Lab 1 - Advanced Machine Learning

Linear Regression

Pierre HOUDOUIN

CentraleSupélec

29 Septembre 2023

General information

- **Assignment** : Alone or in pairs, you will code the algorithms you learnt in Ścikit-learn formalism’
- **Due** : Each assignment has to be sent at most 7 days after the lab session at pierre.houdouin@centralesupelec.fr
- **Grading** : There are 5 lab sessions, each lab session is worth 4 points, the average will count for half of your final grade
- **Questions** : If you have questions, comments or feedbacks about the lab session, feel free to contact me by email

Lesson recap

Linear regression

- We suppose that the relation between the input variables $x_1, ...x_d$ and the output variable y is a linear relation : $y = \beta_0 1 + \beta_1 x_1 + ... + \beta_d x_d$

Lesson recap

Linear regression

- We suppose that the relation between the input variables $x_1, ...x_d$ and the output variable y is a linear relation : $y = \beta_0 1 + \beta_1 x_1 + ... + \beta_d x_d$

Training

- Using a train set of n examples $\mathbf{X}$ with target prediction $\mathbf{Y}$, we look to find the best coefficients $\beta_0, ..., \beta_d$
- What does best means ?

Lesson recap

Linear regression

- We suppose that the relation between the input variables $x_1, \dots x_d$ and the output variable y is a linear relation : $y = \beta_0 1 + \beta_1 x_1 + \dots + \beta_d x_d$

Training

- Using a train set of n examples $\mathbf{X}$ with target prediction $\mathbf{Y}$, we look to find the best coefficients $\beta_0, \dots, \beta_d$
- What does best means ?

Optimization criterion

- We look to minimize an **objective function** involving $\mathbf{X}$ and $\mathbf{Y}$ using a well-suited algorithm

Least square objective function

Objective

$$min_{\boldsymbol{\beta}}||\mathbf{X}\boldsymbol{\beta} - \mathbf{Y}||_2^2$$

Closed-form solution

$$\hat{\boldsymbol{\beta}} = \left(\mathbf{X}^{\mathbf{T}}\mathbf{X}\right)^{-1}\mathbf{X}^{\mathbf{T}}\mathbf{Y}$$

Algorithm

The well suited algorithm is **gradient descent**

Effect

Vanilla

Ridge

Objective

$$min_{\boldsymbol{\beta}}||\mathbf{X}\boldsymbol{\beta} - \mathbf{Y}||_2^2 + \lambda||\boldsymbol{\beta}||_2^2$$

Closed-form solution

$$\hat{\boldsymbol{\beta}} = \left(\mathbf{X}^{\mathbf{T}}\mathbf{X} + \lambda\mathbf{I}\right)^{-1}\mathbf{X}^{\mathbf{T}}\mathbf{Y}$$

Algorithm

The well suited algorithm is **gradient descent**

Effect

Shrinkage

LASSO

Objective

$$min_{\boldsymbol{\beta}}||\mathbf{X}\beta - \mathbf{Y}||_2^2 + \lambda||\beta||_1^1$$

Closed-form solution

Unavailable

Algorithm

The well suited algorithm is **proximal gradient such as Iterative Shrinkage-Thresholding Algorithm**

Effect

Sparsity

Robust

Objective

$$min_{\boldsymbol{\beta}} \rho \left(\mathbf{X}\boldsymbol{\beta} - \mathbf{Y} \right)$$

Closed-form solution

Unavailable

Algorithm

The well suited algorithm is **Iteratively Reweighted Least Square**

Effect

Robustness

Example of potential functions

- **Huber** potential :

$$\rho(e) = \begin{cases} \frac{e^2}{2} & \text{if} |e| \leq \delta \\ \delta |e| - \frac{\delta^2}{2} & \text{if} |e| > \delta \end{cases}$$

- **Bisquare** potential :

$$\rho(e) = \begin{cases} \frac{\delta^2}{6} \left(1 - (1 - \frac{e^2}{\delta^2})^3\right) & \text{if} |e| \leq \delta \\ \frac{\delta^2}{6} & \text{if} |e| > \delta \end{cases}$$