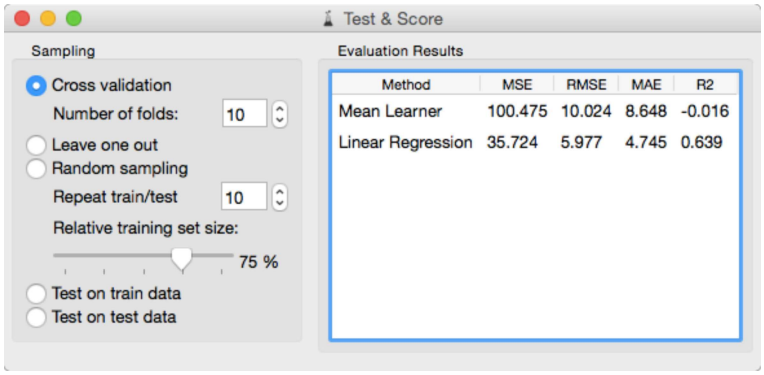
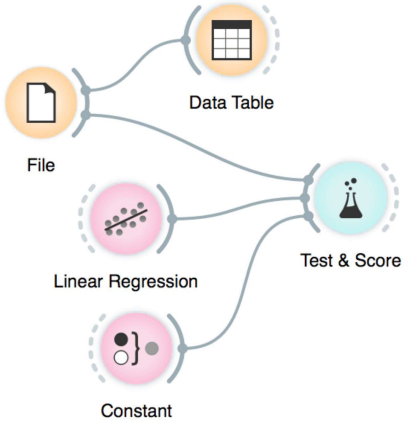
Lesson 20: Prediction of Tissue Age from Level of Methylation

Download the methylation data set from http://file.biolab.si/files/methylation.tab. Predictions of age from methylation profile were investigated by Horvath (2013) Genome Biology 14:R115.

Enough painting. Now for the real data. We will use a data set that includes human tissues from subjects at different age. The tissues were profiled by measurements of DNA methylation, a mechanism for cells to regulate the gene expression. Methylation of DNA is scarce when we are young, and gets more abundant as we age. We have prepared a data set where the degree of methylation was expressed per each gene. Let us test if we can predict the age from the methylation profile - and if we can do this better than by just predicting the average age of subjects in the training set.





Using other learners, like random forests, takes a while on this data set. But you may try to sample the features, obtain a smaller data set, and try various regression learners.

This workflow looks familiar and is similar to those for classification problems. The Test & Score widget reports on statistics we have not seen before. MAE, for one, is the mean average error. Just like for classification, we have used cross-validation, so MAE was computed only on the test data instances and averaged across 10 runs of cross validation. The results indicate that our modeling technique misses the age by about 5 years, which is a much better result than predicting by the mean age in the training set.