
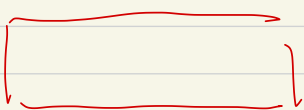


CPSC S36F

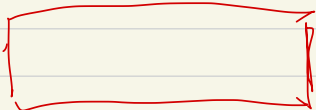
Nov 24, 2025

Idea:

picture 1   $\leftarrow \vec{x}_1 \in \mathbb{R}^d$

picture 2   $\leftarrow \vec{x}_2 \in \mathbb{R}^d$

$\vdots$   $\vdots$

picture n   $\leftarrow \vec{x}_n \in \mathbb{R}^d$

but each picture an element of  $\mathbb{R}^d$

$d \gg n$ . Maybe  $d \approx 12,000,000$

To each picture, we have maybe  $n \approx 1000$   
or 10,000

$(\vec{x}_i, y_i)$   
input  $\nearrow$   $\vec{x}_i$  feature list       $y_i$   $\nwarrow$  output  
what we hope to infer

Want: ridge regression

$$X = \text{pictures} = \underbrace{\begin{bmatrix} \vec{x}_1^T \\ \vec{x}_2^T \\ \vdots \\ \vec{x}_n^T \end{bmatrix}}_{d \approx 12,000,000}$$

1000, n

Literally:

$$(\vec{x}_1, y_1), \dots, (\vec{x}_n, y_n)$$

$$\left[ \begin{array}{l} \text{Remark: } y_i \in \mathbb{R}, \text{ but it's no problem} \\ \text{to have } \vec{y}_i \in \mathbb{R}^p, \text{ and} \\ \{ \text{cat, dog, goldfish} \} \leftrightarrow \{ \vec{e}_1, \vec{e}_2, \vec{e}_3 \} \subset \mathbb{R}^3 \end{array} \right]$$

-3-

Output, what we want to predict for  
a general  $\vec{x} \in \mathbb{R}^d$ , is: find

$\vec{w} = (w_1, \dots, w_d)$  weights, to minimize

$$\min_{\vec{w} \in \mathbb{R}^d} \left\| X \vec{w} - \vec{y} \right\|_{L^2}^2 + \lambda \|\vec{w}\|^2$$

$$X = \text{pictures} = \underbrace{\begin{bmatrix} - \vec{x}_1^T - \\ - \vec{x}_2^T - \\ \vdots \\ - \vec{x}_n^T - \end{bmatrix}}_{\substack{n \\ 1000}} \begin{bmatrix} w_1 \\ \vdots \\ w_d \end{bmatrix}$$

$d \approx 12,000,000$

roughly equal  $\begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}$

$$\lambda \geq 0$$

Normal Eqs

$$X^T X \vec{w} + \lambda \vec{w} = X^T \vec{y}$$

$$(*) \underbrace{(X^T X + \lambda I_d)}_{\text{invert a } d \times d \text{ matrix}} \vec{w} = X^T \vec{y}.$$

invert a  $d \times d$  matrix

$12 \cdot 10^6 \times 12 \cdot 10^6$  bad news...

"Dual Ridge Regression" (not really

the dual): we can solve this as

$$\text{follows: hope that } \underbrace{\vec{w}}_{d \times 1} = \underbrace{X^T}_{d \times n} \underbrace{\vec{\alpha}}_{n \times 1}$$

$$\vec{\alpha} = \begin{bmatrix} \alpha_1 \\ \vdots \\ \alpha_n \end{bmatrix} \leftarrow \text{low dim}$$

Lemma: If  $\lambda > 0$ , then

(1) there is a unique  $\vec{w}$  solving the normal equations (\*)

(2)  $\vec{w} = X^T \vec{\alpha}$  for some  $\vec{\alpha} \in \mathbb{R}^n$ ,

(3)

$$\underbrace{(X X^T + \lambda \mathbb{I}_n)}_{\text{nan matrix}} \vec{\alpha} = \vec{y}$$

Proof: "Plug-and-chug"

Also follow from duality theory

-8-

Notation has different symbols

Proof 2: Use duality:

