

# Exercise 4:

## Derivation of linear regression

$$\mathbf{x}_i = \begin{pmatrix} 1 \\ x_{i1} \\ \vdots \\ x_{iD} \end{pmatrix}, \mathbf{X} = \begin{pmatrix} \mathbf{x}_1^T \\ \mathbf{x}_2^T \\ \vdots \\ \mathbf{x}_N^T \end{pmatrix}, \mathbf{t} = \begin{pmatrix} t_1 \\ t_2 \\ \vdots \\ t_N \end{pmatrix}, \mathbf{w} = \begin{pmatrix} w_0 \\ w_1 \\ \vdots \\ w_D \end{pmatrix}$$

Find a linear regression model  $\mathbf{t} = \mathbf{w}^T \mathbf{x}$  using the training samples

1. Express the sum of squared errors  $E$  as a function of  $\mathbf{w}$
2. Derive the following (a)(b) to find the gradient  $\nabla_{\mathbf{w}} E$  for  $\mathbf{w}$  of the sum of squared errors  $E$

$$(a) \sum_{i=1}^N t_i \mathbf{x}_i = \mathbf{X}^T \mathbf{t}$$

$$(b) \sum_{i=1}^N \mathbf{x}_i \mathbf{w}^T \mathbf{x}_i = \mathbf{X}^T \mathbf{X} \mathbf{w}$$

3. Show the gradient  $\frac{\partial E}{\partial \mathbf{w}}$  in terms of  $\mathbf{x}_i$  (or  $\mathbf{X}$ ),  $\mathbf{t}$
4. (Approximate solution) Show the parameter update equation for the linear regression model using the gradient descent method in terms of  $\mathbf{x}_i$  (or  $\mathbf{X}$ ),  $\mathbf{t}$   
Initial solution  $\mathbf{w}^0$ , t-th update  $\mathbf{w}^t$ , step size parameter  $\eta$
5. (Analytic solution) Show that  $\mathbf{w}$  where  $\frac{\partial E}{\partial \mathbf{w}} = 0$  is

$$\mathbf{w} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{t}$$

# Exercise 4:

## Derivation of linear regression

$\mathbf{x}_i = \begin{pmatrix} 1 \\ x_{i1} \\ \vdots \\ x_{iD} \end{pmatrix}, \mathbf{X} = \begin{pmatrix} \mathbf{x}_1^T \\ \mathbf{x}_2^T \\ \vdots \\ \mathbf{x}_N^T \end{pmatrix}, \mathbf{t} = \begin{pmatrix} t_1 \\ t_2 \\ \vdots \\ t_N \end{pmatrix}, \mathbf{w} = \begin{pmatrix} w_0 \\ w_1 \\ \vdots \\ w_D \end{pmatrix}$  とする。事例群  $\{(\mathbf{x}_i, t_i)\}_{i=1}^N$  を使って線形回帰モデル  $t = \mathbf{w}^T \mathbf{x}$  を求めることを考える。

1. 二乗誤差和  $E$  を  $\mathbf{w}$  の関数で表せ
2. 二乗誤差和  $E$  の  $\mathbf{w}$  についての勾配  $\nabla_{\mathbf{w}} E$  を求めるために、以下を導出せよ
  - (a)  $\sum_{i=1}^N t_i \mathbf{x}_i = \mathbf{X}^T \mathbf{t}$
  - (b)  $\sum_{i=1}^N \mathbf{x}_i \mathbf{w}^T \mathbf{x}_i = \mathbf{X}^T \mathbf{X} \mathbf{w}$
3. 勾配  $\frac{\partial E}{\partial \mathbf{w}}$  を  $\mathbf{x}_i$  (あるいは  $\mathbf{X}$ ),  $\mathbf{t}$  の式で示せ。
4. (近似解法) 線形回帰モデルを最急降下法で求めるときの、パラメータの更新式を  $\mathbf{x}_i$  (あるいは  $\mathbf{X}$ ),  $\mathbf{t}$  の式で示せ。初期解を  $\mathbf{w}^0$ ,  $t$  回目の更新時の回を  $\mathbf{w}^t$ , ステップサイズパラメータを  $\eta$  とする。
5. (解析解)  $\frac{\partial E}{\partial \mathbf{w}} = 0$  なる  $\mathbf{w}$  が  $\mathbf{w} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{t}$  であることを示せ。