Exercise 5 Derivation of Ridge Regression

$44251017\ Huang\ Jiahui$

4.4 リッジ回帰の最急降下法と確率的最急降下法

$\boldsymbol{x}_i = \begin{pmatrix} 1 \\ x_{i1} \\ \vdots \\ x_{iD} \end{pmatrix}, \boldsymbol{X} = \begin{pmatrix} \boldsymbol{x}_1^T \\ \boldsymbol{x}_2^T \\ \vdots \\ \boldsymbol{x}_N^T \end{pmatrix}, \boldsymbol{t} = \begin{pmatrix} t_1 \\ t_2 \\ \vdots \\ t_N \end{pmatrix}, \boldsymbol{w} = \begin{pmatrix} w_0 \\ w_1 \\ \vdots \\ w_D \end{pmatrix}$ とする。事例群 $\{(\boldsymbol{x}_i, t_i)\}_{i=1}^N$ を使ってリッジ回帰による線形モデルを求めることを考える。

1. Derive the update formula of the gradient descent method of ridge regression
2. Derive the update formula of the stochastic gradient descent method of ridge regression

Q1

Ridge regression loss function:

$$E(w) = \frac{1}{2}\|t - Xw\|^2 + \frac{\lambda}{2}\|w\|^2 \tag{1}$$

Gradient:

$$\nabla_w E = X^T(Xw - t) + \lambda w \tag{2}$$

Gradient descent update rule:

$$w_{t+1} = w_t - \eta\left[X^T(Xw_t - t) + \lambda w_t\right] \tag{3}$$

Q2

For a single training example (x_i, t_i), the loss function is:

$$E_i(w) = \frac{1}{2}(t_i - w^T x_i)^2 + \frac{\lambda}{2}\|w\|^2 \tag{4}$$

Gradient:

$$\nabla_w E_i = -(t_i - w^T x_i)x_i + \lambda w = (w^T x_i - t_i)x_i + \lambda w \tag{5}$$

SGD update rule:

$$w \leftarrow w - \eta\left[(w^T x_i - t_i)x_i + \lambda w\right] \tag{6}$$