Machine Learning and Data Mining

Machine Learning, seminar

Maxim Borisyak

September 15, 2016

National Research University
Higher School of Economics

Classwork

Problem 1

- load Titanic dataset;
- plot histograms of features w.r.t. target;
- train Random Forest on train dataset;
- make predictions for the model on test dataset;
- compute accuracy, precision, recall

Problem 2

- load MNIST dataset;
- select 1's and 0's;
- consider each pixel p_{ij} as a feature;
- compute and visualize mutual information:

$$I(p_{ij}, \mathrm{target})$$

- train Random Forest with the same parameters;
- evaluate accuracy, precision and recall.

Problem 3

- load MNIST dataset
- consider following model:
  - convolution with straight line with fixed length and width;
  - for each sample you optimize position and angle of the line, so it fits current sample the best.
- what are the parameters of model here?
- what are hyperparameters?
- optimize hyperparameters.
- evaluate accuracy, precision and recall.