ML Exercise Sheet 4

Trung Pham - 7176267 Viktor Vironski - 4330455
Andy Disser - 5984875

23. Mai 2022

Exercise 1.1

- To be proven :

$$\sum_{n=1}^{N}(x_n - \bar{x})(y_n - \bar{y}) = \sum_{n=1}^{N}(y_n - \bar{y})y_n$$

- Start from the left side

$$\begin{aligned}\sum_{n=1}^{N}(x_n - \bar{x})(y_n - \bar{y}) &= \sum_{n=1}^{N} x_n y_n - x_n\bar{y} - \bar{x}y_n + \bar{x}\bar{y} \\ &= \sum_{n=1}^{N}\left(x_n y_n - x_n\bar{y}\right) - \sum_{n=1}^{N}\bar{x}y_n + \sum_{n=1}^{N}\bar{x}\bar{y}\end{aligned}$$

- We show that

$$-\sum_{n=1}^{N}\bar{x}y_n + \sum_{n=1}^{N}\bar{x}\bar{y} = 0$$

then the claim would be verified

$$\begin{aligned}-\sum_{n=1}^{N}\bar{x}y_n + \sum_{n=1}^{N}\bar{x}\bar{y} &= -\sum_{n=1}^{N}\bar{x}y_n + N\cdot\frac{1}{N}\left(\sum_{i=1}^{N}x_i\right)\cdot\frac{1}{N}\left(\sum_{i=1}^{N}y_i\right) \\ &= -\sum_{n=1}^{N}\frac{1}{N}\left(\sum_{i=1}^{N}x_i\right)y_n + \frac{1}{N}\left(\sum_{i=1}^{N}x_i\right)\cdot\left(\sum_{i=1}^{N}y_i\right) \\ &= -\frac{1}{N}\left(\sum_{n=1}^{N}\sum_{i=1}^{N}x_i y_n - \left(\sum_{i=1}^{N}x_i\right)\left(\sum_{i=1}^{N}y_i\right)\right)\end{aligned}$$

- Because of the distributive property

$$\left(\sum_{i=1}^{N}x_i\right)\left(\sum_{i=1}^{N}y_i\right) = \left(\sum_{i=1}^{N}x_i\right)(y_1 + y_2 + \ldots + y_N) = \sum_{n=1}^{N}\left(\sum_{i=1}^{N}x_i\right)\cdot y_n$$

- That's why

$$
\begin{aligned}
-\frac{1}{N}\left(\sum_{n=1}^{N}\sum_{i=1}^{N} x_i y_n - \left(\sum_{i=1}^{N} x_i\right)\left(\sum_{i=1}^{N} y_i\right)\right) &= -\frac{1}{N}\left(\sum_{n=1}^{N}\sum_{i=1}^{N} x_i y_n - \sum_{n=1}^{N}\sum_{i=1}^{N} x_i y_n\right) \\
&= -\frac{1}{N}\cdot 0 \\
&= 0 \qquad \text{(Which is what we wanted to show)}
\end{aligned}
$$

Exercise 1.2

- Let $\mathbf{x}_i = \begin{pmatrix} x_{i0} & x_{i1} & x_{i2} & \dots & x_{id} \end{pmatrix}^\top$ be the vector of the i-th sample including an additional dimension $x_{i0} = 1$ for all $i = 1, \dots, N$

- Let $\mathbf{y} = \begin{pmatrix} y_1 & y_2 & y_3 & \dots & y_n \end{pmatrix}^\top$, the matrix $\mathbf{X} = \begin{pmatrix} x_{10} & x_{11} & x_{12} & \dots & x_{1d} \\ x_{20} & x_{21} & x_{22} & \dots & x_{2d} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ x_{N0} & x_{N1} & x_{N2} & \dots & x_{Nd} \end{pmatrix}$

- The matrix product $:= \mathbf{X}^\top\mathbf{X} = (e_{ij})_{i,j=0,\dots,d}$ is a $(d+1)\times(d+1)$-matrix whose entry e_{ij} represents the inner product of the i-th and the j-th independent variables, i.e. $e_{ij} = \sum_{k=1}^{N} x_{ki}x_{kj}$

