Recitation #10

Irina Espejo (iem244@nyu.edu)

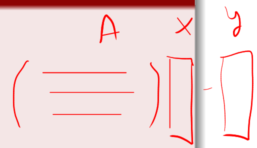
Center for Data Science

DS-GA 1014: Optimization and Computational Linear Algebra for Data Science

recall: least squares

Least squares

The least squares problems can be written as



$$\min_{x \in \mathbb{R}^n} \|Ax - y\|^2$$

where $\underline{A \in \mathbb{R}^{d \times n}}$. Using minimization of convex functions we find that

$$x \text{ is solution} \leftrightarrow A^T Ax = A^T y$$

But... A does not have an inverse!

Moore-Penrose pseudo-inverse

A^{-1}

Let $A \in \mathbb{R}^{d \times n}$ decompose as $A = U\Sigma V^T$, then $A^\dagger = V\Sigma' U^T \in \mathbb{R}^{d \times n}$ is the Moore-Penrose pseudo-inverse of A where

dagger

$$\Sigma'_{ii} = \begin{cases} \frac{1}{\Sigma_{ii}} & \text{if } \Sigma_{ii} \neq 0 \\ 0 & \text{otherwise} \end{cases}, \Sigma'_{ij} = 0 \text{ when } i \neq j$$

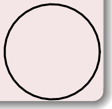
Unregularized least squares

Theorem: The set of solutions of the minimization problem $\min_{x\in\mathbb{R}^d}\|Ax-y\|^2$ is $A^\dagger y + ker(A)$

Ridge regularization

Theorem: for any $\lambda > 0$ the solution of the minimization problem $\min_{x\in\mathbb{R}^d}\|Ax-y\|^2 + \lambda\|x\|^2$ is

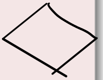
$$x_{\text{ridge}} = (A^TA + \lambda Id)^{-1}A^Ty$$



Lasso regularization

Theorem: for any $\lambda > 0$ the solution of the minimization problem $\min_{x\in\mathbb{R}^d}\|Ax-y\|^2 + \sum_j \lambda|x_j|$ is

$$x_{\text{lasso}} = \text{argmin}_{x\in\mathbb{R}^d}\|Ax-y\|^2 + \lambda|x|$$



practice: ridge regression

Exercise 1

Show that the solution x_{ridge} is given by the formula in the previous slide

$$x_{\text{ridge}} = (A^T A + \lambda Id)^{-1} A^T y$$

$$f(x) = \|Ax - y\|^2 + \lambda \|x\|^2$$

$$\nabla_x f(x) = 2\left(A^T(Ax - y)\right) + 2\lambda x$$

$$\nabla_x f = 0$$

$$A^T(Ax - y) + \lambda x = 0$$

$$(A^T A + \lambda Id)x = A^T y \rightarrow \text{solution} \quad \square$$

Exercise 1

Exercise 1

Exercise 2

Let $x_0 \in \mathbb{R}$ and $f_{x_0} : \mathbb{R} \to \mathbb{R}$ be defined as

fixed

$$f_{x_0}(x) = \frac{1}{2}x^2 - x_0 x + \lambda|x|$$

fixed

Show that for $\lambda \geq 0$, the function f_{x_0} admits a unique minimizer goven by $x^* = \eta(x_0; \lambda)$ where η is the *soft-thresholding function*:

$$\eta(x_0;\lambda) = \begin{cases} x_0 - \lambda & \text{if } x_0 \geq \lambda \\ 0 & \text{if } -\lambda \leq x_0 \leq \lambda \cancel{x_0 + \lambda} \\ \text{if } x_0 \leq -\lambda & \text{if } x_0 < -\lambda \end{cases}$$

$x_0 \le |\lambda|$

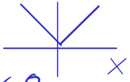
$\frac{x}{|x|}$

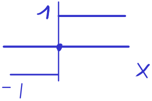
sign(x)

$$\frac{\partial f}{\partial x} = x - x_0 + \lambda\, \Theta(x)$$

Heavyside function

$$\Theta(x) = \begin{cases} -1 & \text{if } x < 0 \\ 0 & \text{if } x = 0 \\ 1 & \text{if } x > 0 \end{cases}$$





