Submission Date: **06.11.2024**

1 Linear Regression

Housing prices in Freiburg are through the roof! The following table shows the cost of buying a house in the Freiburg area given its age in years and its constructed area in square meters.

Age [years]	**Area** [m^2]	**Price** [€]
1	50.73	523,902.67
42	41.83	325,104.45
13	46.54	434,919.86
25	58.27	575,719.18
63	72.53	629,274.54
15	51.47	390,576.98

Using Python:

1. Use linear regression to find the weights ($\boldsymbol{w} = [w_{\text{age}}, w_{\text{area}}]$) for the age and area inputs ($\boldsymbol{x} = [x_{\text{age}}, x_{\text{area}}]$) respectively, in order to predict the price of a house ($\hat{y}$). Use the analytic solution and do not use a bias term.
2. What would the predicted cost of a house ($\hat{y}$) that was built 10 years ago and that has an area of $50.0\,\text{m}^2$ be?
3. If the real value of a 10 year old, $50.0\,\text{m}^2$ house is 427,451.10 €, what are the least-squares and L_1 losses with respect to your prediction?

2 Logistic Regression

1. Starting from negative loglikelihood (binary cross entropy loss), derive the update rule for $\boldsymbol{w}$ for gradient descent. Ignore the bias term for the moment.

   Negative loglikelihood:

   $$J = -\sum_{n=1}^{N} y_n \log p_n + (1 - y_n) \log(1 - p_n) \tag{1}$$

   with $p_n = h_{\boldsymbol{w}}(\boldsymbol{x}_n) = \mathrm{P}(y = 1 \mid X = \boldsymbol{x}_n;\ \boldsymbol{w}) = \sigma(\boldsymbol{x}_n)$
   For sigmoid use

   $$\sigma(\boldsymbol{x}_n) = \frac{\mathrm{e}^{\boldsymbol{x}_n \boldsymbol{w}}}{1 + \mathrm{e}^{\boldsymbol{x}_n \boldsymbol{w}}} \tag{2}$$

   which is equivalent to $\frac{1}{1+e^{-\boldsymbol{x}_n \boldsymbol{w}}}$ (try it).

2. What is Gradient Descent?
   Why do we use it for logistic regression?

3. Using the data from the following table, perform *one* step of gradient descent. Consider the initial parameters to be $\boldsymbol{w}_0 = [0, 0, 0]^\mathsf{T}$ and a learning rate of $\alpha = 0.25$.

x_1	x_2	y
-5	0	0
-3	-2	0
2	5	1
4	1	1

   Note: Think about a good way to include the bias in the data, instead of deriving its update rule on its own.

4. Using the learned weights $\boldsymbol{w}_1$, predict the probability
   $\mathrm{P}(y = 1 \mid X = [-1, 1]^\mathsf{T})$