CPSC 536 Nov 7, 2025

Last time:

Ridge Regression:

$A \in \mathbb{R}^{m \times n}$, $\vec{b} \in \mathbb{R}^m$

$\vec{u} \in \mathbb{R}^n$, var $\vec{v} \in \mathbb{R}^n$

$$\min_{\vec{u}} \left[\underbrace{\underbrace{\|A\vec{u} - \vec{b}\|^2}_{f_1(\vec{u})}}_{\text{model}} + C \underbrace{\underbrace{\|u\|^2}_{f_2(\vec{u})}}_{\text{controls too many parameters}} \right]$$

If $n \geq 4$ then ☹ tend to fit to noise

Saw:

$$f_1(\vec{u} + \varepsilon \vec{v}) = f_1(\vec{u}) + \varepsilon \underbrace{\text{Term}_{1,f_1}}_{2\vec{v} \cdot A^T(A\vec{u} - \vec{b})} + O(\varepsilon^2)$$

Hence:

$f_2(\vec{u} + \varepsilon \vec{v})$ similarly, with

$$f_2(\vec{u}+\varepsilon\vec{v}) = f_2(\vec{u}) + \varepsilon \, \text{Term}_{1,f_2} + O(\varepsilon^2)$$

$$\text{Term}_{1,f_2} = 2\vec{v} \cdot I^T(I^T u - \vec{0})$$

(same as f_1, $A \to I$, $b \to \vec{0}$)

$$\boxed{\begin{array}{l} \text{Rem:} \\ \min \; C_1 \|A_1\vec{u} - \vec{b}_1\|^2 + C_2\|A_2\vec{u} - \vec{b}_2\|^2 + \cdots \\ \qquad + C_k\|A_k\vec{u} - \vec{b}_k\|^2 \\ \text{get similar} \\ \qquad 2\vec{v} \cdot \sum_{i=1}^{k} C_i A_i^T (A_i\vec{u} - \vec{b}_i) \end{array}}$$

Today: (1) Examine Ridge Regression

(2) λ's of Symmetric Matrices

Admin:

I'll write a guide to what to do from Exercises D, E, maybe F.

Now: I'm adding to D for future uses.

E = current unit

F = anything else (infinite kernels, Gaussian, etc.

Ridge Regression:

$$f_1(u) = \|A\vec{u} - \vec{b}\|^2$$

$$f_2(u) = \|I\,\vec{u} - \vec{0}\|^2 = \|u\|^2$$

$$\min: \underbrace{f_1 + C f_2}_{g} \qquad 0 \le C < \infty$$

$$g(\vec{u} + \varepsilon\vec{v}) = g(\vec{u}) +$$

$$\varepsilon(2\vec{v}) \cdot A^T(A\vec{u} - \vec{b})$$
$$+ \varepsilon(2\vec{v}) \cdot C\, I^T(I\vec{u} - \vec{0}) \quad + O(\varepsilon^2)$$

$$\varepsilon(2\vec{v}) \cdot \left(A^T(A\vec{u} - \vec{b}) + C\vec{u}\right)$$

If $g(\vec{u})$ has a min at $\vec{u}^*$,

then

$\forall\, \vec{v} \in \mathbb{R}^n$

$$(2\vec{v}) \cdot \underbrace{\left(\begin{array}{c} \text{1st} \\ \text{variation} \\ \text{stuff} \end{array} \right)}_{\text{must be } \vec{0}} = 0$$

$$\Longrightarrow \quad \left(A^T(A\vec{u} - \vec{b}) + C\vec{u} \right) \overset{\text{must be}}{=} \vec{0}$$

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

$\Rightarrow$

$$\left(A^T A + c I\right)\vec{u} = A^T\vec{b}$$

$$\vec{u} = \left(A^T A + c I\right)^{-1} A^T\vec{b}$$

$c = 0$

"normal equations":

$$A^T A\vec{u} = A^T\vec{b}$$

or if $A^T A$ is "well-conditioned"