Here:  $X, Y$  sets

$$F: X \times Y \rightarrow \mathbb{R}$$

$$\max_{x \in X} \min_{y \in Y} F(x, y) = m(F)$$

$$\leq$$

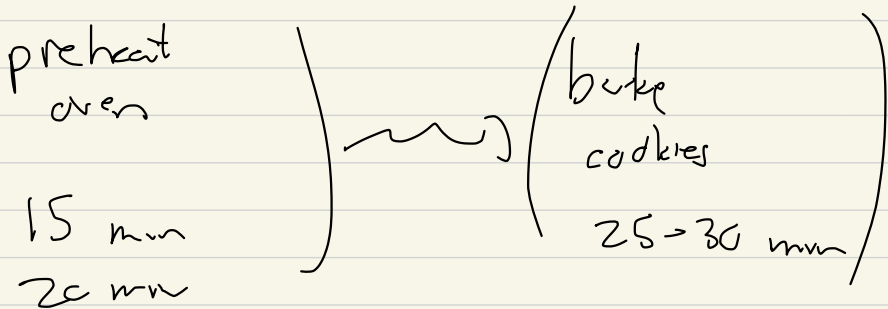
$$\min_{y \in Y} \max_{x \in X} F(x, y) = M(F)$$

If  $m(F) = M(F)$ , then under mild conditions, solving one problem solves the other.

Constrained task

e.g., Baking Cookies: .. preheat oven  
to  $350^{\circ}\text{F}$  ..

Most constraining:



In linear programming!

complementary slackness

“Plug and chug”

Lemma: (1) If  $\lambda > 0$ , then

$$(X^T X + \lambda I_d)$$

and

$$(X X^T + \lambda I_n)$$

are positive definite (symmetric) matrices

(2) therefore both have strictly positive eigenvalues

(3) therefore both are invertible



- 9 -

Warning!

$$\begin{bmatrix} 1 \\ 2 \end{bmatrix} \begin{bmatrix} 1 & 2 \end{bmatrix} = \begin{bmatrix} 1 & 2 \\ 2 & 4 \end{bmatrix} \text{ eigs:}$$

$$\begin{bmatrix} 1 & 2 \\ 2 & 4 \end{bmatrix} \begin{pmatrix} 1 \\ 2 \end{pmatrix} = \begin{bmatrix} 5 \\ 10 \end{bmatrix} = 5 \begin{pmatrix} 1 \\ 2 \end{pmatrix}$$

$$\lambda_2 = 0$$

$$\lambda_1$$

$$\begin{bmatrix} 1 & 2 \\ 2 & 4 \end{bmatrix}^{-1} \text{ doesn't exist}$$

$$\left( \begin{bmatrix} 1 & 2 \\ 2 & 4 \end{bmatrix} + 10^{-100} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \right)^{-1} \text{ exists}$$

So if  $A$  symmetric,  
 pos semi-def, then  
 for any  $\lambda > 0$ ,  $(A + \lambda I)^{-1}$   
 exists, but the "conditioning" or  
 "numerical stability" can be really  
 bad if  $\lambda$  not large enough...

Proof of Lemma page 5:

$XX^T + \lambda I_n$  is invertible so

let

$$\vec{\alpha} = (XX^T + \lambda I_n)^{-1} \vec{y}.$$

Then

$$(XX^T + \lambda I_n) \vec{\alpha} = \vec{y}$$

$$XX^T \vec{\alpha} + \lambda \vec{\alpha} = \vec{y}$$

mult by  $X^T$  on left

$$X^T (X X^T \vec{\alpha} + \lambda \vec{\alpha}) = X^T y$$

$$X^T X X^T \vec{\alpha} + \lambda X^T \vec{\alpha} = X^T y$$

$$X^T X \hat{w} + \lambda \hat{w} = X^T y$$

$$\hat{w} = X^T \vec{\alpha}$$

$$(X^T X + \lambda I_d) \hat{w} = X^T y$$

□

Next time: What is going on?