CPSC 536F Dec 3, 2025

- Ridge Regression Dual Form: given X, y, $\lambda > 0$

$$\min_{\vec{w} \in \mathbb{R}^d} \|Xw - y\|^2 + \lambda \|w\|^2$$

Gives $(X^T X + \lambda I_d) w = X^T y$. Bad if $d \gg n$

Equiv:

$$\min_{(\vec{w}, \vec{z}) \in \mathbb{R}^d \times \mathbb{R}^n} f(w, z) = \|z\|^2 + \lambda \|w\|^2$$

$$z = Xw - y$$

Lagrangian: $g(w, z) = Xw - y - z = 0$

Etc. min-max vs max-min

Conceptualize (Abstract nonsense):

Thm (The Representer Theorem, Simple Form):

H could be $\mathbb{R}^d$, d large
Hilbert Space

Let $\Phi : \mathcal{X} \to H$ be a set theoretic map, and consider data points

$$(x_1, y_1), \ldots, (x_n, y_n) \in \mathcal{X} \times \mathbb{R}$$

Let $\lambda > 0$, and consider $g : H \to \mathbb{R}$

$$g(w) = \lambda \|w\|_H^2 + \sum_{i=1}^{n} \left(\langle \Phi(x_i), w \rangle_H - y_i \right)^2$$

Then minimum at w_0 satisfies

$$(X^* X + \lambda I_H) w_0 = X^* y$$

where $X : H \to \mathbb{R}^n$ given by

$$X(w) = \begin{bmatrix} \langle \Phi(x_1), w \rangle_H \\ \vdots \\ \langle \Phi(x_n), w \rangle_H \end{bmatrix}.$$

One remarkable fact ...

So $\mathbb{R}^d \rightsquigarrow H$ inner product space

($\dim(H) = \infty$, we want a Hilbert space)

$$X^T \rightsquigarrow X^*$$

$$X : \mathbb{R}^d \to \mathbb{R}^n \rightsquigarrow X : H \to \mathbb{R}^n$$

$$\vec{y} = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} \qquad X = \begin{bmatrix} - & x_1^T & - \\ & \vdots & \\ - & x_n^T & - \end{bmatrix}$$

w best coefficients, $w \in \mathbb{R}^d$, $d \gg n$

=

$$\min f(\vec{x})$$

s.t. $\vec{g}(\vec{x}) = \vec{0}$

at optimum (min/max/saddle)

$\nabla f(\vec{x}) = 0$ if no g's

$\nabla f(\vec{x})$ in span $\nabla \vec{g}$, i.e.

$\nabla g_1, \nabla g_2, \ldots, \nabla g_a$

$\nabla f(\vec{x}) + \alpha^T \cdot \nabla \vec{g} = 0$ some $\vec{\alpha}$

$$L(x, \alpha) = f(x) + \alpha \cdot \vec{g}(x)$$

For any $\vec{x}$

$$\max_{\alpha \in \mathbb{R}^a} f(\vec{x}) + \alpha \cdot \vec{g}(\vec{x})$$

$$= +\infty \quad \text{unless} \quad \vec{g}(\vec{x}) = \vec{0}$$

$$\min_{\vec{x}} \max_{\vec{\alpha}} L(\vec{x}, \vec{\alpha}) = \min_{\vec{x} \text{ s.t. } \vec{g}(\vec{x}) = \vec{0}} f(\vec{x})$$

Now consider

$$\max_{\vec{\alpha}} \min_{\vec{x}} L(\vec{x}, \vec{\alpha})$$

if these two are equal, then...

$$\min_{\vec{w} \in \mathbb{R}^d} \overbrace{\lambda \|w\|^2 + \|X\vec{w} - \vec{y}\|^2}^{f(w)\ ☹}$$

rewrite ! Key step (ad hoc) to get a good dual problem

$$\min_{(\vec{w},\vec{z}) \in \mathbb{R}^d \times \mathbb{R}^n} \underbrace{\lambda \|w\|^2 + \|z\|^2}_{f(w,z)\ ☺}$$

s.t. $\vec{z} = X\vec{w} - \vec{y}$, i.e.

