystyle\max_{x \in \mathbb{R}^n : \|x\|_2 = 1} x^T V D \underbrace{V^T x}_{\triangleq\, u}$ w/ $\|u\|_2^2 = u^T u = x^T V V^T x = x$

$$\text{i.e } \max_{u \in \mathbb{R}^n : \|u\|_2 = 1} u^T D u = \max_{u \in \mathbb{R}^n, \|u\|_2 = 1} \sum_{i=1}^{n} \lambda_i u_i^2 \quad (*)$$

$(*)$ $\displaystyle\sum_{i=1}^{n} \lambda_i u_i^2 \leq \lambda_1 \sum_{i=1}^{n} u_i^2 = \lambda_1$ achieved w/ $\underline{u} = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}$ $\left(\text{assuming } D = \begin{pmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{pmatrix}\right)$

$u = V^T x$ so that $x = Vu = \begin{pmatrix} | & | & & | \\ v_1 & v_2 & \cdots & v_n \\ | & | & & | \end{pmatrix} \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} = v_1$

SYMMETRIC POSITIVE DEFINITE MATRICES

Definition.

A symmetric matrice $\mathbf{A}$ is positive definite if it has positive eigenvalues, i.e., $\forall i \in \{1, \cdots, n\} \lambda_i > 0$.

A symmetric matrice $\mathbf{A}$ is positive semidefinite if it has nonnegative eigenvalues, i.e., $\forall i \in \{1, \cdots, n\} \lambda_i \geq 0$.

Convention: $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$

Variational form of extreme eigenvalues for symmetric positive definite matrices $\mathbf{A}$

$$\lambda_1 = \max_{\mathbf{x} \in \mathbb{R}^n : \| \mathbf{x} \|_2 = 1} \mathbf{x}^\top \mathbf{A} \mathbf{x} = \max_{\mathbf{x} \in \mathbb{R}^n} \frac{\mathbf{x}^\top \mathbf{A} \mathbf{x}}{\| \mathbf{x} \|_2^2}$$

$$\lambda_n = \min_{\mathbf{x} \in \mathbb{R}^n : \| \mathbf{x} \|_2 = 1} \mathbf{x}^\top \mathbf{A} \mathbf{x} = \min_{\mathbf{x} \in \mathbb{R}^n} \frac{\mathbf{x}^\top \mathbf{A} \mathbf{x}}{\| \mathbf{x} \|_2^2}$$

Theorem (Sylvester theorem)

For any analytic function f, we have

$$f(\mathbf{A}) = \sum_{i=1}^{n} f(\lambda_i) \mathbf{v}_i \mathbf{v}_i^\top$$

"Proof": Let A be symmetric matrix so that $A = \sum_{i=1}^{n} \lambda_i v_i v_i^T$ $(A \in \mathbb{R}^{n \times n})$ $\{v_i\}_{i=1}^{n}$ orthonormal basis of eigenvect

Then $A^2 = A \times A = \left(\sum_{i=1}^{n} \lambda_i v_i v_i^T\right)\left(\sum_{j=1}^{n} \lambda_j v_j v_j^T\right) = \sum_{i=1}^{n}\sum_{j=1}^{n} \lambda_i \lambda_j v_i \underbrace{v_i^T v_j} v_j^T = \sum_{i=1}^{n} \lambda_i^2 v_i v_i^T$

$\langle v_i, v_j \rangle = \delta_{ij} \triangleq \begin{cases} 1 & \text{if } i=j \\ 0 & \text{else} \end{cases}$

By extension $A^p = \sum_{i=1}^{n} \lambda_i^p v_i v_i^T$

$\mathrm{poly}(A) = \sum_{i=1}^{n} \mathrm{poly}(\lambda_i) v_i v_i^T$

Finally for any f analytic $f(A) = \sum_{i=1}^{n} f(\lambda_i) v_i v_i^T$

Example: $A^{-1} = \sum_{i=1}^{n} \frac{1}{\lambda_i} v_i v_i^T$ (Assuming $\forall i\ \lambda_i \neq 0$)

Note: assume v_1 eigenvector associated to λ_1 wlog $\|v_1\| = \|v_2\| = 1$

——— v_2 ——— ——— to $\lambda_2 \neq \lambda_1$

Note that
$$v_2^T A v_1 = v_2^T (A v_1) = \lambda_1 v_2^T v_1$$
$$= (v_2^T A) v_1 = (v_2^T A^T) v_1 = (A v_2)^T v_1 = \lambda_2 v_2^T v_1$$

