Lecture “Advanced Data Analytics”

Problem Set 2

Simon Scheidegger

This problem set will give you the chance to practice the content from the week three of the lectures of this course (Pandas, linear regression, and Newton’s method).

Exercise 1a

Multiple regression is like linear regression, but with more than one independent value, meaning that we try to predict a value based on **two or more** variables.
Take a look at the data set in **data/cars.csv**, it contains some information about cars.
We want to predict the CO2 emission of a car based on the size of the engine, but with multiple regression we can throw in more variables, like the weight of the car, to make the prediction more accurate.
In Python we have modules that will do the work for us. Start by importing the Pandas module.

```
import pandas
```

The Pandas module allows us to read csv files and return a DataFrame object.
The file is meant for testing purposes only

```
df = pandas.read_csv("cars.csv")
```

Then make a list of the independent values and call this variable X.
Put the dependent values in a variable called y
X = df[['Weight','Volume']]
y = df['CO2']
We will use some methods from the sklearn module, so we will have to import that module as well:

```
from sklearn import linear_model
```

From the sklearn module we will use the LinearRegression() method to create a linear regression object.
This object has a method called fit() that takes the independent and dependent values as parameters and fills the regression object with data that describes the relationship:

```
regr = linear_model.LinearRegression()
regr.fit(X, y)
```

Now we have a regression object that are ready to predict CO2 values based on a car's weight and volume:

Task: predict the CO2 emission of a car where the weight is 2300kg, and the volume is 1300cm3.

Exercise 1b

The coefficient is a factor that describes the relationship with an unknown variable.
Example: if x is a variable, then 2x is x two times. x s the unknown variable, and the number 2 is the coefficient.
In this case, we can ask for the coefficient value of weight against CO2, and for volume against CO2. The answer(s) we get tells us what would happen if we increase, or decrease, one of the independent values.

Task: Print the coefficient values of the regression object.

Exercise 1c

The result array represents the coefficient values of weight and volume.

Weight: 0.00755095
Volume: 0.00780526

These values tell us that if the weight increase by 1kg, the CO2 emission increases by 0.00755095g.
And if the engine size (Volume) increases by 1 cm3, the CO2 emission increases by 0.00780526 g.
I think that is a fair guess, but let test it!
We have already predicted that if a car with a 1300cm3 engine weighs 2300kg, the CO2 emission will be approximately 107g.

Task: Do a prediction on the model to test what if we increase the weight with 1000kg?

Exercise 2 (Newton's Method)

Solve the following set of nonlinear equations with Newton's Method (a self-implemented version, as provided in the lecture – no python package!):

$$\begin{aligned} 5x_1^2 - x_2^2 &= 0 \\ x_2 - 0.25(\sin x_1 + \cos x_2) &= 0 \end{aligned}$$

As a starting guess, use

$$x_1^{(0)} = x_2^{(0)} = 0$$

Exercise 3 (Newton's Method)

Solve the problem with Newton's Method (a self-implemented version, as provided in the lecture – no python package!):

$\sqrt{34} = ?$

Formulate the problem as and find the root of the function $f(x) = x^2 - 34$

As initial guess: use 6.