EECS 127 Lecture 8.

Feb 13, 2020

Admin:

- HW due tomorrow!

Today:
- Noise
- Sensitivity / Perturbation Analysis
- Ridge regression.

- $A\vec{x} = \vec{y}$ $\quad A \in \mathbb{R}^{n \times n}$, invertible.

If $\vec{y} \to \vec{y} + \vec{\delta y}$, and because of this $\vec{x} \to \vec{x} + \vec{\delta x}$

how big is $\vec{\delta x}$?

$\frac{\|\vec{\delta x}\|_2}{\|\vec{x}\|_2}$: want to understand "relative change"

$$A(\vec{x} + \vec{\delta x}) = (\vec{y} + \vec{\delta y})$$

$$\Rightarrow A\vec{\delta x} = \vec{\delta y} \quad \Rightarrow \quad \vec{\delta x} = A^{-1} \cdot \vec{\delta y}$$

$$\|\vec{\delta x}\|_2 = \|A^{-1} \cdot \vec{\delta y}\|_2 \leq \|A^{-1}\|_2 \cdot \|\vec{\delta y}\|_2 \quad (\Delta)$$

by definition of spectral norm.

$$\|A\|_2 = \max_{\|\vec{y}\|} \frac{\|A\vec{y}\|_2}{\|\vec{y}\|_2}$$

How to bound norm $\|\vec{x}\|_2$?

$$A\vec{x} = \vec{y}.$$

$$\|\vec{y}\|_2 = \|A\vec{x}\|_2 \leq \|A\|_2 \cdot \|\vec{x}\|_2.$$

$$\frac{\|\delta \vec{x}\|_2}{\|\vec{x}\|_2} \overset{(\Delta)}{\leq} \frac{\|A^{-1}\|_2 \cdot \|\delta \vec{y}\|_2}{\|\vec{x}\|_2} \leq \frac{\|A^{-1}\|_2 \|\delta \vec{y}\|_2}{\|\vec{y}\|_2 \cdot \frac{1}{\|A\|_2}} = \|A\|_2 \|A^{-1}\|_2 \left[\frac{\|\delta \vec{y}\|_2}{\|\vec{y}\|_2}\right]$$

$$\|A\|_2 = \sigma_{max}$$

$$\|A^{-1}\|_2 = \frac{1}{\sigma_{min}}$$

$$\|A\|_2 \cdot \|A^{-1}\|_2 = \boxed{\frac{\sigma_{max}}{\sigma_{min}}}$$

Condition Number of A.

Least-Squares: $\vec{x} = (A^T A)^{-1} A^T \vec{b}$

Normal eqn's: $(A^T A)\vec{x} = A^T \cdot \vec{b}$

$$K(A^T A) = \frac{\sigma_{max}(A^T A)}{\sigma_{min}(A^T A)}$$

Consider: $\vec{x} = (A^T A + \lambda I)^{-1} A^T b$.

Introducing Ridge Regression.

- Shift property of eigenvalues.

A is a matrix. $\vec{v}$ is an e-vector, e-val. [illegible] μ

$(A + \lambda I)\vec{v} = A\vec{v} + \lambda \vec{v} =$ [illegible] $\mu \cdot \vec{v} + \lambda \vec{v} = (\mu + \lambda) \cdot \vec{v}$

Regularization

Optimization problem:

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

Penalty term = Regularization.

$$f(x) = \|A\vec{x} - \vec{b}\|_2^2 + \lambda^2 \|\vec{x}\|_2^2$$

$$= (A\vec{x} - \vec{b})^T (A\vec{x} - \vec{b}) + \lambda^2 \vec{x}^T \vec{x}$$

f(x) [illegible] $= (A\vec{x} - \vec{b})^T (A\vec{x} - \vec{b}) + \lambda^2 \cdot \vec{x}^T \vec{x}$.

$$= \vec{x}^T A^T A \cdot \vec{x} + \vec{b}^T \vec{b} - \vec{x}^T A^T \vec{b} - 2\vec{b}^T A \vec{x} + \lambda^2 \cdot \vec{x}^T \vec{x}.$$

$$\nabla f(\vec{x}) = 2A^T A \vec{x} - 2(\vec{b}^T A)^T + 2\lambda^2 \vec{x}$$

Set $= 0$.