Hence $\underbrace{(\lambda_1 - \lambda_2)}_{\neq 0} v_2^T v_1 = 0$ hence $v_2^T v_1 = 0$

In addition:
$$A^{1/2} = \sum_{i=1}^{n} \sqrt{\lambda_i}\, v_i v_i^T = V \Lambda^{1/2} V^T \qquad \exp(A) = \sum_{i=1}^{n} e^{\lambda_i} v_i v_i^T$$

A SDP

SYSTEM OF SYMMETRIC DEFINITE EQUATIONS

Consider the system $\mathbf{y} = \mathbf{A}\mathbf{x}$ with $\mathbf{A}$ symmetric positive definite

Proposition.

Let $\{\mathbf{v}_i\}$ be the eigenvectors of $\mathbf{A}$.

$$\mathbf{x} = \sum_{i=1}^{n} \frac{1}{\lambda_i} \langle \mathbf{y}, \mathbf{v}_i \rangle \mathbf{v}_i$$

Assume some observation error $\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{e}$, with $\mathbf{e}$ unknown, and we reconstruct $\mathbf{x}$ as $\widetilde{\mathbf{x}}$ by applying $\mathbf{A}^{-1}$

Proposition (Worst case bounds)

$$\frac{1}{\lambda_1^2} \mid\mid \mathbf{e} \mid\mid_2^2 \leq \|\mathbf{x} - \tilde{\mathbf{x}}\|_2 \leq \frac{1}{\lambda_n^2} \mid\mid \mathbf{e} \mid\mid_2^2.$$

Proposition (Average case bound) If $\mathbf{e} \sim \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{I})$, then

$$\mathbb{E}\left[\|\mathbf{x} - \tilde{\mathbf{x}}\|_2\right] = \sigma^2 \sum_{i=1}^{n} \frac{1}{\lambda_i^2}$$

Proof: A symmetric definite in $\mathbb{R}^{n\times n}$

In that case $y=Ax$ has a unique solution $x = A^{-1}y = \left(\sum_{i=1}^{n} \frac{1}{\lambda_i} v_i v_i^T\right) y = \sum_{i=1}^{n} \frac{1}{\lambda_i} v_i (v_i^T y)$

$$= \sum_{i=1}^{n} \frac{1}{\lambda_i} \langle v_i, y\rangle v_i$$

$\blacksquare$

Proof: $y = Ax + e$ (A symmetric definite)

$$\tilde{x} \triangleq A^{-1}y = A^{-1}(Ax+e) = x + A^{-1}e \qquad (*) \qquad \text{note that } A^{-1} = \sum_{i=1}^{n} \frac{1}{d_i} v_i v_i^T$$

The reconstruction error is $x - \tilde{x}$

$$\|x-\tilde{x}\|_2^2 = \|A^{-1}e\|_2^2 = e^T A^{-T} A^{-1} e = e^T \underbrace{\left(\sum_{i=1}^{n} \frac{1}{d_i} v_i v_i^T\right)\left(\sum_{j=1}^{n} \frac{1}{d_j} v_j v_j^T\right)}_{\sum_{i=1}^{n} \frac{1}{d_i^2} v_i v_i^T} e = e^T \sum_{i=1}^{n} \frac{1}{d_i^2} v_i v_i^T e$$

$$= \sum_{i=1}^{n} \frac{1}{d_i^2} \underbrace{e^T v_i v_i^T e}_{\triangleq \langle e, v_i \rangle^2}$$

Hence $\|x-\tilde{x}\|_2^2 \leq \frac{1}{d_n^2} \sum_{i=1}^{n} \langle e, v_i \rangle^2 = \frac{1}{d_n^2} \|e\|_2^2$

Similarly $\|x-\tilde{x}\|_2^2 \geq \frac{1}{d_{max}^2} \|e\|_2^2$

influence of noise itself

effect of A^{-1}

SYSTEM OF SYMMETRIC DEFINITE EQUATIONS

Consider the system $\mathbf{y} = \mathbf{A}\mathbf{x}$ with $\mathbf{A}$ symmetric positive definite

Proposition.

