Note

We would like to do task 1 as video.

Task 1

Let

$$X = \begin{pmatrix} 1 & -2 \\ 1 & -1 \\ 1 & 0.5 \\ 1 & 0.6 \\ 1 & 5 \\ 1 & 7 \end{pmatrix} X^\top = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 \\ -2 & -1 & 0.5 & 0.6 & 5 & 7 \end{pmatrix}$$

For the loss to be minimal, we need to solve

$$\hat{\beta}^{ls} = (X^\top X)^{-1} X^\top y$$

For this we have

$$(X^\top X)^{-1} = \begin{pmatrix} 6 & 10.1 \\ 10.1 & 79.61 \end{pmatrix}^{-1} = \frac{20}{7513} \begin{pmatrix} 79.61 & -10.1 \\ -10.1 & 6 \end{pmatrix}$$

and

$$X^\top y = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 \\ -2 & -1 & 0.5 & 0.6 & 5 & 7 \end{pmatrix} \begin{pmatrix} 0 \\ 0 \\ 1 \\ 0 \\ 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 3 \\ 12.5 \end{pmatrix}$$

Plugging this into the equation above, we get

$$\hat{\beta}^{ls} = \frac{20}{7513} \begin{pmatrix} 79.61 & -10.1 \\ -10.1 & 6 \end{pmatrix} \begin{pmatrix} 3 \\ 12.5 \end{pmatrix} \approx \begin{pmatrix} -0.28 \\ 0.12 \end{pmatrix}$$

This leads to the predicted values

$$y = X\beta = \begin{pmatrix} 1 & -2 \\ 1 & -1 \\ 1 & 0.5 \\ 1 & 0.6 \\ 1 & 5 \\ 1 & 7 \end{pmatrix} \begin{pmatrix} -0.28 \\ 0.12 \end{pmatrix} \approx \begin{pmatrix} -0.54 \\ -0.42 \\ -0.24 \\ -0.23 \\ 0.3 \\ 0.53 \end{pmatrix}$$

Setting the deciding threshold between the two classes to 0.5, where every value below 0.5 belongs to the class "0" and "1" else, yields

$$\hat{y} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 1 \end{pmatrix}$$

Linear regression is not suitable for classification due the difficult decision of setting a threshold that decides which value belongs to which class. Setting the value too high or too low could lead to a "onesided" result, where most values belong to only one class instead of being distributed to some extent. Our result above is such an example of setting the threshold too high, because only one value is predicted to be in class "1".

Task 2

1.)

$$\begin{aligned}
\frac{\delta L^{nll}(\beta)}{\delta \beta_i} &= \frac{\delta}{\delta \beta_i}\left(-\sum_{i=1}^{n} y_i \log(p(x_i)) + (1-y_i)\log(1-p(x_i))\right) \\
&= -\sum_{i=1}^{n} y_i \frac{1}{p(x_i)} p(x_i)(1-p(x_i))\phi(x_i) + (1-y_i)\frac{1}{p(x_i)-1} p(x_i)(1-p(x_i))\phi(x_i) \\
&= -\sum_{i=1}^{n} y_i(1-p(x_i))\phi(x_i) - (1-y_i)p(x_i)\phi(x_i) \\
&= -\sum_{i=1}^{n} (y_i - p(x_i))\phi(x_i) \\
&= \sum_{i=1}^{n} (p(x_i) - y_i)\phi(x_i)
\end{aligned}$$

2.)

$$\begin{aligned}
\frac{\delta^2 L^{nll}(\beta)}{\delta \beta_i^2} &= \frac{\delta}{\delta \beta_i}\sum_{i=1}^{n}(p(x_i) - y_i)\phi(x_i) \\
&= \sum_{i=1}^{n} \phi(x_i)p(x_i)(1-p(x_i)\phi(x_i)
\end{aligned}$$

Task 3

Choose $f'(x, y) = f(x, y) - f(x, 0)$, then one gets for $y' \in \{0, 1\}$

$$\begin{aligned}
P(y|x) &= \frac{e^{f(x,y)-f(x,0)}}{e^{f(x,0)-f(x,0)} + e^{f(x,1)-f(x,0)}} \\
&= \frac{\frac{e^{f(x,y)}}{e^{f(x,0)}}}{e^{0} + \frac{e^{f(x,1)}}{e^{f(x,0)}}} \\
&= \frac{e^{f(x,y)}}{e^{f(x,0)} + e^{f(x,1)}} \\
&= \frac{e^{f(x,y)}}{\sum_{y'} e^{f(x,y')}}
\end{aligned}$$

$\square$