Exercise 2

$\frac{\partial f}{\partial x} = 0$ $\qquad x - x_0 + \lambda \theta(x) = 0$

$$\boxed{x + \lambda \theta(x) = x_0}$$

- If $x^* > 0$, $\leftrightarrow$ $\theta(x^*) = 1$

$$x^* = x_0 - \lambda > 0$$

$$x_0 > \lambda$$

If $x_0 > \lambda$ then $x^* = x_0 - \lambda$

- If $\theta(x^*) = -1 \rightarrow x^* < 0$

$x_0 < -\lambda$

$$x^* - \lambda = x_0$$

$$x^* = x_0 + \lambda < 0$$

Exercise 2

- $\theta(x^*) = 0 \rightarrow x^* \leq 0$ $\boxed{\theta(x) = 0}$

  $x^* + \cancel{\lambda \cdot 0} = x_0 \leq 0$

otherwise

Exercise 2

Exercise 3

Let $A \in \mathbb{R}^{n\times d}$ be a matrix such that its columns are orthonormal. Show that the Lasso solution $x_{\text{lasso}} = \text{argmin}_{x\in\mathbb{R}^d} \|Ax - y\|^2 + \sum_j \lambda|x_j|$ satisfies

$$x_{j\,\text{lasso}} = \eta(x_{j\,LS};\lambda), \text{for all } j \in \{1,..,d\}$$

"x_0"

where $x_{LS} = A^T y$ "x_0"

$A : A^T A = I_d$

$$\boxed{\nabla f = 2(A^T(Ax-y)) + \lambda \vec{\theta}(x)}$$

$$\vec{\theta}(x) = \begin{pmatrix} \\ \hline \\ \end{pmatrix} \rightarrow \theta_j(x_j) = \begin{cases} -1 & \text{if } x_j < 0 \\ 0 & \text{if } x_j = 0 \\ 1 & \text{if } x_j > 0 \end{cases}$$

Exercise 3

$$\cancel{2A^TA}^{\,Id}\, x + \lambda \vec{\theta}(x) = 2A^T y$$

$x^* = \begin{pmatrix} - \\ - \\ - \\ - \\ - \end{pmatrix} \to$ Let's find x_j^*

j: $\quad \cancel{2}\, x_j + \frac{\lambda}{2}\theta_j(x_j) = \cancel{2}(A^Ty)_j$

Solve for $x_j \to x_j^*$

Ex 2: $\quad x + \lambda\theta(x) = x_0$

new "λ" $\quad \lambda/2$

new "x_0" $\quad (A^Ty)_j$

$$\left.\begin{matrix} \\ \\ \end{matrix}\right\} \quad x_j^* = \eta\left((A^Ty)_j \,;\, \lambda/2\right)$$

Exercise 3

Vectorial solution

$$x^* = \eta(A^T y; 1/2) \quad \square$$

Exercise 3

practice: pseudo-inverse

Exercise 4

Show that the Moore-Penrose pseudo-inverse $A^\dagger \in \mathbb{R}^{d\times n}$ is the only matrix such that

1. $AA^\dagger A = A$
2. $A^\dagger A A^\dagger = A^\dagger$
3. $AA^\dagger \in \mathbb{R}^{d\times n}$ are symmetric matrices

We will prove this in two steps:

- 4.1: Show that the Moore-Penrose pseudo-inverse fulfils properties 1, 2 and 3.
- 4.2: Now show that it is unique.

Exercise 4.1

(1) $A = U \Sigma V^T$ recall that U, V are orthogonal

$A^+ = V \Sigma' U^T$

$$AA^+A = (U\Sigma \cancel{V^T})(\cancel{V}\Sigma' \cancel{U^T})(\cancel{U}\Sigma V^T)$$

$$= U \Sigma \Sigma' \Sigma V^T = U \Sigma V^T = A$$

(3) $AA^+ = U\Sigma \cancel{V^T}\cancel{V}\Sigma' U^T = U\Sigma\Sigma' U^T =$

$$(AA^+)^T = (U\Sigma\Sigma' U^T)^T = (\Sigma' U^T)^T (U\Sigma)^T = U\Sigma'\Sigma U^T$$

is symmetric

Exercise 4.1

2

$$
\begin{aligned}
A^+ A A^+ &= (V \Sigma' \cancel{U^T})(\cancel{U} \Sigma \cancel{V^T})(\cancel{V} \Sigma' U^T) \\
&= V \Sigma' \Sigma \Sigma' U^T \\
&= V \Sigma' U^T \\
&= A^+
\end{aligned}
$$

Exercise 4.2

4.2.2 By definition of A^+ and SVD is unique

4.2.1 Clear from 1, 2, 3