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 $A\in\mathbb{R}^{d\times n}$. 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

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

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

recall: least squares

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

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

Lasso regularization

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

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

practice: ridge regression

Exercise 1

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

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

Exercise 1

Exercise 1

practice: lasso regression

Exercise 2

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

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

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*:

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

Exercise 2

Exercise 2

Exercise 2

practice: lasso regression

Exercise 3

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

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

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

Exercise 3

Exercise 3

Exercise 3

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. $A^{\dagger}AA^{\dagger} = A^{\dagger}$
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

Exercise 4.1

Exercise 4.2