3. Exercise Sheet
Introduction to Machine Learning (WiSe 2020/21)

Anton Lammert
Matrikelnummer: ??????

Henrik Leisdon
Matrikelnummer: ??????

Tony Zöppig
Matrikelnummer: ??????

Bauhaus-Universität Weimar

January 28, 2021

Exercise 1

$$\sigma(x) = \frac{1}{1+e^{-x}} \tag{1}$$

(a) $\sigma(-x) = \frac{1}{1+e^{x}} = \frac{1+e^x-e^x}{1+e^x} = \frac{1+e^x}{1+e^x} - \frac{e^x}{1+e^x} = 1 - \frac{1}{\frac{1}{e^x}+1} = 1 - \frac{1}{e^{-x}+1} = 1 - \sigma(x)$

(b) $y = \sigma(x) = \frac{1}{1+e^{-x}}$
$\implies 1 + e^{-x} = \frac{1}{y}$
$\implies e^{-x} = \frac{1}{y} - 1$
$\implies -x = \ln\left(\frac{1}{y} - 1\right)$
$\implies x = -\ln(\frac{1-y}{y}) = \ln((\frac{1-y}{y})^{-1}) = \ln(\frac{y}{1-y})$

(c)

$$\frac{\partial\sigma(x)}{\partial x} = \sigma(x)\cdot(1-\sigma(x)) = \frac{1}{1+e^{-x}}\cdot(1-\frac{1}{1+e^{-x}}) = \frac{1}{1+e^{-x}} - \frac{1}{(1+e^{-x})^2}$$

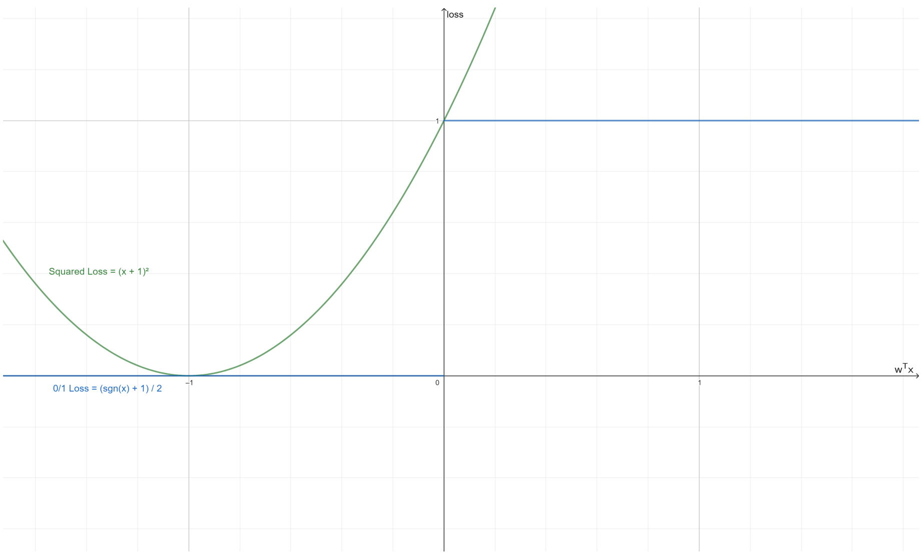
$$\frac{\partial\sigma(x)}{\partial x} = \frac{\partial(1+e^{-x})^{-1}}{\partial x} = (-1)\cdot(1+e^{-x})^{-2}\cdot(-e^{-x}) = \frac{e^{-x}}{(1+e^{-x})^2}$$

$$= \frac{1+e^{-x}}{(1+e^{-x})^2} - \frac{1}{(1+e^{-x})^2} = \frac{1}{1+e^{-x}} - \frac{1}{(1+e^{-x})^2} = \sigma(x)\cdot(1-\sigma(x))$$

