CPSC 536F Nov 19, 2025

- max-min, min-max for $\mathcal{R}_A$
- SVD: $A = U \Sigma V^T$, low rank approximation.
- Ridge regression

Classic: Many data points, few features

New: Few data points, many features

e.g. 100 pictures of many pixels

Dual form (not the usual dual)

- Start! "Nonparametric" Stats & Kernel Functions (2 important examples, 1 is not really about

Some details:

k^{th} largest eigenvalue of A

$$= \max_{\substack{\dim(\bar{V}) = k \\ \bar{V} \subset \mathbb{R}^n}} \left(\min_{\substack{\vec{v} \in \bar{V} \\ \vec{v} \neq \vec{0}}} R_A(\vec{v}) \right)$$

k^{th} smallest

$$= \min_{\substack{\dim(\bar{V}) = k \\ \bar{V} \subset \mathbb{R}^n}} \left(\max_{\substack{\vec{v} \in \bar{V} \\ \vec{v} \neq \vec{0}}} R_A(\vec{v}) \right)$$

SVD: Consider

$$A^T A = \text{(we hope)} \; V \Sigma^T \Sigma V^T$$

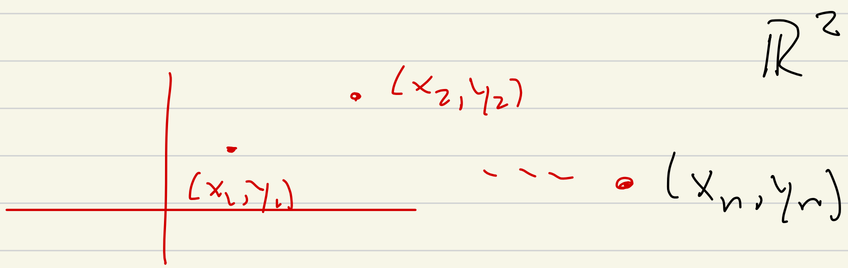
Regression: $A \in \mathbb{R}^{m \times n}$

$$f(\vec{x}) = (A\vec{x} - \vec{b}) \cdot (A\vec{x} - \vec{b}) + C\,\vec{x} \cdot \vec{x}$$

min attched at $\vec{x} = \vec{u} \implies$

$$\underbrace{A^T A \vec{u} + C \vec{u}}\, = A^T b$$

$$(A^T A + C I)\,\vec{u} = A^T b$$



Model: $y = w_1 f_1(x) + w_2 f_2(x) + w_3 f_3(x)$

$$A\vec{w} = \vec{b}, \quad A = \begin{bmatrix} | & | & | \\ f_1 & f_2 & f_3 \\ | & | & | \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \\ w_3 \end{bmatrix} = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}$$

What if

$$A = \begin{bmatrix} | & | & | \\ f_1 & f_2 & f_3 \\ | & | & | \end{bmatrix} \in \mathbb{R}^{n \times 3}$$

$$\in \mathbb{R}^{n \times}$$

There is an ON eigenbasis $\vec{u}_1, \vec{u}_2, \ldots, \vec{u}_n$ for any $A \in \mathbb{R}^{n \times n}$ symmetric, with corresponding eigenvalues $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_n$

$$A \begin{bmatrix} | & & | \\ \vec{u}_1 & \ldots & \vec{u}_n \\ | & & | \end{bmatrix} = \begin{bmatrix} | & & | \\ \vec{u}_1 & & \vec{u}_n \\ | & & | \end{bmatrix} \begin{bmatrix} \lambda_1 & & & \\ 0 & \lambda_2 & & \\ 0 & & \ddots & \\ 0 & & & \lambda_n \end{bmatrix}$$

$$A U = U \underline{\Lambda}, \quad \underline{\Lambda} = \begin{bmatrix} \lambda_1 & & & \\ & \ddots & & \\ & & \ddots & \\ & & & \lambda_n \end{bmatrix}$$

and

$U U^T = U^T U = I$, U is <u>orthogonal matrix</u>

Now let

$$A = \begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{bmatrix}$$

$$(U\vec{x}) \cdot (U\vec{y}) = \vec{x} \cdot (U^T U \vec{y})$$

U ortho

$$= \vec{x} \cdot \vec{y}$$

So (dot products) preserved under U

(length) " " "

(cos of angles between 2 vectors) " " U

min/max , max/min $\qquad \lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_n$

① If $A = \begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{bmatrix}$

then

$$\lambda_1 = \max_{\substack{\vec{v} \in \mathbb{R}^n \\ (\vec{v} \neq \vec{0})}} R_A(\vec{v})$$

$$= \max_{\substack{\bar{V} \subset \mathbb{R}^n \\ \dim(\bar{V}) = 1}} \left(\min_{\vec{v} \in \bar{V}} R_A(\vec{v}) \right)$$

$$\lambda_2 = \max_{\substack{\bar{V} \subset \mathbb{R}^n \\ \dim(\bar{V}) = 2}} \left(\min_{\vec{v} \in \bar{V}} R_A(\vec{v}) \right)$$

Proof:

Say that $\vec{v} = \begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{pmatrix}$

