**Exercises Week 2**

1. For this exercise you will need to install the python package sklearn. This package contains all the machine learning functions we will use in the first part of the course.
2. Load the data in ‘adni\_adas13\_train.csv’. ADAS13 is a cognitive score which we try to predict from the other data. This is our ‘Y’. The other columns contain our features (‘X’) and are made up by volumes of brain regions (Ventricles, Hippocampus, Entorhinal, Fusiform), the participants’ AGE, brain glucose metabolism (FDG), brain amyloid burden (AV45) and a genetic marker (APOE4). You may need to specify a column number containing the index (or *rownames*) to properly load the data. Use the ‘head()’ function to see how the data was loaded. Use the pandas function ‘describe()’ to obtain an overview of the dataset. This allows us to see min/max values and percentiles for each feature. If min/max is quite different between input variables, it is a good idea to *normalize* the features.
3. Split the data frame into features (X) and response (Y).
4. Use the sklearn function StandardScaler to normalize/scale your features to 0 mean and standard deviation 1.0.
5. Train a linear regression and apply the trained model to the test data in ‘adni\_adas13\_test.csv’. Use the numpy function corrcoef to compute the correlation and the sklearn function mean\_squared\_error to compute mean squared error between predictions and observations (for this check out the ‘Regression metrics’ section of the sklearn API). In addition, make a scatter plot between observed and predicted data for the test set.
6. Repeat point Ex. 5, but now use a Ridge regression rather than a linear regression (no need to plot the scatter plot, though). Use three settings for alpha: 10000, 100, 1. How do the β coefficients change? What about MSE and the correlation?
7. Challenge: Run the Ex. 6 with alpha values from 105, 104.75, 104.5, …, 100. Plot the development of MSE and correlation depending on alpha, and also plot the development of your β coefficients depending on alpha. (Hint: for better visualization it may be beneficial to plot alpha values on a log-scale.)