Hyperparameter Optimization using Hyperopt

Yassine Alouini - Paul Coursaux
03/11/2016

@qucit

@YassineAlouini

About us

Yassine

- Data Scientist @ Qucit
- Centrale Paris & Cambridge
- Quora's Top Writer 2016

Paul

- Data Scientist @ Qucit
- Centrale Paris
- Market finance in London
- Horse riding

Outline

1. Hyperparameters in Machine Learning
2. How to Choose Hyperparameters ?
3. Tree-structured Parzen Estimation Approach
4. Live-coding Example

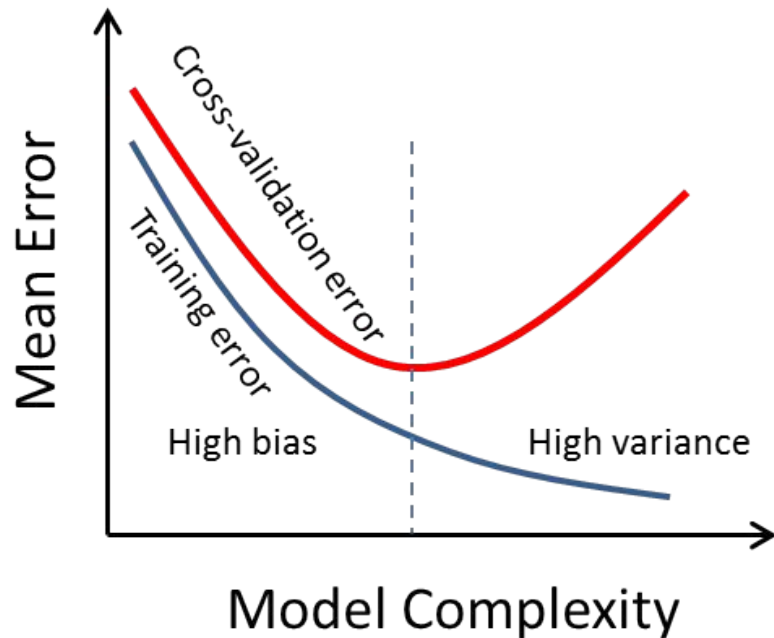# 1. Hyperparameters in Machine Learning

What are hyperparameters ?

Parameters:

Rent = $a_1 \times$ surface +

$a_2 \times$ distance to city center +

...

Hyperparameters:

$RMSE_{LASSO}$ = RMSE +
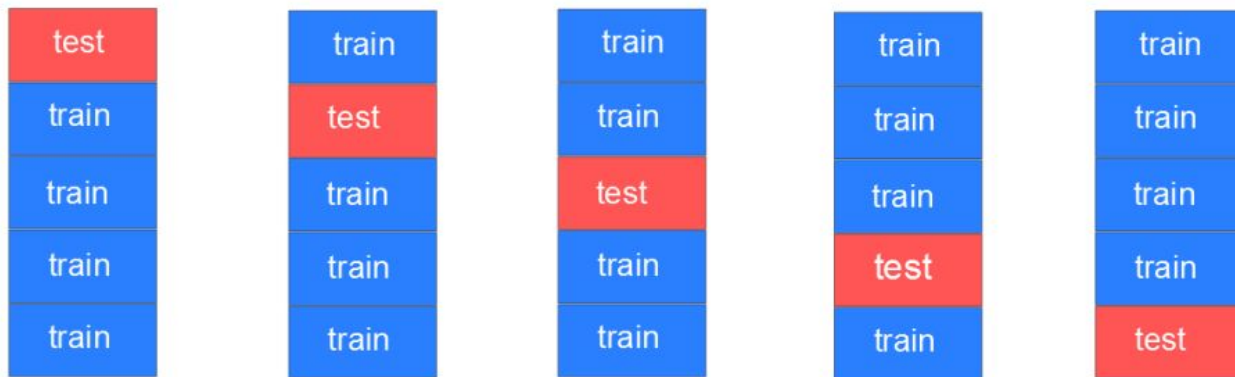
$\alpha \times (|a_1| + ...)$

