Exercise 4:
Derivation of Linear Regression

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

1. 二乗誤差和 E を $\boldsymbol{w}$ の関数で表せ
2. 二乗誤差和 E の $\boldsymbol{w}$ についての勾配 $\nabla_{\boldsymbol{w}} E$ を求めるために、以下を導出せよ
   (a) $\sum_{i=1}^N t_i \boldsymbol{x}_i = \boldsymbol{X}^T\boldsymbol{t}$
   (b) $\sum_{i=1}^N \boldsymbol{x}_i \boldsymbol{w}^T \boldsymbol{x}_i = \boldsymbol{X}^T\boldsymbol{X}\boldsymbol{w}$
3. 勾配 $\frac{\partial E}{\partial \boldsymbol{w}}$ を $\boldsymbol{x}_i$(あるいは $\boldsymbol{X}$), $\boldsymbol{t}$ 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 $\boldsymbol{x}_i$(あるいは $\boldsymbol{X}$), $\boldsymbol{t}$ の式で示せ。初期解を $\boldsymbol{w}^0$, t 回目の更新時の回を $\boldsymbol{w}^t$, ステップサイズパラメータを η とする。
5. (解析解) $\frac{\partial E}{\partial \boldsymbol{w}} = 0$ なる $\boldsymbol{w}$ が $\boldsymbol{w} = (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{t}$ であることを示せ。

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

1. 二乗誤差和 E を $\boldsymbol{w}$ の関数で表せ
   $E(\boldsymbol{w}) = \sum_{i=1}^N (t_i - \boldsymbol{w}^T \boldsymbol{x}_i)^2$ もしくは $E(\boldsymbol{w}) = ||\boldsymbol{t} - \boldsymbol{X}\boldsymbol{w}||_2^2$
2. 二乗誤差和 E の $\boldsymbol{w}$ についての勾配 $\nabla_{\boldsymbol{w}} E$ を求めるために、以下を導出せよ
   (a) $\sum_{i=1}^N t_i \boldsymbol{x}_i = \boldsymbol{X}^T \boldsymbol{t}$

$$\boldsymbol{X}^T\boldsymbol{t} = \begin{pmatrix} \boldsymbol{x}_1^T \\ \boldsymbol{x}_2^T \\ \vdots \\ \boldsymbol{x}_N^T \end{pmatrix}^T \begin{pmatrix} t_1 \\ t_2 \\ \vdots \\ t_N \end{pmatrix} = \begin{pmatrix} \boldsymbol{x}_1 \ \boldsymbol{x}_2 \ \dots \ \boldsymbol{x}_N \end{pmatrix} \begin{pmatrix} t_1 \\ t_2 \\ \vdots \\ t_N \end{pmatrix} = \textstyle\sum_{i=1}^N t_i\boldsymbol{x}_i$$

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

$$\begin{aligned} \boldsymbol{X}^T\boldsymbol{X}\boldsymbol{w} &= \begin{pmatrix} \boldsymbol{x}_1^T \\ \boldsymbol{x}_2^T \\ \vdots \\ \boldsymbol{x}_N^T \end{pmatrix}^T \begin{pmatrix} \boldsymbol{x}_1^T \\ \boldsymbol{x}_2^T \\ \vdots \\ \boldsymbol{x}_N^T \end{pmatrix} \begin{pmatrix} w_0 \\ w_1 \\ \vdots \\ w_D \end{pmatrix} = \begin{pmatrix} \boldsymbol{x}_1 \ \boldsymbol{x}_2 \ \dots \ \boldsymbol{x}_N \end{pmatrix} \begin{pmatrix} \boldsymbol{x}_1^T\boldsymbol{w} \\ \boldsymbol{x}_2^T\boldsymbol{w} \\ \vdots \\ \boldsymbol{x}_N^T\boldsymbol{w} \end{pmatrix} \\ &= \begin{pmatrix} \boldsymbol{x}_1 \ \boldsymbol{x}_2 \ \dots \ \boldsymbol{x}_N \end{pmatrix} \begin{pmatrix} \boldsymbol{w}^T\boldsymbol{x}_1 \\ \boldsymbol{w}^T\boldsymbol{x}_2 \\ \vdots \\ \boldsymbol{w}^T\boldsymbol{x}_N \end{pmatrix} = \sum_{i=1}^N \boldsymbol{x}_i\boldsymbol{w}^T\boldsymbol{x}_i \end{aligned}$$

