Introduction to Machine Learning
Lab 2: Polynomial Regression

Hongteng Xu

March 13, 2022

1 Motivation

- Try to implement various training methods for polynomial regression

- Implement commonly-used data normalization methods and check their influences on the MSE of testing data.

- Quantitatively analyze the influence of the polynomial's order on the model complexity and the learning results.

2 Tasks

1. Given the problem

$$\min_{\boldsymbol{w}} \|\boldsymbol{y} - \boldsymbol{X}\boldsymbol{w}\|_2^2, \tag{1}$$

   where $\boldsymbol{X} = \{x_n^{d-1}\}_{n,d=1}^{N,D} \in \mathbb{R}^{N \times D}$ is the Vandermonde matrix, implement the function "training" to achieve its closed-form solution.

2. Given the problem in (1), implement the function "training_sgd" to learn the model via stochastic gradient descent.

3. Given the above data matrix $\boldsymbol{X} = [\boldsymbol{x}_1, ..., \boldsymbol{x}_D]$, implement different normalization strategies:

$$\tilde{\boldsymbol{X}} = \boldsymbol{X}\mathrm{diag}(\boldsymbol{a}), \quad \boldsymbol{a} = [\frac{1}{\|\boldsymbol{x}_1\|_p}, ..., \frac{1}{\|\boldsymbol{x}_D\|_p}]^T, \quad p = 1, 2, \infty, \tag{2}$$

   where $\boldsymbol{a}$ is called scaling vector.

4. Given the normalization method and the training algorithm, try to combine them together, achieving a SGD algorithm based on normalized data.

5. Implement a testing method for the model trained on the normalized data.