$\nabla \vec{x}^T A^T A \vec{x} = 2A^T A \cdot \vec{x}$

$\nabla \vec{c}^T \vec{x} = \vec{c}$

- Set to 0.

$$(A^T A + \lambda^2 I_{n \times n}) \cdot \vec{x} = A^T \vec{b}$$

$$\vec{x} = (A^T A + \lambda^2 I)^{-1} \cdot A^T \vec{b}.$$

Another interpretation:

" $\vec{x}$ is not too large."

$$\vec{x} \approx \vec{0}$$

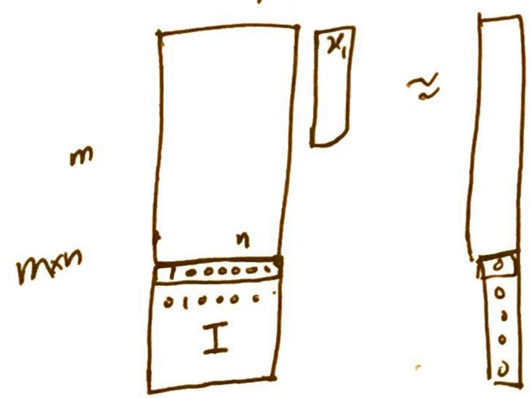
$$10^{10} \vec{x} \approx \vec{0}$$

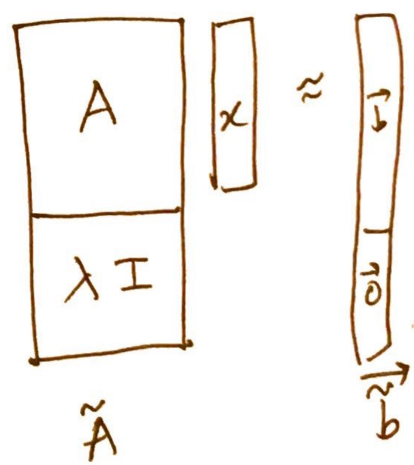
$$10 \vec{x} \approx \vec{0}$$

$$0.1 \vec{x} \approx \vec{0}$$

Least-Squares

$$A \vec{x} \approx \vec{b}$$

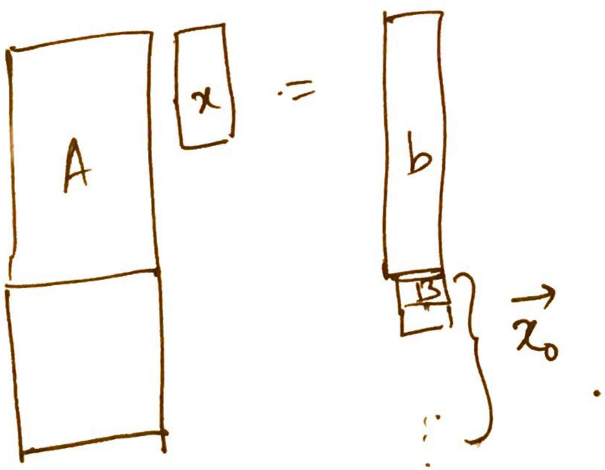




LS: $\vec{x}_{LS} = (\tilde{A}^T \tilde{A})^{-1} \tilde{A}^T \vec{b}$

$$= \left(\begin{bmatrix} A^T & \lambda I \end{bmatrix} \begin{bmatrix} A \\ \lambda I \end{bmatrix} \right)^{-1} \begin{bmatrix} A^T & \lambda I \end{bmatrix} \begin{bmatrix} \vec{b} \\ \vec{0} \end{bmatrix}$$

$$= (A^T A + \lambda^2 I)^{-1} A^T \vec{b}$$



Tikhonov Regularization.

$$\min_{\vec{x}} \| W_1 (A\vec{x} - \vec{b}) \|_2^2 + \| W_2 (\vec{x} - \vec{x}_0) \|_2^2$$

$$\begin{bmatrix} W_1 A \\ W_2 \end{bmatrix} \vec{x} = \begin{bmatrix} W_1 \vec{y} \\ W_2 \vec{x}_0 \end{bmatrix}$$