$\vec{v} = v_1 \vec{e}_1 + \dots + v_n \vec{e}_n$

$\vec{e}_1, \dots, \vec{e}_n$ standard basis vectors

$A \vec{e}_i = \lambda_i \vec{e}_i$

since $A = \begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{bmatrix}$

$$R_A(\vec{v}) = \frac{(A\vec{v}) \cdot \vec{v}}{\vec{v} \cdot \vec{v}}$$

$$= \frac{\lambda_1 v_1^2 + \lambda_2 v_2^2 + \dots + \lambda_n v_n^2}{v_1^2 + v_2^2 + \dots + v_n^2}$$

since: $(A\vec{v}) \cdot \vec{v} = \left(\begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{bmatrix} \begin{bmatrix} v_1 \\ \vdots \\ v_n \end{bmatrix} \right) \cdot \begin{bmatrix} v_1 \\ \vdots \\ v_n \end{bmatrix}$

$$= \begin{bmatrix} \lambda_1 v_1 \\ \lambda_2 v_2 \\ \vdots \\ \lambda_n v_n \end{bmatrix} \cdot \begin{bmatrix} v_1 \\ \vdots \\ v_n \end{bmatrix}$$

So

$$R_A(\vec{v}) = \lambda_1 \left(\frac{v_1^2}{v_1^2 + \dots + v_n^2} \right)$$

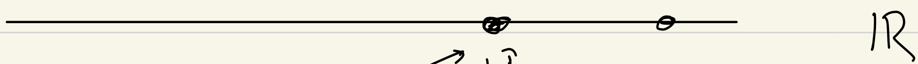
$$+ \lambda_2 \left(\frac{v_2^2}{v_1^2 + v_2^2 + \dots + v_n^2} \right)$$

$$+ \dots$$

$$+ \lambda_n \left(\frac{v_n^2}{v_1^2 + \dots + v_n^2} \right)$$

= a weighted ("probabilistic" "convex") linear combo of $\lambda_1, \dots, \lambda_n$

Rem: For a general symmetric $A \in \mathbb{R}^{n \times n}$



$$\lambda_n \leq \ldots \leq \lambda_4 \leq \lambda_3 \leq \lambda_2 \leq \lambda_1$$

So...

let $\bar{V} \subset \mathbb{R}^n$ be a subspace

of dimension 2, $\bar{V} = \operatorname{span}(\vec{w}_1, \vec{w}_2)$

Claim: $\exists\, \vec{w} \in \bar{V}$ of the form

$$\vec{w} = \begin{bmatrix} 0 \\ w_2 \\ w_3 \\ \vdots \\ w_n \end{bmatrix}$$

Why: $\alpha_1, \alpha_2 \in \mathbb{R}$

$(\alpha_1 \vec{w}_1 + \alpha_2 \vec{w}_2) \perp \vec{e}_1$

$\alpha_1 (\vec{w}_1 \cdot \vec{e}_1) + \alpha_2 (\vec{w}_2 \cdot e_2) = 0$

one equation, 2 unknowns

Rem: If $\dim(\bar{V}) = 3$

$= \text{Span}(\vec{w}_1, \vec{w}_2, \vec{w}_3)$

then $\exists\ \vec{w} \in \bar{V}$ s.t.

$$\vec{w} = \begin{bmatrix} 0 \\ 0 \\ w_3 \\ w_4 \\ \vdots \\ w_n \end{bmatrix}$$

Why? we want $\alpha_1, \alpha_2, \alpha_3 \in \mathbb{R}$ s.t.

$\alpha_1 \vec{w}_1 + \alpha_2 \vec{w}_2 + \alpha_3 \vec{w}_3 \perp \vec{e}_1, \vec{e}_2$

$\alpha_1 (\vec{w}_1 \cdot \vec{e}_1) + \alpha_2 (\vec{w}_2 \cdot \vec{e}_1) + \alpha_3 (\vec{w}_3 \cdot \vec{e}_1) = 0$

-- $\vec{e}_2$ -- $\vec{e}_2$ --

2 linear equations, 3 unknowns ...

<u>Upshot:</u> If $\vec{V} \subset \mathbb{R}^n$, $\dim(\vec{V}) = d$, then $\exists\ \vec{w} \in \vec{V}$

s.t.

$$R_A(\vec{w}) = R_{\begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{bmatrix}}(\vec{w})$$

= convex combination of $\lambda_d, \lambda_{d+1}, \ldots, \lambda_n$, namely

$$\vec{w} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ w_d \\ w_{d+1} \\ \vdots \\ v_n \end{bmatrix}, \quad R_A(\vec{w}) = \begin{aligned} &\lambda_d \left(\quad \right) \\ + &\lambda_{d+1} \left(\quad \right) \\ &\vdots \qquad \vdots \\ + &\lambda_n \left(\quad \right) \end{aligned}$$

non-neg reals, sum = 1

$$R_A(\vec{w}) \le \lambda_d$$

since $\lambda_d \geq \lambda_{d+1} \geq \cdots \geq \lambda_n$

Hence

$$\lambda_d \le \min_{\vec{w} \in \overline{W}} R_A(\vec{w})$$

for any $\overline{W} \subset \mathbb{R}^n$

$\dim(\overline{W}) = d$.

And ... to be continued

on Friday ...