EECS 127

April 2, 2020

- Projects: next week
- P/NP or S/U.

The L1 norm.

Least-squares: Minimizes L2 norm of error.

Min-norm: ~~Min norm,~~ minimizes L2 norm of $\|\vec{x}\|_2$

What does minimizing the L1 norm give us?

① Min L1-norm

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

- Objective is convex.
- Constraints are linear.
- Non-differentiable objective.

A: wide

Assume Full-row rank.

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

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

Transformation:

$$|x_i| = x_i^+ + x_i^-$$
$$x_i^+ = x_i \text{ if } x_i > 0, \quad 0 \text{ otherwise.}$$
$$x_i^- = -x_i \text{ if } x_i < 0, \quad 0 \text{ otherwise.}$$

Note: $x_i^+ \geq 0, \ x_i^- \geq 0$.

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

$\therefore$ We can write the program as:

$$\min \sum_{i=1}^{n} x_i^+ + \sum_{i=1}^{n} x_i^-$$
$$\text{s.t. } A\vec{x}^+ - A\vec{x}^- = \vec{b}$$
$$\vec{x}^+ \geq 0, \quad \vec{x}^- \geq 0.$$

Think of:

$$\vec{y} = \begin{bmatrix} \vec{x}^{+} \\ +\vec{x}^{-} \end{bmatrix} \qquad [A \mid A] \begin{bmatrix} \vec{x}^{+} \\ +\vec{x}^{-} \end{bmatrix} = \vec{b}$$

Then:

$$\boxed{\begin{array}{l} \min\limits_{\vec{y}} \; [1 \; 1 \; 1 \cdots 1]\, \vec{y} \\ [A \mid -A] \begin{bmatrix} \vec{y} \end{bmatrix} = \vec{b} \end{array}}$$

An LP.

By writing $x_i = x_i^{+} - x_i^{-}$, we are adding some "freedom", since we are not including the conditional constraints in the program.

But the minimization will always choose values such that only one of x_i^{+}, x_i^{-} is non zero. Why?

Suppose we have $x_i^{+} > 0$ and $x_i^{-} > 0$ at optimal.

Consider the solution. $x_i^{+} - \varepsilon$, $x_i^{-} + \varepsilon$

It also satisfies the constraints

WLG say. $x_i^{+} > x_i^{-} > 0$ at optimum.

then consider $x_i^{+} - x_i^{-} = x_{i(new)}^{+}$

$x_{i(new)}^{-} = 0.$

Still satisfy constraints, but objective. $(x_{i(new)}^{+} + x_{i(new)}^{-})$ is strictly smaller.

So this change of variables does not affect optimum value.

This trick can be used to convert 1-norms from objective & into LPs!

2 Now consider the parallel to Least squares. i.e.

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

So this least squares parallel is also an LP in the same way as before.

3 Median as a robust version of the "average":
For a bunch of scalars if we want to find their average we write this as:

Points: $x_1, x_2 \dots x_n$.

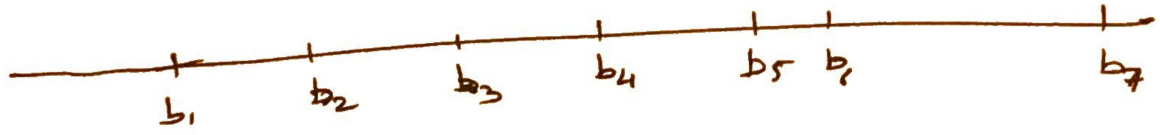
$$\min_{x} \sum_{i=1}^{n} (x - x_i)^2 \rightarrow \text{optimal } x^* \text{ is } \frac{\sum_{i=1}^{n} x_i}{n} \text{ (mean.}$$

. Avg income in room.

What if we consider: $\min_{x} \sum_{i=1}^{n} |x - b_i|$ (same as x_i above, just change of notation.)

Points $b_1, b_2 \dots, b_n$

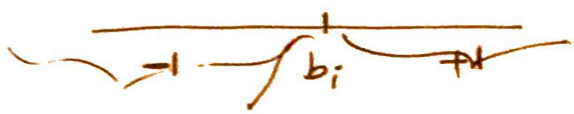
Say these are ordered.



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

Non-differentiable points: $b_1, b_2 \ldots b_7$.

