DATA.ML.100 Introduction to Pattern Recognition and Machine Learning
TAU Computing Sciences
Exercise - Week 6: *Decision trees (disease dataset)*

Be prepared for the exercise sessions (watch the demo lecture). You may ask TAs to help if you cannot make your program to work, but don't expect them to show you how to start from the scratch.

1. **Disease – Scikit-Learn regressors** (40 points)

   Download data about measuring severity of a certain disease (the larger is the value the worse is the case).

   - `disease_train.txt`
   - `disease_X_test.txt`
   - `disease_y_train.txt`
   - `disease_y_test.txt`

   (a) Baseline (10 points)

   Compute regression baseline by using the training data mean value as the prediction for all test samples.

   Print baseline MSE.

   Note: You can use MSE available in the sklearn.metrics sub-package.

   (b) Linear model (10 points)

   Use `sklearn.linear_model` to fit a linear model to the data.
   Use the linear model to predict the values for the test data and print the test set MSE.

   (c) Decision tree regressor (10 points)

   Use `sklearn.tree.DecisionTreeRegressor` to fit a decision tree to your data.
   Use the model to predict the values for the test data and print the test set MSE.

   (d) Random forest regressor (10 points)

   Use `sklearn.ensemble.RandomForestRegressor` to fit a random forest to your data.
   Use the model to predict the values for the test data and print the test set MSE.

   Return the following items:

   - Python code (single file): `<surname>_disease.py`
   - A full desktop screenshot that includes a terminal window executing your code and printing each method name and the obtained MSE:
     `<surname>_disease_desktop.png`