CPSC 531F March 23, 2021

SVD - Singular Value Decomposition

Think of SVD!

- If $A \in M_{n,n}(\mathbb{R})$ is symmetric,

 then A has ON eigenbasis,

 $A\vec{v}_i = \lambda_i \vec{v}_i$, $\vec{v}_1, \ldots, \vec{v}_n$ ON

 then (spectral theorem)

$$\boxed{A = \sum_{i=1}^{n} \lambda_i \vec{v}_i \vec{v}_i^T}$$

(used this for $A = A_G$, G graph,

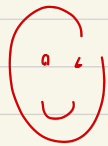
especially nice d-regular graphs).

If A is not symmetric?

- General A has Jordan canonical form

$$\begin{pmatrix} \lambda & 1 & & \\ & \lambda & 1 & \\ & & \ddots & \ddots \\ & & & \lambda \end{pmatrix}$$

recurrences, ODE with const coefs



most other times

- "Most $A \in M_n(\mathbb{R})$" have distinct eigenvalues

Reasonable things for A not symmetric

① If A is self-adjoint wrt $(,)$ inner product, then A has real eigenvalues, ON eigenbasis, ON $\leftarrow (,)$

② Perron-Frob for A with non-neg entries (A irreducible) you get a nice λ_1 = means "capacity" info theory

③ If A is general $M_n(\mathbb{R})$ you have biorthogonal decomposition

(4) For any $A \in M_{m,n}(\mathbb{R})$ there is SVD.

- Reduce information (dimension reduction)
- Principal Components

⋮

(4) & (1) have associated "variational principles"

Consider $A \in M_{m,n}(\mathbb{R})$!

$$f(\vec{u},\vec{v}) := \| A - \underbrace{\vec{u}\,\vec{v}^T}_{\text{rank 1 matrix}} \|^2_{Frob}$$

minimize f.

"Aside"

Imagine you are given:

$$\vec{x}_1, \vec{x}_2, \ldots, \vec{x}_m \in \mathbb{R}^n$$

n large, m large. Say you want:

$k < n$ (probably k small, maybe $k=1$)

want

$$\vec{y}_1, \ldots, \vec{y}_m \in \mathbb{R}^k$$

such that $\vec{y}_i$ "represents $\vec{x}_i$"

"as best as possible"

More precisely:

we want

$$\mathcal{L} : \mathbb{R}^k \longrightarrow \mathbb{R}^n \quad (\text{linear})$$

s.t.

$$\boxed{F(\vec{y}_1, \dots, \vec{y}_m, \mathcal{L}) = \sum_{i=1}^{m} |\vec{x}_i - \mathcal{L}\vec{y}_i|^2}$$

why? is this a good measure of approximation

is minimized over all $\vec{y}_1, \dots, \vec{y}_m \in \mathbb{R}^k$

and $\mathcal{L} : \mathbb{R}^k \longrightarrow \mathbb{R}^n$.

"Reducing from $\mathbb{R}^n$ to $\mathbb{R}^k$"

Thm: Such $\vec{y}_1, \dots, \vec{y}_m, \mathcal{L}$ exist

(given $\vec{x}_1, \dots, \vec{x}_m \in \mathbb{R}^n$ and $k \in \mathbb{N}$)

AND ACTUALLY

$$\vec{y}_i = \hat{\mathcal{L}} \, \vec{x}_i$$

FOR A LINEAR TRANSFORMATION

$$\hat{\mathcal{L}} : \mathbb{R}^n \to \mathbb{R}^k$$

!!!

We find $\hat{\mathcal{L}}$, $\mathcal{L} : \mathbb{R}^k \to \mathbb{R}^n$

via SVD.

(today: mostly do $k=1$)

Why

$$F = \sum_{i=1}^{m} \left| \vec{x}_i - \mathcal{L}\, \vec{y}_i \right|_2^2$$

Sum of squares error ?

(Also typical of - linear regression

- modeling

$f(x)$ via $a + bx + cx^2$ ...

You want $f(\ ) =$ mean of error

1. non-negative
2. zero $\iff$ exact fit

- Maybe this F is good for Gaussian

Warning! In some applications ...

$$F = \max_i \| \vec{x}_i - R\vec{y}_i \|_p$$

maybe $p = 1, \infty$

Σ ...