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



So now:

$$\frac{d}{dx}\left(\sum_{i=1}^{n} |x - b_i|\right)$$

$= -7$ if $x < b_1$

$= 1 - 6 = -5$ $\quad b_1 < x < b_2$

$= -3$ $\quad b_2 < x < b_3$

b_4 is a good candidate for a local minimum.

So $\displaystyle\sum_{i=1}^{7}|x-b_i| = |x-b_4| + \underbrace{\sum_{\substack{i=1\\ i\neq 4}}^{7}|x-b_i|}_{\substack{\text{the same for all}\\ x\in(b_3,b_5)}}$

~~whatever you add one~~

whatever you add on one side, you remove from the other.

So $x=b_4$ must be minimizer.

$b_4=$ median!

So minimizing sum of L1-norms gives median.

Median is robust ... b_7 can go as far out as it wants

3

What about the parallel to ridge regression?

Can we regularize with a 1-norm?

$$\min_{\vec{x}} \quad \|A\vec{x}-\vec{b}\|_2^2 + \lambda\|\vec{x}\|_1$$

<u>LASSO.</u>

Can also be written as:

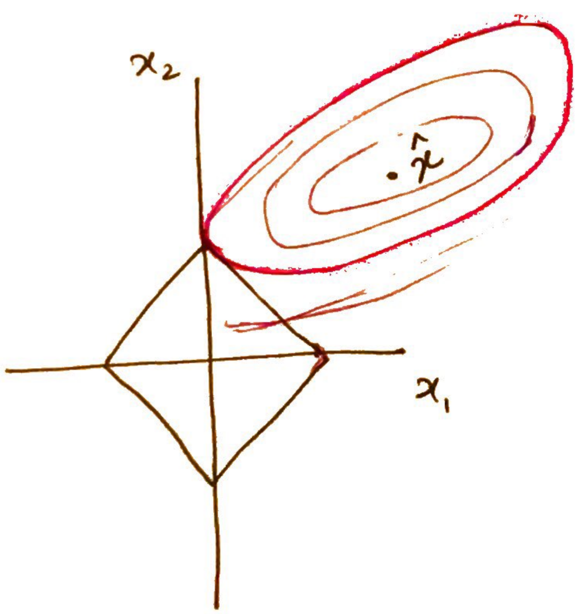
$$\min \quad \|A\vec{x}-\vec{b}\|_2^2$$

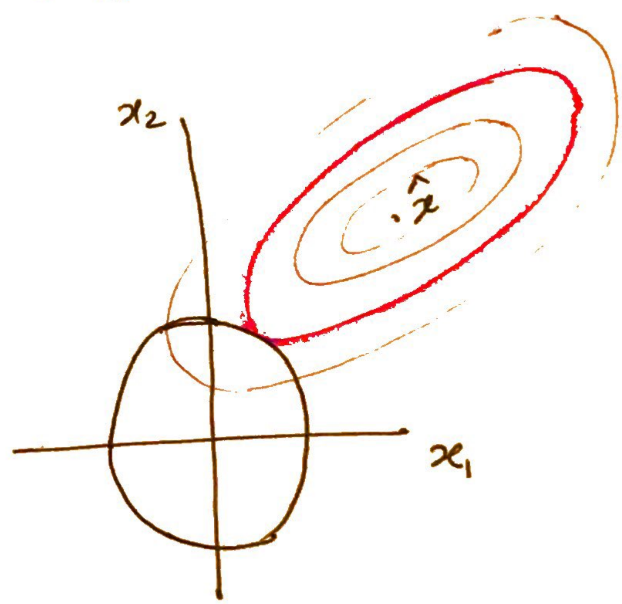
$$\text{s.t.} \ \|\vec{x}\|_1 \le t.$$

Compared to Ridge.

$$\min \quad \|A\vec{x}-\vec{b}\|_2^2$$

$$\text{s.t.} \ \|\vec{x}\|_2 \le t$$





Let $\hat{x} = \underset{x}{\operatorname{argmin}} \ \|A\vec{x}-\vec{b}\|_2^2$ with no constraints.

$$\|A\vec{x}-\vec{b}\|_2^2 = l$$

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

Quadratic $A^TA \cdot PSD \Rightarrow$ Level sets are an ellipse.

So when the regulaiized is L1, you will meet at a corner. Corners are on the axis, so some coefs will be set to 0.

This is why LASSO is said to give a "sparse" solution