127

- L1 norm
  - min-norm
  - least-squares
  - ridge

① min ℓ_1 norm

want to transform into LP

$$\min \|\vec{x}\|_1 \quad \text{s.t. } A\vec{x} = \vec{b}$$

→ convex objective

→ constraints linear

→ non-differentiable

↳ A wide, full row rank (∞ solns)

$$\sum_{i=1}^{n} |x_i| = \|\vec{x}\|_1$$

$$x_i = x_i^+ - x_i^-$$

$$|x_i| = x_i^+ - x_i^-$$

$$x_i^+ = \begin{cases} x_i & \text{if } x_i > 0 \\ 0 & \text{o/w} \end{cases} \qquad x_i^- = \begin{cases} -x_i & \text{if } x_i < 0 \\ 0 & \text{o/w} \end{cases}$$

$$x_i^+, x_i^- \geq 0$$

Rewrite Program:

$$\min \sum_{i=1}^{n} x_i^+ + \sum_{i=1}^{n} x_i^-$$

$$A(\vec{x}^+ - \vec{x}^-) = \vec{b}$$

$$x_i^+ \geq 0$$

$$x_i^- \geq 0$$

Claim: This new program will always choose only one of x_i^+ or x_i^- nonzero.

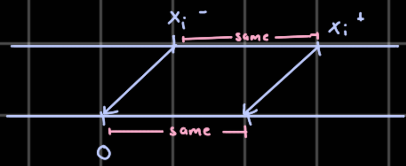
Suppose $x_i^+ > 0$ & $x_i^- > 0$

WLOG: $x_i^+ > x_i^-$

Consider $x_{i\,(new)}^+ = x_i^+ - x_i^-$

& $x_{i\,(new)}^- = 0$

Notice: constraint on gap between x_i^+ & x_i^-, not actual values



$x_i^+(\text{new}) + x_i^-(\text{new}) = x_i^+ - x_i^- + 0$

$< x_i^+ + x_i^-$

↳ contradiction

② Parallel to least-squares

$$\min_{\vec{x}} \|A\vec{x} - \vec{b}\|_1$$

↳ $\vec{e} \doteq A\vec{x} - \vec{b}$

$$\min \|\vec{e}\|_1$$
$$\text{s.t. } A\vec{x} - \vec{e} = \vec{b}$$

↳ follow same strat as min-norm to get LP

3) Points $\vec{b}_1, \ldots, \vec{b}_m$

Consider $\min \sum_{i=1}^{m} (\vec{x} - \vec{b}_i)^2$

↳ optimum $\vec{x}_* = \dfrac{\sum \vec{b}_i}{m}$

ℓ_1 min-norm parallel:

$$\min_{\vec{x}} \sum_{i=1}^{n} |\vec{x} - \vec{b}_i|$$

↳ scalar case: $\vec{x}_* \in \mathbb{R}$ $\quad b_i \in \mathbb{R}$ $\quad 1 \le i \le m$

$$\min_{x} \sum_{i=1}^{m} |x - b_i|$$

eg: $b_1, b_2, \ldots, b_7$ ordered ie:

$x > b_1$, 2

$\frac{d}{dx}|x-b_1| = 1$

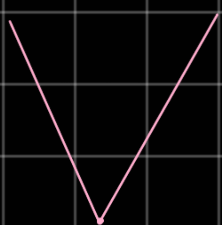
$1 + (-6) = -5$

↳ $\frac{d}{dx}\left(\sum_{i=1}^{n} |x - b_i|\right) = \begin{cases} -7 & x < b_1 \\ -5 & b_1 < x < b_2 \\ \vdots & \vdots \\ 5 & b_6 < x < b_7 \\ 7 & b_7 < x \\ \text{DNE} & \end{cases}$

$|x - b_i| = \begin{cases} x - b_i & \text{if } x > b_i \\ b_i - x & \text{if } x \le b_i \end{cases}$

non-diffable, but not everywhere

$\frac{d}{dx}|x - b_i| = \begin{cases} 1 & \text{if } x > b_i \\ -1 & \text{if } x < b_i \\ \text{undefined/DNE} & \text{if } x = b_i \end{cases}$



non-diff'able here

Critical points: $\frac{d}{dx} = 0$, $\frac{d}{dx} =$ undefined, boundaries

$$\sum_{i=1}^{7} |x - b_i| = |x - b_4| + \underbrace{\sum_{i \ne 4} |x - b_i|}_{\forall x \in (b_3, b_5)}$$

sum $|x-b_3|+|x-b_5|$ = const

↳ same for b_2, b_6 (ie, you can pair these up s.t. they cancel out)

so then we get that the minimum is the median ("robust compared to the mean")

↳ minimum $= |x-b_4|$

Note: this pairing up doesn't work in the vector case, only the scalar case

LASSO

$$\min_{\vec{x}} \|A\vec{x}-\vec{b}\|_2 + \lambda\|\vec{x}\|_1$$

→ penalization when $\vec{x}$ far from 0.

↳ Exploring simple scalar case: sparsity:

$$\min_x \frac{1}{2}\sum_{i=1}^{n}(a_i x - b_i)^2 + \lambda\|x\|_1$$

Recall, least-sq.

$$\min_x \frac{1}{2}\sum_{i=1}^{n}(a_i x - b_i)^2$$

$$\vec{x}_* = \frac{\sum a_i b_i}{\sum a_i^2}$$

$$\vec{x}_* = (A^TA)^{-1}A^T\vec{b}$$

$$\frac{d}{dx}f(x) = \begin{cases} \sum(a_i x - b_i)a_i + \lambda & \text{if } x>0 \\ \sum(a_i x - b_i)a_i - \lambda & \text{if } x<0 \\ \text{DNE} & \text{if } x=0 \end{cases}$$

critical points: $x=0$ (non-diff'able there)

set derivative $= 0$:

if $x>0$, then $\sum(a_i^2 x - a_i b_i) + \lambda = 0$

$$x\sum a_i^2 - \sum a_i b_i + \lambda = 0$$

$$x = \frac{\sum a_i b_i - \lambda}{\sum a_i^2} \quad \text{if } \sum a_i b_i > \lambda$$

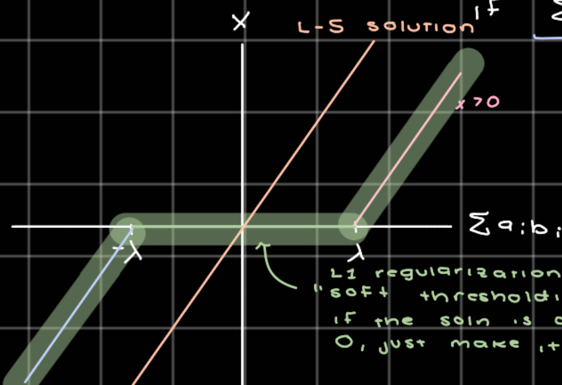
↳ check: is cond satisfied? (ie is $x>0$?)

↳ denom is $\sum$ of ^{2}'s ⇒ will always be non-neg → when will the inequality be strict? ie $x>0$ and NOT $x \geq 0$?

if $\sum a_i b_i > \lambda$

if $x<0$, then $x = \dfrac{\sum a_i b_i + \lambda}{\sum a_i^2}$ if $\sum a_i b_i < -\lambda$

↳ similarly reason



L1 regularization is doing "soft thresholding" - ie, if the soln is close to 0, just make it 0