Week 1

How to Win a Data Science Competition: Learn from Top Kagglers

Introduction & Recap

This week we will introduce you to competitive data science. You will learn about competitions' mechanics, the difference between competitions and a real life data science, hardware and software that people usually use in competitions. We will also briefly recap major ML models frequently used in competitions.

Less

Key Concepts

- Describe competition mechanics
- Compare real life applications and competitions
- Summarize reasons to participate in data science competitions
- Describe main types of ML algorithms
- Describe typical hardware and software requirements
- Analyze decision boundaries of different classifiers
- Use standard ML libraries

Less

Welcome to "How to win a data science competition"

- **Video:** About the University 1 min
- **Reading:** About the University 10 min
- **Video:** Introduction 1 min
- Pre-survey on HSE online courses 10 min
- **Reading:** Welcome! 10 min
- **Video:** Meet your lecturers 2 min
- **Video:** Course overview 7 min
- **Reading:** Week 1 overview 10 min

Competition mechanics

- **Video:** Competition Mechanics 6 min
- **Video:** Kaggle Overview [screencast] 7 min
- **Video:** Real World Application vs Competitions 5 min
- **Practice Quiz:** Practice Quiz 4 questions

Recap of main ML algorithms

- **Video:** Recap of main ML algorithms 9 min
- **Reading:** Disclaimer 10 min

- **Practice Quiz:** Recap 4 questions
- **Quiz:** Recap 6 questions
- **Reading:** Explanation for quiz questions 10 min
- **Notebook:** Will performance of GBDT model drop dramatically if we remove the first tree? 15 min
- **Reading:** Additional Materials and Links 10 min

Software/Hardware requirements

- **Video:** Software/Hardware Requirements 5 min — Resume
- **Practice Quiz:** Software/Hardware 3 questions
- **Notebook:** Pandas basics
- **Programming Assignment:** Pandas basics 3h
- **Quiz:** Graded Soft/Hard Quiz 4 questions
- **Reading:** Explanation for quiz questions 10 min
- **Reading:** Additional Material and Links 10 min

Feature Preprocessing and Generation with Respect to Models

In this module we will summarize approaches to work with features: preprocessing, generation and extraction. We will see, that the choice of the machine learning model impacts both preprocessing we apply to the features and our approach to generation of new ones. We will also discuss feature extraction from text with Bag Of Words and Word2vec, and feature extraction from images with Convolution Neural Networks.

Less

Key Concepts

- Explain how employed model impacts choice of preprocessing
- Summarize feature preprocessings for numeric and categorical features
- Summarize feature generation approaches for datetime and coordinates
- Summarize approaches to deal with missing values
- Outline the pipeline of applying Bag of Words
- Compare Bag of Words and Word2vec
- Explain how to extract CNN descriptors from images

Less

Feature preprocessing and generation with respect to models

- **Video:** Overview 6 min — Resume
- **Video:** Numeric features 13 min
- **Video:** Categorical and ordinal features 10 min
- **Video:** Datetime and coordinates 8 min
- **Video:** Handling missing values 10 min
- **Practice Quiz:** Feature preprocessing and generation with respect to models 4 questions

Quiz: Feature preprocessing and generation with respect to models 4 questions

Reading: Explanation for quiz questions 10 min

Reading: Additional Material and Links 10 min

Feature extraction from text and images

Video: Bag of words 10 min

Video: Word2vec, CNN 13 min

Practice Quiz: Feature extraction from text and images 4 questions

Quiz: Feature extraction from text and images 4 questions Due Sep 28, 1:59 AM CDT

Reading: Explanation for quiz questions 10 min

Reading: Additional Material and Links 10 min

Final Project Description

This is just a reminder, that the final project in this course is better to start soon! The final project is in fact a competition, in this module you can find an information about it.

Key Concepts

- Analyze the final project task and requirements to compete it

More

Final project

Reading: Final project 10 min

Video: Final project overview 4 min Resume

Discussion Prompt: Meet and Greet 5 min

Reading: Final project advice #1 10 min

Survey

Survey 10 min