2 Simple Linear Regression

Exercise 2.1

1. We want to use a simple linear regression to approximate a cloud of points (x_i, y_i) for $1 \leq i \leq n$. Which criterion should be minimized to obtain the best linear approximation ? Precise clearly which parameters we want to estimate with this criterion.
2. Give the closed-form expression of the slope $\hat{\beta}_1$ which is obtained by minimizing the criterion above. Deduce the closed-form expression of the Y-intercept $\hat{\beta}_0$.
3. Draw the cloud of points $P_i = (x_i, y_i)$ given by

	P_1	P_2	P_3	P_4
x_i	1	2	4	5
y_i	-2	-1	4	3

   Calculate the solution of the simple linear regression. Draw it on your scheme.

Exercise 2.2

1. Generate $n = 500$ points x_i equally spaced between 1 and 100.
2. Generate y_i such that $y_i = 2x_i + 25.2 + \varepsilon_i$ where ε_i is an univariate normal random sample with mean 0 and variance $\sigma^2 = 104.04$.
3. Plot the cloud of points (x_i, y_i).
4. Solve the linear regression problem with the function "linear_model.LinearRegression()" from the library scikit-learn :

```
from sklearn import linear_model
```

5. Plot the regression line over the cloud of points.
6. Compute and plot the residuals and the standardized residuals. Comment.
7. Compute an estimate of σ^2 based on the residuals. Comment.
8. From the output of the linear regresssion function, give the coefficient of determination. Comment.
9. Compute by yourself the t-statistics for testing the significance of the estimated slope. Comment.

   The threshold $t_{\frac{\alpha}{2}}$ can be computed as

```
talpha2=scipy.stats.t.ppf(1-alpha/2, n - 2)
```

10. Repeat all the steps for the model $y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$ where $\beta_0 = 25.2$ and $\beta_1 = 0$, then $\beta_1 = 0.05$.
11. Repeat all the steps for the model $y_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \varepsilon_i$ where $\beta_0 = 25.2$, $\beta_1 = 2$ and $\beta_2 = 0.02$. Comment the standardized residuals.