Exercise 2

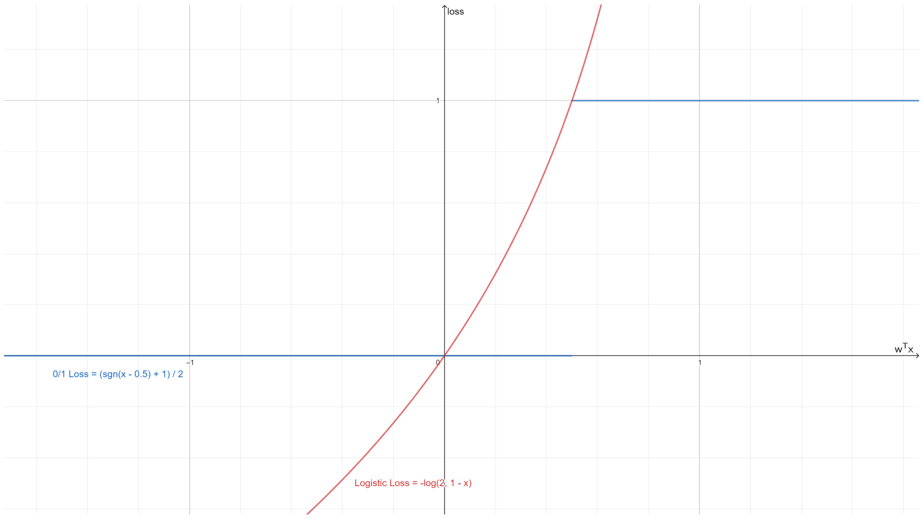
(a) $y(\mathbf{x}) = \mathbf{w}^T\mathbf{x}, l_2(y(\mathbf{x}), -1) = (y(\mathbf{x})+1)^2 = (\mathbf{w}^T\mathbf{x}+1)^2$

(b)



(c) $l_\sigma(z(\mathbf{x}), 0) = -\log(1 - z(\mathbf{x})) = -\log(1 - \mathbf{w}^T\mathbf{x})$

(d) We assume that log refers to the binary logarithm $\log_2$ and thus use it in the following.



Exercise 3

(a) The training error will be 0 since the closest point for each input from the training set is the point itself that was previously stored. Thus the classifier will always return the correct result for all training examples.

(b) Since we know that the training error of the 1NN-classifier is always 0 we know that the validation error needs to be 0.36 since the average training and validation error would not be 0.18 otherwise. This however means that the 1NN-classifier has a higher error on the validation set then the logistic regression. Because of this the logistic regression would be preferred.

Exercise 4

(a) ☐ remain constant

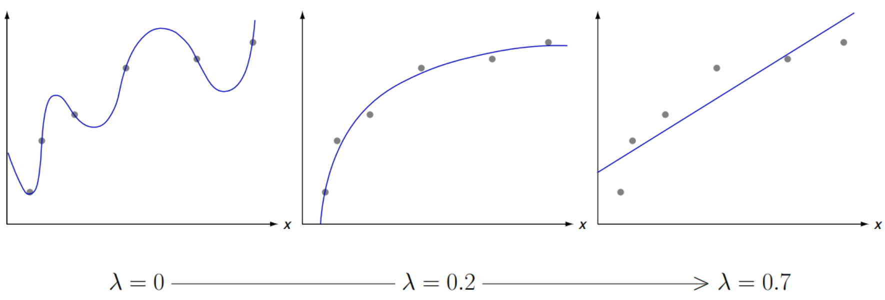
☒ steadily increase

☐ steadily decrease

☐ increase initially, then eventually start decreasing in an inverted U shape

☐ decrease initially, then start increasing in a U shape

The value of RSS: $tr(\mathbf{w})$ will steadily increase because the regularization term increases steadily and by doing so penalizes high weights $\mathbf{w}_i$. If λ is zero the classification function can fit the training data perfectly. By increasing λ it cannot fit the training data perfectly anymore since the degree of the classification function is reduced and the RSS increases steadily. This can be seen in this illustration from the lecture as well:



(b) ☐ remain constant

☐ steadily increase

☐ steadily decrease

☐ increase initially, then eventually start decreasing in an inverted U shape

☒ decrease initially, then start increasing in a U shape

The value of RSS: $tr(\mathbf{w})$ decrease initially, then start increasing in a U shape. The reason for this is that the classification fits the training data but most likely not the test data perfectly if $\lambda = 0$. By increasing λ it cannot fit the training data perfectly anymore but it can fit the test data better. If λ get larger after it fit the test data best then the RSS will increase since the regularization is now to strong and the degree of the classification function is now to low to fit the test and training data well. This can be seen in this illustration from the lecture as well: