Linear Regression (Part 1)

Deep into MLF

Hao Ni
University College London
The Alan Turing Institute

Overview

Housing Data Example

Task: How to predict the median house value based on the factors, e.g. longitude, latitude, median age of the housing, etc.

Input X:

	longitude	latitude	housing_median_age	total_rooms	total_bedrooms	population	households	median_income
0	-122.23	37.88	41.0	880.0	129.0	322.0	126.0	8.3252
1	-122.22	37.86	21.0	7099.0	1106.0	2401.0	1138.0	8.3014
2	-122.24	37.85	52.0	1467.0	190.0	496.0	177.0	7.2574
3	-122.25	37.85	52.0	1274.0	235.0	558.0	219.0	5.6431
4	-122.25	37.85	52.0	1627.0	280.0	565.0	259.0	3.8462

Output Y:

median_house_value
452600.0
358500.0
352100.0
341300.0
342200.0

Figure: Housing data downloaded via the link have 20640 samples [1].

Setup of Linear regression

Linear Regression Setting

Dataset: Let $\mathcal{D} = \{(x_i, y_i) \in \mathbb{R}^d \times \mathbb{R}\}_{i=1}^N$ such that

$$y_i = f_\theta(x_i) + \varepsilon_i,$$

where $\theta := (\theta_i^{(1)}, \cdots, \theta_i^{(d)})^T \in \mathbb{R}^d$ and $x_i := (x_i^{(1)}, \cdots, x_i^{(d)}) \in \mathbb{R}^d$ and $f_\theta : \mathbb{R}^d \to \mathbb{R}$ is a linear function such that

$$f_\theta(x) = \sum_{j=1}^{d} \theta^{(j)} x^{(j)} \leftarrow \text{(Linear Model)}.$$

Further Assumptions: Assume that the noise term ε_i has Gaussian distribution and iid with $\mathbb{E}[\varepsilon_i | x_i] = 0$.

We also adopt the matrix form for $\mathcal{D} = (X, Y)$, where

$$X = \begin{pmatrix} x_1^{(1)}, & x_1^{(2)}, & \cdots, & x_1^{(d)} \\ \vdots & \vdots & & \vdots \\ x_N^{(1)}, & x_N^{(2)}, & \cdots, & x_N^{(d)} \end{pmatrix} \text{ and } Y = \begin{pmatrix} y_1 \\ \vdots \\ y_N \end{pmatrix}.$$

Question: What is the loss function of linear regression to choose the best estimator for θ ?
Answer: Mean squared error of the model estimated output, i.e.

$$L(\hat{Y}, Y) = \frac{1}{N} \sum_{i=1}^{N} |\hat{y}_i - y_i|^2,$$

where $\hat{Y} = \begin{pmatrix} \hat{y}_1, \cdots, \hat{y}_N \end{pmatrix}^T$ is the model estimated output.
Remark: Note that under the above model assumption, minimizing the loss function of OLS is equivalent to maximizing the log-likelihood ratio of the observation of the output Y conditional on X.

Linear Regression (Ordinary Least Squares)

Ordinary Least Squares (OLS)

Data	$\mathcal{D} = \{(x_i, y_i) \in \mathbb{R}^d \times \mathbb{R}\}_{i=1}^N$.
Model:	$y = f_{\theta}(x) + \varepsilon = x\theta + \varepsilon$.
Loss Function:	$L(\theta \mid X, Y) = (Y - X\theta)^T(Y - X\theta) \to \min$.
Optimization:	$\hat{\theta} = (X^T X)^{-1} X^T y$.
Prediction:	$\hat{y}_* = x_* \hat{\theta}$.
Validation:	Compute RMSE, R^2, the adjusted R^2, p-value.

Conclusion

- This video introduces Linear regression under the framework of supervised learning covered in Step 2.2.
- This video is based on Chapter 3 of my book [2] (Chinese version) and [3] (English version).
- One can refer to the code examples of linear regression to the Chapter3_LinearRegression folder at the link `https://github.com/deepintomlf/mlfbook`.
- Next steps: the derivation of the optimal linear coefficient estimator (Step 2.6) and numerical experiments (Step 2.7).

Thanks for your attention!

References I

Aurélien Géron.
Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow: Concepts, tools, and techniques to build intelligent systems.
" O'Reilly Media, Inc.", 2019.

Hao Ni, Xin Dong, Jinsong Zheng, and Guangxi Yu.
An Introduction to Machine Learning in Quantitative Finance (Chinese version).
Tsinghua University Press, 2021.

Hao Ni, Xin Dong, Jinsong Zheng, and Guangxi Yu.
An Introduction to Machine Learning in Quantitative Finance (English version).
World Scientific, 2021.