Practical work 04 – 09.10.2018
Linear Regression

Summary for the organisation :

— Submit the solutions of the practical work before Monday 12h00 next week in Moodle.
— Preferred modality : archive with iPython notebook(s).
— Alternative modality : pdf report with annotated code and outputs.
— The file name must contain the number of the practical work, followed by the names of the team members by alphabetical order, for example `02_dupont_muller_smith.zip`.
— Put also the name of the team members in the body of the notebook (or report).
— Only one submission per team.

Exercice 1 Get the data

— Get 'lausanne-appart.xlsx' from moodle
— Open a new iPython notebook
— Read columns 0 (x) and column 2 (y) from the excel file 'lausanne-appart.xlsx'
— Visualize the data

a) histogram to visualize the distribution of the renting price
b) histogram to visualize the distribution of the living area
c) scatter plot of living area as a function of renting price

Exercice 2 Normal equations for linear regression

In the case of a monovariable linear regression, we want to discover the parameters θ of a simple linear model defined with

$$h_\theta(\mathbf{x}) = \theta_0 + \theta_1 x \tag{1}$$

The "best" θ's are the ones that will minimise the squared difference between the gotten outputs and the target outputs defined with

$$J(\theta) = \frac{1}{2N} \sum_{n=1}^{N} (h_\theta(\mathbf{x}_n) - y_n)^2 \tag{2}$$

A closed form solution to this problem is given by the following normal equation :

$$\theta = (X^T X)^{-1} X^T \vec{y} \tag{3}$$

With X and $\vec{y}$ defined as

$$X = \begin{pmatrix} 1 & x_{1,1} & & \dots & & x_{1,D} \\ 1 & & \ddots & & & \\ 1 & \vdots & & x_{n,d} & & \vdots \\ 1 & & & & \ddots & \\ 1 & x_{N,1} & & \dots & & x_{N,D} \end{pmatrix} \tag{4}$$

$$y = \begin{pmatrix} y_1 \\ \vdots \\ y_N \end{pmatrix} \tag{5}$$

a) Implement Equation 3 assuming that x is the living area and y is the renting price. Use numpy for the vector operations. Plot the computed line on top of the scatter plot of exercise 1.

b) Compute the overall cost value according to Equation 2.

Exercice 3 Batch gradient descent for linear regression

Implement the batch gradient descent algorithm for the previous problem. As seen in the theory, the update rules are

$$\theta_0 \leftarrow \theta_0 - \alpha \frac{1}{N} \sum_{n=1}^{N} (h_\theta(\mathbf{x}_n) - y_n) \tag{6}$$

$$\theta_1 \leftarrow \theta_1 - \alpha \frac{1}{N} \sum_{n=1}^{N} (h_\theta(\mathbf{x}_n) - y_n) x_{n,1} \tag{7}$$

Remark You need to iterate several times over the training set. If you have problems of convergence, you need to use a smaller value of α. Values such as 0.000001 are common.

a) Plot the cost value (Equation 2) as a function of the iterations. What do you observe ?

b) Imagine a stopping criterion, i.e. when do we stop iterating on the training set ?

c) Plot the computed line $h_\theta(\mathbf{x})$ on top of the scatter plot of exercise 1.

d) Compute the final cost value according to Equation 2 and compare it to the one of exercise 2. What can you conclude ?

Exercice 4 Stochastic gradient descent for linear regression

Implement the stochastic gradient descent algorithm for the previous problem. As seen in the theory, the update rules are

$$\theta_i \leftarrow \theta_i - \alpha(h_\theta(\mathbf{x}_n) - y_n)x_{n,i} \tag{8}$$

a) Plot the computed line $h_\theta(\mathbf{x})$ on top of the scatter plot of exercise 1.

b) How many samples do you need to visit for reaching the convergence ?

c) What kind of stopping criterion could we use here ?

d) Compute the final cost value according to Equation 2 and compare it to the one of exercise 2 and 3. What can you conclude ?

Exercice 5 Optional – Mini-batch gradient descent for linear regression

Implement the mini-batch gradient descent algorithm for the previous problem, adding a parameter B defining the size of the mini-batch. Check that when $B = N$, you fall back on the batch gradient descent solution, and when $B = 1$, you get the behaviour of stochastic gradient descent.

Exercice 6 Optional – multi-variable linear regression

a) Implement one of the gradient descent algorithm (ex. 3-5) for the multi-variable linear regression assuming x_1 being the living area and x_2 the square of the living area. Plot the computed curve (second order) on top of the scatter plot of exercise 1.

b) Implement one of the gradient descent algorithm (ex. 3-5) for the multi-variable linear regression assuming x_1 being the living area and x_2 the number of bedrooms.