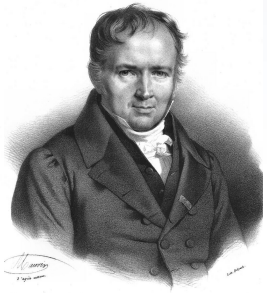
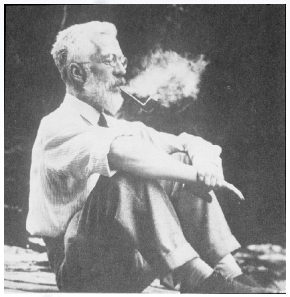
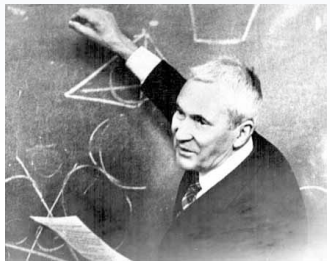
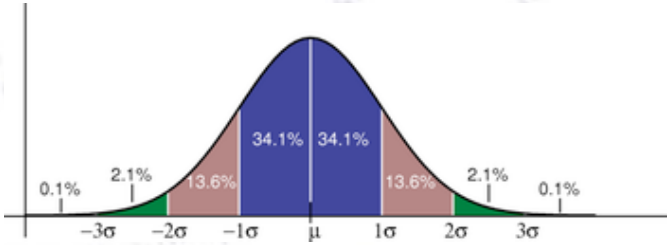
Applied ML

BDT & NN exercises in Week2



Troels C. Petersen (NBI)



"Statistics is merely a quantisation of common sense - Machine Learning is a sharpening of it!"

Exercises today

Following Week1, you should now be able to run a basic BDT and NN. We also introduced the following concepts:

- Regression vs. Classification (vs. Multi-classification).
- Loss functions (e.g. L1 vs. L2 vs. LogCosh in regression).
- Splitting data (Train, Valid, Test) and Cross Validation.
- Stochastic Gradient Descent (by using batches).

In order to make sure you understand these, today's exercise is to:

1. Make a **regression** algorithm with BDT and NN.
   Note: The HousingPrices data is hard with an NN (choose only non-NaNs).
2. Try to change between the three "classic" regression loss functions, and see what difference this makes on the resulting performance distributions (i.e. 1D: (Predicted - True) / True, 2D: Predicted vs. True).
3. Try Cross Validation on a small dataset, and calculate the performance uncertainty.
4. Try to change the batch size, and get a feel for the difference that it makes.

If you get 1+2 done, you're in good shape :-)

A note on coding…

I know that "first versions" don't look nice, but consider "cleaning up" a bit, and ensure that you've got good, simple, readable, and documented code.