$$\vec{g}(\vec{w},\vec{z}) = 0, \text{ where}$$

$$\vec{g}(\vec{w},\vec{z}) = X\vec{w} - \vec{y} - \vec{z}$$

$$\max_{\vec{\alpha}} \left(\min_{(w,z)} \underbrace{L((w,z);\alpha)}_{\|z\|^2 + \lambda\|w\|^2 + \vec{\alpha}\cdot\vec{g}(\vec{w},\vec{z})} \right)$$

Fix $\vec{\alpha} \in \mathbb{R}^a$

Which (w,z) pairs minimize

$L((w,z);\alpha) =$

$$\underbrace{\vec{z}\cdot\vec{z}}_{1} + \underbrace{\lambda\,\vec{w}\cdot\vec{w}}_{2} + \underbrace{\vec{\alpha}\left(X\vec{w}-\vec{y}-\vec{z}\right)}_{3}$$

$\vec{\alpha}$ fixed, $\|\vec{w}\|, \|\vec{z}\| \to \infty$

$L \to \infty$

So ... say min is at $\omega = \vec{\omega}_0$, $z = \vec{z}_0$

Vary: fix $\vec{z}_1$, $\vec{\omega}_1$

Consider

$$\vec{\omega}(\varepsilon) = \vec{\omega}_0 + \varepsilon \vec{\omega}_1$$

$$\vec{z}(\varepsilon) = \vec{z}_0 + \varepsilon \vec{z}_1$$

---

(1) $\vec{z}(\varepsilon) \cdot \vec{z}(\varepsilon)$

$$= \left(\vec{z}_0 + \varepsilon \vec{z}_1\right) \cdot \left(\vec{z}_0 + \varepsilon \vec{z}_1\right)$$

$$= \vec{z}_0 \cdot \vec{z}_0 + \left(2 \vec{z}_0 \cdot \vec{z}_1\right) \varepsilon + O\left(\varepsilon^2\right)$$

ε term is $2 \vec{z}_0 \cdot \vec{z}_1$

(2) ε term is $2 \vec{\omega}_0 \cdot \vec{\omega}_1$

③ $\vec{\alpha}\left(X(\vec{w}_0+\varepsilon\vec{w}_1)-\vec{y}-(\vec{z}_0+\varepsilon\vec{z}_1)\right)$

ε term $\quad \vec{\alpha}\left(X\vec{w}_1-\vec{z}_1\right)$

So at min z_0, w_0; $\forall\ \vec{z}_1, \vec{w}_1$;

$$2\vec{z}_0\circ\vec{z}_1+2\lambda\,\vec{w}_0\cdot\vec{w}_1+\vec{\alpha}\left(X\vec{w}_1-\vec{z}_1\right)=\vec{0}$$

$$=\left(2\vec{z}_0-\vec{\alpha}\right)\cdot\vec{z}_1+\left(2\lambda\vec{w}_0+X^T\vec{\alpha}\right)\vec{w}_1=\vec{0}$$

So $=\vec{0}\quad\forall\ \vec{z}_1,\ \forall\vec{w}_1$

$$\begin{cases}2\vec{z}_0-\vec{\alpha}=0\\ 2\lambda\vec{w}_0+X^T\vec{\alpha}=0\end{cases}$$

$$\begin{cases}\vec{z}_0=\vec{\alpha}/2 & \text{(less important)}\\ \vec{w}_0=\left(\frac{-1}{2\lambda}\right)X^T\vec{\alpha}\end{cases}$$

☺

$\Rightarrow$ $\vec{\omega}_0$ is proportional to $X^T \vec{\alpha}$

Plug & chug!

We guess $\vec{w}^*$ at min is

$$X^T \vec{\alpha}_{\text{plug \& chug}}$$

Duality!

$\vec{w}^*$ is $\left(\frac{-1}{2\lambda}\right) X^T \vec{\alpha}_{\text{dual}}$

$$= X^T \left(\left(\frac{-1}{2\lambda}\right) \vec{\alpha}_{\text{dual}} \right)$$