Schant Demircian
Sören Krüger
Leon Happ

Small note: We use your numbering of the tasks in our comments for readability, but they were given differently in the exercise sheet.

Exercise 3

3.1

We will use the linearity of the inner and outer product.

Comment for 3.1a) Your solution is more transparent than the sample solution, since you also included the steps in your calculation, which is great. Otherwise your solution is of course correct and your result is the same as in the sample solution.

a)

It is

$$\frac{\partial}{\partial b}\left[\sum_{i=1}^{N}(w^T\cdot x_i+b-y_i)^2\right]=\sum_{i=1}^{N}(2(w^T\cdot x_i+b-y_i))=2Nb+2\sum_{i=1}^{N}(w^T\cdot x_i-y_i)$$
$$\overset{!}{=}0.$$

Thus,

$$\begin{aligned}\hat{b}&=\frac{1}{N}\sum_{i=1}^{N}(y_i-w^T\cdot x_i)=-\frac{1}{N}\sum_{i=1}^{N}w^T\cdot x_i\\&=-\frac{1}{2}w^t\cdot(\mu_1+\mu_{-1}),\end{aligned}\tag{1.1}$$

where we used the important property

$$\frac{1}{N}\sum_{i:y_i=\pm1}x_i=\frac{1}{2}\mu_{\pm1}.\tag{1.2}$$

b)

It is

$$\begin{aligned}\frac{\partial}{\partial w}\left[\sum_{i=1}^{N}(w^Tx_i+\hat{b}-y_i)^2\right]&\overset{a)}{=}\sum_{i=1}^{N}(2x_i(w^T\cdot x_i)+2x_i\hat{b}-2x_iy_i)\\&\overset{\text{Hinweis},(1.1)}{=}2\sum_{i=1}^{N}((x_ix_i^T)\cdot w-\frac{1}{2}\left(x_i(\mu_1+\mu_{-1})^T\right)\cdot w-x_iy_i)\\&\overset{!}{=}0.\end{aligned}$$

Thus

$$\frac{1}{N}\sum_{i=1}^{N}x_i\left(x_i-\frac{1}{2}(\mu_1+\mu_{-1})\right)^T\cdot\hat{w}=\frac{1}{N}\sum_{i=1}^{N}x_iy_i.\tag{1.3}$$

By multiplying out the LHS we can simplify it further

$$\begin{aligned}
&\frac{1}{N}\sum_{i=1}^{N} x_i \left(x_i - \frac{1}{2}(\mu_1 + \mu_{-1})\right)^T \\
&= \frac{1}{N}\sum_{i=1}^{N} x_i x_i^T - \frac{1}{2N}\sum_{i=1}^{N} x_i \mu_1^T - \frac{1}{2N}\sum_{i=1}^{N} x_i \mu_{-1}^T \\
&\overset{(1.2)}{=} \frac{1}{N}\sum_{i=1}^{N} x_i x_i^T - \frac{1}{4}\mu_1\mu_1^T - \frac{1}{4}\mu_{-1}\mu_1^T - \frac{1}{4}\mu_1\mu_{-1}^T - \frac{1}{4}\mu_{-1}\mu_{-1}^T \\
&= \frac{1}{N}\sum_{i=1}^{N} x_i x_i^T - \frac{1}{4}(\mu_1\mu_1^T + \mu_1\mu_{-1}^T + \mu_1\mu_{-1}^T + \mu_{-1}\mu_{-1}^T). \qquad (1.4)
\end{aligned}$$

On the other hand it is by definition

$$\begin{aligned}
S_W &= \frac{1}{N}\sum_{i=1}^{N}(x_i - \mu_{y_i})(x_i - \mu_{y_i})^T \\
&= \frac{1}{N}\sum_{i=1}^{N}(x_i x_i^T - x_i\mu_{y_i}^T - \mu_{y_i}x_i^T + \mu_{y_i}\mu_{y_i}^T) \\
&\overset{(1.2)}{=} \frac{1}{N}\sum_{i=1}^{N} x_i x_i^T - \frac{1}{2}(\mu_1\mu_1^T + \mu_{-1}\mu_{-1}^T) - \frac{1}{2}(\mu_1\mu_1^T + \mu_{-1}\mu_{-1}^T) + \frac{1}{2}(\mu_1\mu_1^T + \mu_{-1}\mu_{-1}^T) \\
&= \frac{1}{N}\sum_{i=1}^{N} x_i x_i^T - \frac{1}{2}(\mu_1\mu_1^T + \mu_{-1}\mu_{-1}^T)
\end{aligned}$$

Comment for 3.1b) Your solution is once more equivalent to the sample solution, although the ordering of some of the steps have been different. The way you calculated the left-hand side of the equation seems more transparent than the way it's presented in the sample solution. Maybe you could write the final result (the final equation) explicitly, but not a big thing.

and

$$\begin{aligned}
\frac{1}{4}S_B &= \frac{1}{4}(\mu_1 - \mu_-1)(\mu_1 - \mu_{-1})^T \\
&= \frac{1}{4}\mu_1\mu_1^T + \frac{1}{4}\mu_{-1}\mu_{-1}^T - \frac{1}{4}\mu_1\mu_{-1}^T - \frac{1}{4}\mu_{-1}\mu_1^T,
\end{aligned}$$

whereby we obtain

$$\begin{aligned}
S_W + \frac{1}{4}S_B &= \frac{1}{N}\sum_{i=1}^{N} x_i x_i^T - \frac{1}{4}(\mu_1\mu_1^T + \mu_1\mu_{-1}^T + \mu_1\mu_{-1}^T + \mu_{-1}\mu_{-1}^T) \\
&= (1.4).
\end{aligned}$$

Furthermore, the RHS of (1.3) becomes exactly

$$\begin{aligned}
\frac{1}{N}\sum_{i=1}^{N} x_i y_i &= \frac{1}{N}\sum_{i:y_i=1} x_i + \frac{1}{N}\sum_{i:y_i=-1}(-x_i) \\
&\overset{(1.2)}{=} \frac{1}{2}(\mu_1 - \mu_{-1}).
\end{aligned}$$

c)

Replugging in the definition of S_B we get

$$(S_W + \frac{1}{4}S_B)\cdot\hat{w} = S_W\cdot\hat{w} + (\mu_1 - \mu_-1)\underbrace{\frac{1}{4}(\mu_1 - \mu_{-1})^T\cdot\hat{w}}_{=:\tau'}.$$

This yields

$$\begin{aligned}\hat{w} &= S_W^{-1}\underbrace{(\frac{1}{2} - \tau')}_{=:\tau}(\mu_1 - \mu_{-1}) \\ &= \tau S_W^{-1}(\mu_1 - \mu_{-1}).\end{aligned}$$

Comment for 3.1c) The way you solved this part is completely equivalent to the sample solution.

General comment for 3.1: You derived the necessary equations in a manner completely similar to the sample solution, good job! Our comments for the next exercise is to be found under the code.