Technische Hochschule Rosenheim
Prof. Johannes Jurgovsky

Exercise 4

Learning Goals

- Simple Linear Regression
- Multiple Linear Regression
- Polynomial Regression

1. ●●○ [1] **Simple Linear Regression** You are working for a high-end chocolate manufacturer and are considering the launch of a new very high-end chocolate bar. To decide on the ideal selling price, you have conducted an experiment in 6 stores and offered the new product for different prices for the same two weeks. You can find the number of products sold in the table below.

Store	1	2	3	4	5	6
Price [€]	20	16	15	16	13	10
Units sold	0	3	7	4	6	10

   (a) Draw a scatter plot showing this data.
   (b) Determine the parameters b_0 and b_1 for a Simple Linear Regression model to predict sold units based on price.
   (c) Add the regression line to the scatter plot.
   (d) Compute the Coefficient of Determination.

2. ●○○ **Multiple Linear Regression** You changed your job and are now working for a cheese manufacturer. Your product management department needs a better understanding of why certain cheese sell very well and others sell very slowly. They have identified three properties of the cheese they believe influence the units sold, all rated on a scale from 0 to 10:

   - darkness of the color of the cheese: 0=pale as the moon; 10=brown as mud
   - smelliness of the cheese: 0=not smelly at all; 10=definitely needs to be stored outside the house
   - texture of the cheese: 0=so soft and fluid it may walk away by itself;10=hard as a rock

Cheese	1	2	3	4	5	6	7	8
Color	2	8	4	10	4	7	1	3
Smelliness	4	10	3	0	6	3	1	5
Texture	5	1	5	10	6	6	1	3
Units sold	1635	261	3615	2453	523	4212	1121	2321

   You decide to use multiple linear regression to build a model based on this data. As doing the actual calculations is a lot of work, we will only express the problem so that we can use a computer to perform the calculations.

[1] Difficulty indicators: ○○○ : very easy, ●○○ : easy, ●●○ : moderate, ●●● : hard

(a) Frame this problem as a multiple regression problem in vector form $\mathbf{y} = \mathbf{X}\beta + \epsilon$.

(b) How many parameters does the model have and what are their interpretations?

3. ●●○ **Polynomial Regression** Once again you changed your job. You have gone back into academia working for a biology department. Your task is to analyze the success of cats: your research group believes that fatter cats are worse at catching mice and skinnier mice are better at avoiding cats. In addition, younger cats seem to catch more mice (maybe they are faster or have better eyes), unless they are very young and inexperienced. Your colleagues have conducted a number of experiments already, using a few cats and a lot of mice, and generated the following table showing how often a mouse managed to escape cats of certain weights.

c: Weight of Cat (kg)	2.5	2.3	2.8	1.9	3.1
m: Weight of Mouse (g)	123	232	111	265	198
a: Age of Cat (Years)	2	5	13	4	8
e: Escape Rate (count)	15	8	24	1	15

You already know that a (multiple) linear model with independent variables c, m and a does not fit the data well (this should be easy to see from the relationship between escape rate and age of the cat – first it goes down, then it goes up again), so you have decided to try a polynomial model of degree 3.

(a) State the regression function this model uses.

(b) How many parameters does this model use?

(c) How many independent variables does the multiple linear regression have that this polynomial model can be transformed into?

(d) Frame this polynomial problem as a multiple regression problem in vector form $\mathbf{y} = \mathbf{X}\beta + \epsilon$.

4. ●●○ **Simple Linear Regression** Derive the parameters b_0, b_1 of a Simple Linear Regression model for a set of examples $D = \{(x_i, y_i)\}_{i=1}^{n}$ using the method of least squares:

$$\hat{y}_i = b_0 + b_1 \cdot x_i$$

$$argmin_{b_0,b_1} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$$