Let $\{\mathbf{v}_i\}$ be the eigenvectors of $\mathbf{A}$.

$$\mathbf{x} = \sum_{i=1}^{n} \frac{1}{\lambda_i} \langle \mathbf{y}, \mathbf{v}_i \rangle \mathbf{v}_i$$

Assume some observation error $\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{e}$, with $\mathbf{e}$ unknown, and we reconstruct $\mathbf{x}$ as $\tilde{\mathbf{x}}$ by applying $\mathbf{A}^{-1}$

Proposition (Worst case bounds)

$$\frac{1}{\lambda_1^2} \| \mathbf{e} \|_2^2 \leq \|\mathbf{x} - \tilde{\mathbf{x}}\|_2 \leq \frac{1}{\lambda_n^2} \underbrace{\| \mathbf{e} \|_2^2}.$$

Proposition (Average case bound) If $\mathbf{e} \sim \mathcal{N}(\mathbf{0}, \sigma^2\mathbf{I})$, then

$$\mathbb{E}\left[\|\mathbf{x} - \tilde{\mathbf{x}}\|_2\right] = \sigma^2 \sum_{i=1}^{n} \frac{1}{\lambda_i^2} = \overbrace{n\sigma^2}^{\mathbb{E}(\|e\|_2^2)} \underbrace{\frac{1}{n} \sum_{i=1}^{n} \frac{1}{\lambda_i^2}}$$

SINGULAR VALUE DECOMPOSITION

What happens for non-square matrice?

Theorem (Singular value decomposition)

Let $\mathbf{A} \in \mathbb{R}^{m\times n}$ with $\operatorname{rank}(\mathbf{A}) = r$. Then $\mathbf{A} = \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^T$ where

- $\mathbf{U} \in \mathbb{R}^{m\times r}$ such that $\mathbf{U}^\intercal\mathbf{U} = \mathbf{I}_r$ (orthonormal columns) (in general $UU^T \neq I$)
- $\mathbf{N} \in \mathbb{R}^{n\times r}$ such that $\mathbf{V}^\intercal\mathbf{V} = \mathbf{I}_r$ (orthonormal columns) (in general $VV^T \neq I$)
- $\boldsymbol{\Sigma} \in \mathbb{R}^{r\times r}$ is diagonal with *positive entries*

$$\boldsymbol{\Sigma} \triangleq \begin{bmatrix} \sigma_1 & 0 & 0 & \cdots \\ 0 & \sigma_2 & 0 & \cdots \\ \vdots & & \ddots & \\ 0 & \cdots & \cdots & \sigma_r \end{bmatrix}$$

and $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r > 0$. The σ_i are called the *singular values*

•

We say that $\mathbf{A}$ is full rank is $r = \min(m, n)$

•

We can write $\mathbf{A} = \sum_{i=1}^{r} \sigma_i \mathbf{u}_i \mathbf{v}_i^\intercal$

Proof: $A \in \mathbb{R}^{m\times n}$; from $A^TA \in \mathbb{R}^{n\times n}$, is symmetric and can be written $A^TA = \sum_{i=1}^{n} \lambda_i v_i v_i^T$

Note $\forall x \in \mathbb{R}^n$ $x^TA^TAx = \|Ax\|_2^2 \geq 0$ so that $\forall i \in [\![1;n]\!]$ $\lambda_i \geq 0$

Assume $\operatorname{rank}(A) = r$ then $\operatorname{rank}(A^TA) = r$ (prove it!) so that $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_r > 0$ $\lambda_{r+1} \cdots \lambda_n = 0$

For $i \in [\![1;r]\!]$ set $u_i \stackrel{\Delta}{=} \dfrac{1}{\sqrt{\lambda_i}} A v_i$

① Check $\{u_i\}_{i=1}^{r}$ are orthonormal

$$\forall\ 1 \leq i \leq j \leq r \qquad u_i^T u_j = \frac{1}{\sqrt{\lambda_i\lambda_j}} v_j^T A^T A v_i = \frac{1}{\sqrt{\lambda_i\lambda_j}} v_j^T \left(\sum_{\ell=1}^{n} \lambda_\ell v_\ell v_\ell^T\right) v_i = \begin{cases} 1 & \text{if } j=i \\ 0 & \text{else} \end{cases}$$