- Let $\mathbf{w}^* = \begin{pmatrix} w_0 & w_1 & w_2 & \dots & w_d \end{pmatrix}^\top$ be the given vector, we have

$$
\begin{aligned}
& & \mathbf{w}^* &= (\lambda\mathbf{I} + \mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\mathbf{y} \\
\Leftrightarrow & & (\lambda\mathbf{I} + \mathbf{X}^\top\mathbf{X})\mathbf{w}^* &= \mathbf{X}^\top\mathbf{y} && (1) \\
\Leftrightarrow & & \begin{pmatrix} e_{00}+\lambda & e_{01} & e_{02} & \dots & e_{0d} \\ e_{10} & e_{11}+\lambda & e_{12} & \dots & e_{1d} \\ e_{20} & e_{21} & e_{22}+\lambda & \dots & e_{2d} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ e_{d0} & e_{d1} & e_{d2} & \dots & e_{dd}+\lambda \end{pmatrix} \begin{pmatrix} w_0 \\ w_1 \\ w_2 \\ \vdots \\ w_d \end{pmatrix} &= \begin{pmatrix} \sum_{k=1}^{N} x_{k0}y_k \\ \sum_{k=1}^{N} x_{k1}y_k \\ \sum_{k=1}^{N} x_{k2}y_k \\ \vdots \\ \sum_{k=1}^{N} x_{kd}y_k \end{pmatrix} && (2)
\end{aligned}
$$

- Consequently, it holds $\sum_{k=1}^{d} w_k e_{ik} + \lambda w_i = \sum_{k=1}^{N} x_{ki}y_k$ for all $i = 0, \dots, d$ (*)

- Now we need to show that the partial derivatives $\dfrac{\partial E(\mathbf{w}^*)}{\partial w_i} = 0$ for all $i = 0, \dots, d$

- It holds for each $i = 0, \ldots, d$:

$$
\begin{aligned}
\frac{\partial E(\mathbf{w}^*)}{\partial w_i} &= \frac{1}{2} \sum_{n=1}^{N} 2(-x_{ni})(y_n - w_0 x_{n0} - w_1 x_{n1} - w_2 x_{n2} - \ldots - w_d x_{nd}) + \frac{1}{2} \cdot 2\lambda w_i \\
&= \sum_{n=1}^{N} x_{ni}(-y_n + w_0 x_{n0} + w_1 x_{n1} + w_2 x_{n2} + \ldots + w_d x_{nd}) + \lambda w_i \\
&= \quad x_{1i}(-y_1 + w_0 x_{10} + w_1 x_{11} + w_2 x_{12} + \ldots + w_d x_{1d}) \\
&\quad + x_{2i}(-y_2 + w_0 x_{20} + w_1 x_{21} + w_2 x_{22} + \ldots + w_d x_{2d}) \\
&\quad + x_{3i}(-y_3 + w_0 x_{30} + w_1 x_{31} + w_2 x_{32} + \ldots + w_d x_{3d}) \\
&\qquad \vdots \qquad\qquad \vdots \qquad\qquad\qquad \vdots \qquad\qquad \vdots \\
&\quad + x_{Ni}(-y_N + w_0 x_{N0} + w_1 x_{N1} + w_2 x_{N2} + \ldots + w_d x_{Nd}) + \lambda w_i \\
&= -\sum_{k=1}^{N} x_{ki} y_k + w_0 \sum_{k=1}^{N} x_{ki} x_{k0} + w_1 \sum_{k=1}^{N} x_{ki} x_{k1} + \ldots + w_d \sum_{k=1}^{N} x_{ki} x_{kd} + \lambda w_i \\
&= -\sum_{k=1}^{N} x_{ki} y_k + w_0 e_{i0} + w_1 e_{i1} + \ldots + w_d e_{id} + \lambda w_i \\
&= -\sum_{k=1}^{N} x_{ki} y_k + \sum_{k=1}^{N} w_k e_{ik} + \lambda w_i \\
&= -\lambda w_i + \lambda w_i \quad \text{(because of (*))} \\
&= 0
\end{aligned}
$$

- Therefore $\mathbf{w}^*$ minimizes $E(\mathbf{w})$, which is wanted to show