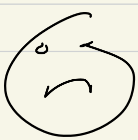
If columns A are features, and

if

$$A = \begin{bmatrix} 5\,\text{m} & 500\,\text{cm} \\ 2\,\text{m} & 200\,\text{cm} \\ 30\,\text{m} & 3000\,\text{cm} \\ -1\,\text{m} & -100\,\text{cm} \end{bmatrix}$$

then these 2 columns are

linearly dependent features,



If same but with noise

$$A = \begin{bmatrix} 5\,m & 500\,cm + \varepsilon_1 \\ 2\,m & 200\,cm + \varepsilon_2 \\ 30\,m & 3000\,cm + \varepsilon_3 \\ -1\,m & -100\,cm + \varepsilon_4 \\ & \vdots \end{bmatrix}$$

ε_i = noise

Fitting to $\begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \end{bmatrix}$ ☹ (exercise ?)

Key:

$$\vec{u} = \left(A^T A + C\mathbb{I}\right)^{-1} A^T \vec{b}$$

maybe C big enough s.t.

$$A^T A + C\mathbb{I} = C\left(\frac{A^T A}{C} + \mathbb{I}\right)$$

is invertible, so inverse is

$$\frac{1}{C}\left(\mathbb{I} + \frac{A^T A}{C}\right)^{-1}$$

$$= \frac{1}{C}\left(\mathbb{I} - \left(\frac{A^T A}{C}\right) + \left(\frac{A^T A}{C}\right)^2 - \cdots\right)$$

PageRank damping factor is analogous ...

A symmetric, $A \in \mathbb{R}^{n \times n}$,

looking at

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

want to minimize, maximize

Basic idea: we believe that eigenvalues of A all real,

$$\lambda_n(A) \le \ldots \le \lambda_2(A) \le \lambda_1(A).$$

Find $\lambda_1(A)$ largest:

$$\max_{\vec{u} \neq \vec{0}} R_A(\vec{u}) = \lambda_1(A)$$

and wherever max occurs, say $\vec{u}^*$, we'll have

$$A\vec{u}^* = \lambda_1 \vec{u}^*.$$

We said, thru simplifying,

say $\vec{u}^*$ is maximum

$$\vec{u}^* \rightsquigarrow \frac{\vec{u}^*}{\|\vec{u}^*\|_{L^2}} = \text{unit vector.}$$

$\vec{u}^*$

So:

$$\max\left\{ (A\vec{u})\cdot\vec{u} \;\middle|\; \begin{array}{c} \|\vec{u}\|_{L^2} = 1 \\ \vec{u}\cdot\vec{u} = 1 \end{array} \right\}$$

maximum is attained at

$\vec{u}^*$.

Rem: If $\vec{v} \perp \vec{u}^*$, then $\|\vec{u}^* + \varepsilon\vec{v}\|$ almost 1

Class Ends

$$\|M\| < 1$$

$$(I - M)^{-1}$$

$$= I + M + M^2 + M^3 + \dots$$

$$\|M\| = \alpha$$

$$I + M + M^2 + \dots + M^k$$

$$+ \underbrace{\left(M^{k+1} + \dots\right)}_{\text{bounded by } \alpha^{k+1} \cdot \frac{1}{1-\alpha}}$$

$$\|\vec{u}^* + \varepsilon \vec{v}\|^2 \qquad = (\quad)^T(\quad)$$

$$(\vec{u}^* + \varepsilon\vec{v}) \cdot (\vec{u}^* + \varepsilon\vec{v})$$

$$= \vec{u}^* \cdot \vec{u}^* + 2\varepsilon \underbrace{\vec{u}^* \cdot \vec{v}}_{\text{if } \vec{u}^* \cdot \vec{v} = 0} + \varepsilon^2 \vec{v} \cdot \vec{v}$$

$$= 1 + 2\varepsilon \cdot 0 + \varepsilon^2 C$$

$$\frac{1}{1 + \varepsilon^2 C} = 1 - \varepsilon^2 C + O(\varepsilon^4)$$