Machine Learning and Data Mining

Introduction

Maxim Borisyak

September 1, 2016

National Research University
Higher School of Economics

Logistics

Course structure

- Lecture:
    - theoretical material.

- Seminar:
    - practice with technologies for ML (read libraries);
    - excersises;
    - Q&A;
    - student presentations.

- Excersises from seminar becomes homework.

- Course meterials will appear at **github.com/HSE-LaMBDA/Machine-Learning-Data-Mining-2016**

- Please join **course google group**

To pass this course you will need to:

- do homework (3-5 during course);
- make either:
    - Machine Learning oriented project;
    - a presentation on course related topic.

Syllabus

There are planned 3 parts of the course:

1. good, old-fashioned Machine Learning;
2. Deep Learning
3. Big Data

Good, old-fashioned Machine Learning

- Machine Learning 101;
- more or less advanced theory;
- practice with:
    - numpy,
    - scipy,
    - sklearn,
    - …

Since Deep Learning area has:

- *specific theory,*
- *has its own specific tools, libraries and sometimes requires specific hardware it forms separate part of the course.*

- minimal theory:
  - Neural Networks 101;
  - optimization 101;
- symbolic computations (theano);
- massively parallel computations (theano on GPU);
- Deep Learning.

Big Data

*We will read "Big Data" as **distributed algorithms for Machine Learning and Data Mining**.*

- distributed computations (Spark);
- SGD, again ...;
- recommendation algorithms.

I am not fully aware of your background. Thus I am asking you to particapate in a survey so I can make some adjestments.

Summary

Summary

- lecture + seminar;
- 3-5 homeworks + either project or presentation;
- 3 parts:
    - "Shallow" Learning,
    - Deep Learning,
    - "Big" Learning;
- you can influence course structure via the survey.