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

Anton Lammert
Matrikelnummer: ——

Henrik Leisdon
Matrikelnummer: ——

Tony Zöppig
Matrikelnummer: ——

Bauhaus-Universität Weimar

December 4, 2020

Exercise 2

(a) **(a1)**: Objects
(a2): Feature Space
(a3): γ
(a4): α
(a5): Classes
(a6): y

(b) The role that we are playing in task 6a corresponds to γ. The function that we should implement in task 6b corresponds to α.

Exercise 3

Car	Wartburg	Moskvich	Lada	Trabi
Age (year)	5	7	15	28
Mileage (km)	30 530	90 000	159 899	270 564
Stopping distance (meter)	50	79	124	300

(a) needed: $\hat{w} = \textbf{argmin}\ \mathrm{RSS}(w)$

$$\mathrm{RSS}(w_0, w_1) = \sum_{i=1}^{n}(y_i - w_0 - w_1 \cdot x_i)^2$$
$$\text{need minimum} \rightarrow \text{second derivative} = 0$$
$$\frac{\partial}{\partial w_0}\sum_{i=1}^{n}(y_i - w_0 - w_1 \cdot x_i)^2 = 0$$
$$\Rightarrow w_0 = \bar{y} - w_1 \cdot \bar{x}$$
$$\frac{\partial}{\partial w_1}\sum_{i=1}^{n}(y_i - w_0 - w_1 \cdot x_i)^2 = 0$$
$$\Rightarrow w_1 = \frac{\sum_{i=1}^{n}(x_i-\bar{x})\cdot(y_i-\bar{y})}{\sum_{i=1}^{n}(x_i-\bar{x})^2}$$

$x-$ Age
$y-$ Stopping distance
Mean value Age: $\bar{x} = \frac{5+7+15+28}{4} = 13.75$
Mean value Stopping distance: $\bar{y} = \frac{50+79+124+300}{4} = 138.25$

$$w_1 = \frac{\sum_{i=1}^{4}(x_i-\bar{x})\cdot(y_i-\bar{y})}{\sum_{i=1}^{n}(x_i-\bar{x})^2} = 10,5868$$
$$\mathrm{w}_0 = \bar{y} - w_1 \cdot \bar{x} = -7,3185$$

The function $f(x) = -7,3185 + 10,5868 \cdot x$ minimizes the RSS.

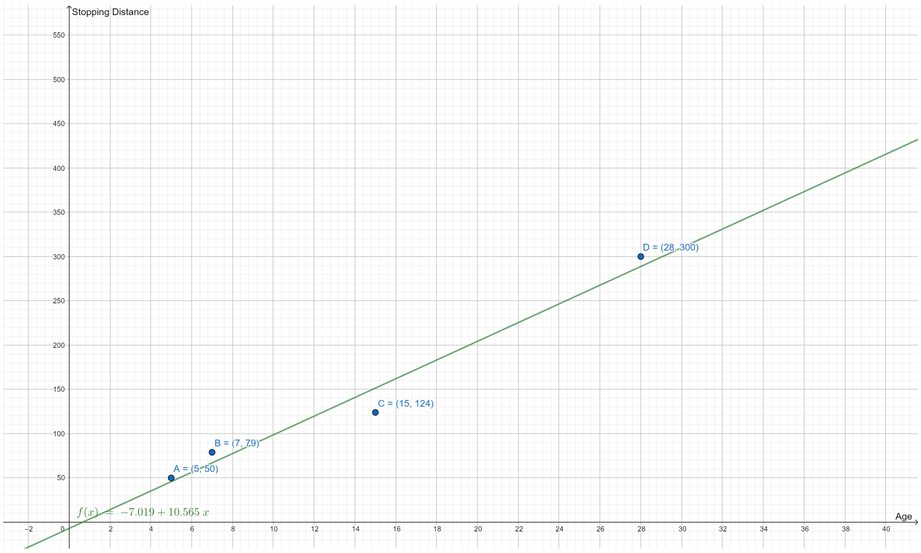
(b) $f(x) = -7,3185 + 10,5868 \cdot x$
$f(15) = 151,484$
The average stopping distance that would be extrapolated by the linear function is $151,484m$.

(c) Higher-Dimensional Feature Space:

$$\begin{aligned}
\text{RSS}(w) &= (y - Xw)^T (y - Xw) \\
\frac{\partial \text{RSS}}{\partial w} &= -2X^T (y - Xw) = 0 \\
\Rightarrow w &= (X^T X)^{-1} X^T y \\
X &= \begin{bmatrix} 1 & 5 & 30530 \\ 1 & 7 & 90000 \\ 1 & 15 & 159899 \\ 1 & 28 & 270564 \end{bmatrix} \\
\Rightarrow w &= \begin{bmatrix} -6,4263 \\ 12,8164 \\ -0,0002 \end{bmatrix}
\end{aligned}$$

$f(x, y) = -6,426 + 12,816 \cdot x - 0,0002 \cdot y$
$f(15, 159899) = 149,197$
Using the multivariable function the average stopping distance that would be extrapolated is $149,197m$.

(d) The scatterplot shows the linear regression for the age as computed in task (a).



(e) For known value-result pairs the result using the extrapolation might not be equal to the exact known result. The reason for this is that the function tries to minimize the distance to all known feature points but due to this behaviour it might not go through all points exactly. If there are data outliers they have a strong influence on the resulting extrapolation and might obscure the results. Furthermore high dimensional functions can not always be approximated well by functions of lower dimensionality which means that sometimes more variables might be needed to find a better suited approximation.

Exercise 4

(a) We have the attributes $A_1, ..., A_p$. For each attribute A_i we have one of m_i values that we can choose which results in the formula: $n(p) = \prod_{i=0}^{p} m_i$ for the number of different possible examples.

(b) Lets say each Attribute A_i has $m_i + 1$ values, where the additional value is the wildcard ?. Then the total number of hypotheses is given by: $\|H_p\| = (\prod_{i=1}^{p} (m_i + 1)) + 1$. We add 1 because of the empty hypothesis with only contradictions.

(c) Using the formula we know that: $\|H_{p+1}\| = (\prod_{i=1}^{p+1}(m_i + 1)) + 1$. In this case we can see that the previous number of hypothesis was roughly multiplied with $(m_{p+1} + 1)$ which means that we have roughly $(m_{p+1} + 1)$ times as many hypothesis as before.