Really — choosing F as we do

leads to an easy computation —

maybe related to ... SDV

$$\| A - u v^T \|_{Frob} \ldots$$

Variational argument:

Say that $\vec{u}, \vec{v}$ chosen so that

$$\| A - \vec{u}\,\vec{v}^T \|_{Frob} \quad \text{is smallest}$$

(last time $\vec{u}^*, \vec{v}^*$) $\quad \vec{u} \in \mathbb{R}^m, \ \vec{v} \in \mathbb{R}^n$.

Then for any $\vec{\omega} \in \mathbb{R}^n$

$$g(\varepsilon) = \| A - \vec{u}\,(\vec{v} + \varepsilon \vec{\omega})^T \|^2_{Frob}$$

has a minimum at $\varepsilon = 0$;

claim:

$$\left. g(\varepsilon) = c_0 + \varepsilon c_1 + \varepsilon^2 c_2 \ \right\} \text{depends on } \vec{\omega}$$

and hence $\mathcal{E}_1 = 0$ (for all $\vec{w}$)

$=$

Frob norm: $\|B\|_{Frob} = \sqrt{\sum |b_{ij}|^2}$

$= \sqrt{Tr(BB^T)}$

$= \sqrt{Tr(B^TB)}$

Use

$Tr(BC) = Tr(CB)$

$B \in M_{\cancel{k},\ell}$, $C \in M_{\ell,k}$

$$g(\varepsilon) = Tr\left(\overbrace{\left(\underbrace{A - \vec{u}\vec{v}^T}_{} - \varepsilon \vec{u}\vec{w}^T\right)}^{T} \left(\underbrace{A^T - \vec{v}\vec{u}^T}_{} - \underbrace{\varepsilon \vec{w}\vec{u}^T}_{\varepsilon}\right) \right)$$

ε term:

$Tr\left((A - \vec{u}\vec{v}^T) \vec{w}\vec{u}^T \right)$

Same

$Tr(B) = Tr(B^T)$

$$g(\varepsilon) = C_0 + \varepsilon C_1 + \varepsilon^2 C_2 \text{ then}$$

$$C_1 = (-2)\,\mathrm{Tr}\left((A - \vec{u}\vec{v}^T)\,\vec{w}\,\vec{u}^T \right)$$

$$C_1 = 0 \implies \text{for any } \vec{w}\,!$$

$$\mathrm{Tr}\left((A - \vec{u}\vec{v}^T)\,\vec{w}\,\vec{u}^T \right) = 0$$

$$(*)\quad \mathrm{Tr}\Big(\underset{m\times n}{A}\, \underbrace{\vec{w}\,\vec{u}^T}_{n\times m} \Big) = \mathrm{Tr}\Big(\underbrace{\vec{u}\,\vec{v}^T}_{m\times n}\, \vec{w}\, \underbrace{\vec{u}^T}_{n\times m} \Big)$$

$$\mathrm{Tr}(BC) = \mathrm{Tr}(CB)$$

$$\mathrm{Tr}(\vec{u}^T A \vec{w})$$

$$\vec{v}^T w = (\vec{v}\cdot\vec{w})$$

$$\mathrm{Tr}(\vec{u}\,\vec{u}^T) = \mathrm{Tr}(\vec{u}^T u)$$

So

2 real numbers are equal

$$\boxed{(*) \quad \vec{u}^T A \vec{w} = (\vec{v} \cdot \vec{w})(\vec{u} \cdot \vec{u})}$$

So!

(i) if $\vec{w} \perp \vec{v}$ then $\vec{v} \cdot \vec{w} = 0$,

$$(A^T u) \cdot \vec{w} = (u^T A)\vec{w} = 0$$

So $$\vec{u}^T A = \beta \vec{v}^T$$

$$A^T \vec{u} = \beta \vec{v}$$

Find β!

$$\underbrace{(A^T\vec{u})\cdot\vec{v}}_{} = \beta(\vec{v}\cdot\vec{v})$$

$$\vec{u}^T A \vec{v}$$

$$(*) \qquad \parallel$$

$$\vec{u}^T A \vec{w}\Big|_{\vec{w}=\vec{v}} = (\vec{v}\cdot\vec{v})(\vec{u}\cdot\vec{u})$$

$$\Rightarrow \text{ if } \vec{v}\cdot\vec{v}=0$$

$$\beta = \vec{u}\cdot\vec{u}.$$

varied $\vec{v}$

Hence: $A - u\vec{v}^T$ via $\vec{v} + \varepsilon\vec{w}$

$$\Rightarrow \boxed{A^T \vec{u} = \beta \vec{v}}$$

1

$\vec{u}\cdot\vec{u}$

2 Vary $A - \vec{u}\vec{v}^T$ $\vec{u} + \varepsilon\vec{w}'$

and conclude ...

Break 4 minutes

For any $\vec{w}'$

$$\mathrm{tr}(\ \ ) = \boxed{\underbrace{\vec{v}^T A \vec{w}'}_{\text{row}} = (u \cdot w')(v \cdot v)}$$

$=$

$$\| A - u\vec{v}^T \|_{\text{Frob}} = \| (A - u\vec{v}^T)^T \|_{\text{Frob}}$$

$$\mathrm{Tr}(BB^T) = \mathrm{Tr}(B^T B) = \sum |b_{ij}|^2$$

$$\| A^T - \vec{v}\,\vec{u}^T \|_{\text{Frob}}$$

Make same conclusion (reminiscent of LP duality)

$$A \to A^T, \quad \vec{v} \to \vec{u}, \quad \vec{u} \to \vec{v}$$

$$\boxed{\begin{array}{c} ② \quad A\vec{v} = \alpha\,\vec{u} \\ \uparrow \\ \vec{v}\cdot\vec{v} \end{array}}$$

In particular:

$$A A^T \vec{u} = A \beta \vec{v}$$

$$\underbrace{(A A^T)}_{\substack{\text{symmetric} \\ m \times m \\ \text{mat}}} \vec{u} = \beta A \vec{v} = \beta \alpha \vec{u}$$

$$(AA^T)^T = (A^T)^T A^T = AA^T$$

1 extra

$$A^T\vec{u} = \beta\vec{v}$$

$$\beta = \vec{u}\cdot\vec{u}$$

$$(AA^T)\vec{u} = \underbrace{\beta\alpha\vec{u}}_{\lambda\vec{u}}$$

2 extra

$$A\vec{v} = \underset{\vec{v}\cdot\vec{v}}{\underset{\uparrow}{\alpha}}\vec{u}$$

$$(A^TA)\vec{v} = \underbrace{\alpha\beta\vec{v}}_{\lambda\vec{v}}$$

Other approach!

$$(AA^T)\vec{u} = \lambda\vec{u}$$

$$\lambda = \alpha \cdot \beta$$

$$(A^TA)\vec{v} = \lambda\vec{v}$$

$$= (u \cdot u)(v \cdot v)$$

Claim! λ is the largest eigenvalue of both A^TA and AA^T.

"Reverse Engineering" $A \in M_{m,n}(\mathbb{R})$

look at $B = A^TA$.

B is symmetric, has ON basis

$\vec{v}_1, \vec{v}_2, \dots, \vec{v}_n, \qquad A^T A \vec{v}_i = \lambda_i \vec{v}_i$

Claim: $\lambda_i \geq 0$!

$$(\lambda_i \vec{v}_i) \cdot \vec{v}_i = (A^T A \vec{v}_i) \cdot \vec{v}_i$$

$$\|$$

$$\lambda_i \underbrace{(\vec{v}_i \cdot \vec{v}_i)}_{\text{positive}}$$

$= 1$

ON eigenbasis

$$= \vec{v}_i^T A^T A \vec{v}_i$$

$$= (\vec{v}_i^T A^T)(A \vec{v}_i)$$

$$= (A \vec{v}_i) \cdot (A \vec{v}_i)$$

$$\geq 0$$

So $\lambda_i \geq 0$

Arrange λ_i's:

$$\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n \geq 0$$

Def: $\sqrt{\lambda_i}$ denoted σ_i

are the singular-values of A,

① Why SV-decomposition

② Dimension reduction

$$\vec{x}_1, \cdots \vec{x}_m \in \mathbb{R}^n \longrightarrow \mathbb{R}^{\text{low dim}}$$

Class Ends