



How to Win a Data Science Competition: Learn from Top Kagglers
National Research University
Higher School of Economics

Overview

Week 1

Week 2

Week 3

Week 4

Week 5

Grades

Notes

Discussion Forums

Messages

Resources

Course Info

Week 4

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

Week 4

Discuss this week's modules here.
96 threads · Last post a day ago

Go to forum

Hyperparameter Optimization



In this module we will talk about hyperparameter optimization process. We will also have a special video with practical tips and tricks, recorded by four instructors.

Key Concepts

- List most important hyperparameters in major models; describe their impact
- Understand the hyperparameter tuning process in general
- Arrange hyperparameters by their importance

Less

Hyperparameter tuning

Reading: Week 4 overview 10 min

Video: Hyperparameter tuning I 7 min

Resume

Video: Hyperparameter tuning II 12 min

Notebook: How to find sufficient `n_estimators` in Random Forest 10 min

Video: Hyperparameter tuning III 13 min

Practice Quiz: Practice quiz 3 questions

Quiz: Graded quiz 4 questions Due Oct 19, 1:59 AM CDT

Reading: Comments on quiz 10 min

Reading: Additional material and links 10 min

Tips and tricks

Video: Practical guide 16 min

Video: KazAnova's competition pipeline, part 1 18 min

Video: KazAnova's competition pipeline, part 2 17 min

Notebook: How to use macros in Jupyter 10 min

Reading: Additional materials and links 10 min

Advanced feature engineering II



In this module we will learn about a few more advanced feature engineering techniques.

Key Concepts

- Understand matrix factorization technique for feature extraction
- Create features based on tSNE
- Explain concept of feature interactions
- Generate and make up new features based on statistics and nearest neighbors

Less

Advanced features II

▶ **Video:** Statistics and distance based features 5 min

Resume

▶ **Video:** Matrix factorizations 6 min

▶ **Video:** Feature Interactions 5 min

▶ **Video:** t-SNE 5 min

📋 **Quiz:** Graded Advanced Features II Quiz 6 questions Due Oct 19, 1:59 AM CDT

📖 **Reading:** Comments on quiz 10 min

📖 **Reading:** Additional Materials and Links 10 min

Optional Honors Content ?

Advanced features II programming assignment

📓 **Notebook:** KNN features implementation

📓 **Honors Programming Assignment:** KNN features implementation 3h Due Oct 19, 1:59 AM CDT

Ensembling



Nowadays it is hard to find a competition won by a single model! Every winning solution incorporates ensembles of models. In this module we will talk about the main ensembling techniques in general, and, of course, how it is better to ensemble the models in practice.

⤴ [Less](#)

Key Concepts

- Describe and compare ensembling methods
- List and describe plan of validation schemes for ensembling
- Compare validation schemes
- Successfully implement ensembling in practical task

⤴ [Less](#)

Ensembling

▶ **Video:** Introduction into ensemble methods 6 min

Resume

▶ **Video:** Bagging 9 min

▶ **Video:** Boosting 16 min

▶ **Video:** Stacking 16 min

▶ **Video:** StackNet 14 min

▶ **Video:** Ensembling Tips and Tricks 14 min

📋 **Reading:** Validation schemes for 2-nd level models 10 min

📓 **Notebook:** Ensembling implementation notebook

📓 **Programming Assignment:** Ensembling implementation 3h Due Oct 19, 1:59 AM CDT

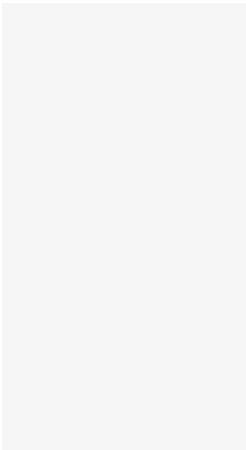
📋 **Practice Quiz:** Ensembling 4 questions



📋 **Quiz:** Ensembling 6 questions Due Oct 19, 1:59 AM CDT

📖 **Reading:** Comments on quiz 10 min




📖 **Reading:** Additional materials and links 10 min


📖 **Reading:** Final project advice #4 10 min



 Optional Honors Content 

Catboost

-  **Video:** CatBoost 1 7 min
-  **Video:** CatBoost 2 7 min
-  **Honors Programming Assignment:** CatBoost basics 3h

Due Oct 19, 1:59 AM CDT
-  **Notebook:** CatBoost notebook 1h