3. 勾配 $\frac{\partial E}{\partial \boldsymbol{w}}$ を $\boldsymbol{x}_i$(あるいは $\boldsymbol{X}$), $\boldsymbol{t}$ の式で示せ。

$$\begin{aligned} \frac{\partial E}{\partial \boldsymbol{w}} &= \frac{\partial}{\partial \boldsymbol{w}} \sum_{i=1}^N (t_i - \boldsymbol{w}^T\boldsymbol{x}_i)^2 \\ &= \frac{\partial}{\partial \boldsymbol{w}} \sum_{i=1}^N t_i^2 - 2t_i\boldsymbol{w}^T\boldsymbol{x}_i + (\boldsymbol{w}^T\boldsymbol{x}_i)^2 \\ &= \sum_{i=1}^N -2t_i\boldsymbol{x}_i + 2\boldsymbol{x}_i\boldsymbol{w}^T\boldsymbol{x}_i \quad \left(\because \frac{\partial \boldsymbol{x}^T\boldsymbol{a}}{\partial \boldsymbol{x}} = \frac{\partial \boldsymbol{a}^T\boldsymbol{x}}{\partial \boldsymbol{x}} = \boldsymbol{a}\right) \\ &= 2\sum_{i=1}^N \boldsymbol{x}_i\boldsymbol{w}^T\boldsymbol{x}_i - t_i\boldsymbol{x}_i \end{aligned}$$

$$\therefore \frac{\partial E}{\partial \boldsymbol{w}} = 2\sum_{i=1}^N \boldsymbol{x}_i\boldsymbol{w}^T\boldsymbol{x}_i - t_i\boldsymbol{x}_i$$

また、2. より

$$\therefore \frac{\partial E}{\partial \boldsymbol{w}} = 2\left(\boldsymbol{X}^T\boldsymbol{X}\boldsymbol{w} - \boldsymbol{X}^T\boldsymbol{t}\right)$$

とも表せる。

別の解答として、

$$\begin{aligned}
\frac{\partial E}{\partial \boldsymbol{w}} &= \frac{\partial}{\partial \boldsymbol{w}}||\boldsymbol{t}-\boldsymbol{X}\boldsymbol{w}||_2^2 = \frac{\partial}{\partial \boldsymbol{w}}(\boldsymbol{t}-\boldsymbol{X}\boldsymbol{w})^T(\boldsymbol{t}-\boldsymbol{X}\boldsymbol{w}) \\
&= \frac{\partial}{\partial \boldsymbol{w}}\boldsymbol{t}^T\boldsymbol{t} - \boldsymbol{t}^T\boldsymbol{X}\boldsymbol{w} - \boldsymbol{w}^T\boldsymbol{X}^T\boldsymbol{t} + \boldsymbol{w}^T\boldsymbol{X}^T\boldsymbol{X}\boldsymbol{w} \\
&= -(\boldsymbol{t}^T\boldsymbol{X})^T - \boldsymbol{X}^T\boldsymbol{t} + 2\boldsymbol{X}^T\boldsymbol{X}\boldsymbol{w} \quad \left(\because \frac{\partial \boldsymbol{x}^T\boldsymbol{a}}{\partial \boldsymbol{x}} = \frac{\partial \boldsymbol{a}^T\boldsymbol{x}}{\partial \boldsymbol{x}} = \boldsymbol{a}, \frac{\partial \boldsymbol{x}^T\boldsymbol{A}^T\boldsymbol{A}\boldsymbol{x}}{\partial \boldsymbol{x}} = 2\boldsymbol{A}^T\boldsymbol{A}\boldsymbol{x}\right) \\
&= 2\left(\boldsymbol{X}^T\boldsymbol{X}\boldsymbol{w} - \boldsymbol{X}^T\boldsymbol{t}\right)
\end{aligned}$$

$$\therefore \frac{\partial E}{\partial \boldsymbol{w}} = 2\left(\boldsymbol{X}^T\boldsymbol{X}\boldsymbol{w} - \boldsymbol{X}^T\boldsymbol{t}\right)$$

としてもよい。

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 $\boldsymbol{x}_i$(あるいは $\boldsymbol{X}$), $\boldsymbol{t}$ の式で示せ。初期解を $\boldsymbol{w}^0$, t 回目の更新時の回を $\boldsymbol{w}^t$, ステップサイズパラメータを η とする。

   パラメータの更新式は $\boldsymbol{w}^{t+1} \leftarrow \boldsymbol{w}^t - \eta\nabla_{\boldsymbol{w}}E$ となるため、3. から、

$$\boldsymbol{w}^{t+1} \leftarrow \boldsymbol{w}^t - 2\eta\sum_{i=1}^{N}\boldsymbol{x_i}(\boldsymbol{w}^t)^T\boldsymbol{x}_i - t_i\boldsymbol{x}_i$$

   もしくは、

$$\boldsymbol{w}^{t+1} \leftarrow \boldsymbol{w}^t - 2\eta(\boldsymbol{X}^T\boldsymbol{X}\boldsymbol{w}^t - \boldsymbol{X}^T\boldsymbol{t})$$

5. (解析解) $\frac{\partial E}{\partial \boldsymbol{w}} = 0$ なる $\boldsymbol{w}$ が $\boldsymbol{w} = (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{t}$ であることを示せ。

$$\begin{aligned}
\frac{\partial E}{\partial \boldsymbol{w}} = 2\left(\boldsymbol{X}^T\boldsymbol{X}\boldsymbol{w} - \boldsymbol{X}^T\boldsymbol{t}\right) &= \boldsymbol{0} \\
\iff \boldsymbol{X}^T\boldsymbol{X}\boldsymbol{w} &= \boldsymbol{X}^T\boldsymbol{t}
\end{aligned}$$

   ここで、$\boldsymbol{X}^T\boldsymbol{X}$ の逆行列が存在するならば、

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