The impact of hyperparameters

2. How to choose hyperparameters ?

Cross validation

Enable to choose the hyperparameter(s) with the best generalization capabilities making an efficient use of the data
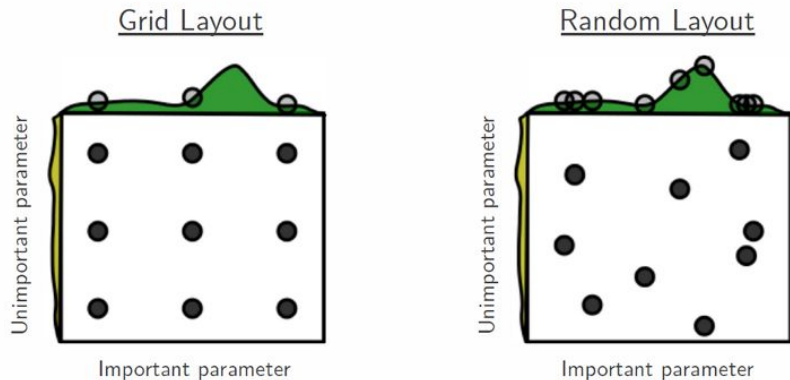
How to choose the points to cross-validate?
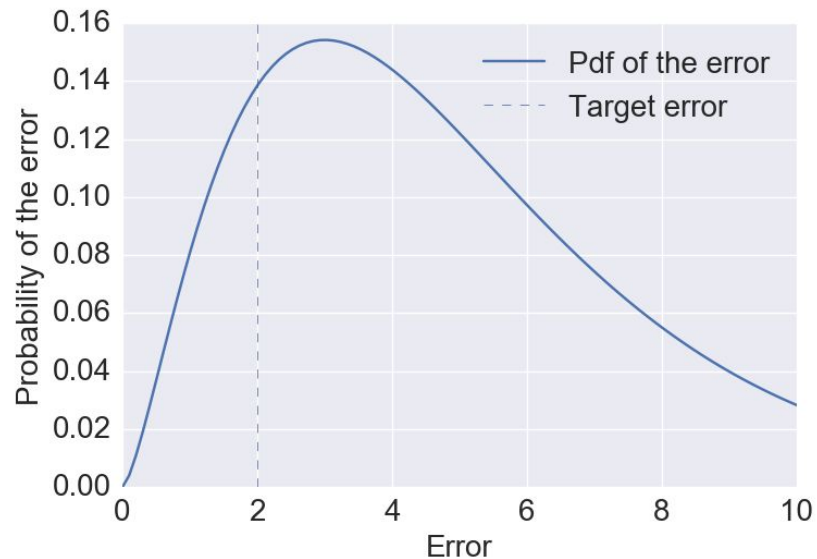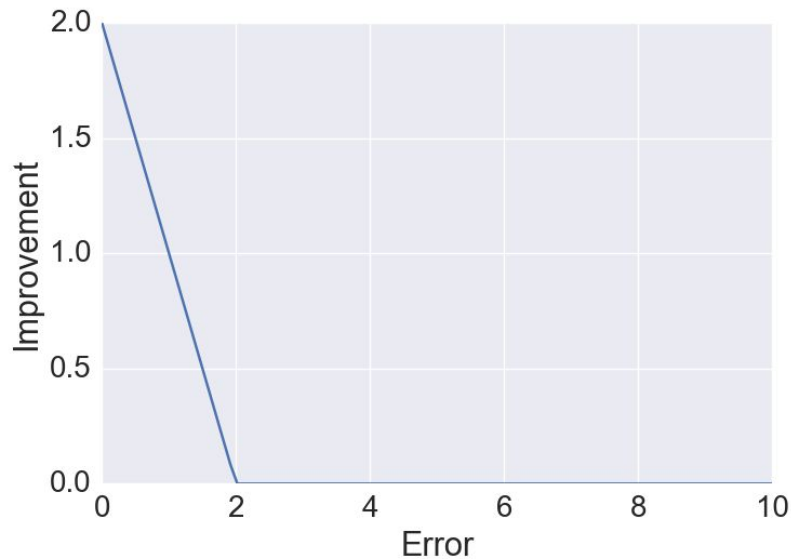
Grid search

Random search

3. Tree-structured Parzen Estimation Approach

Sequential Model-based Global Optimization

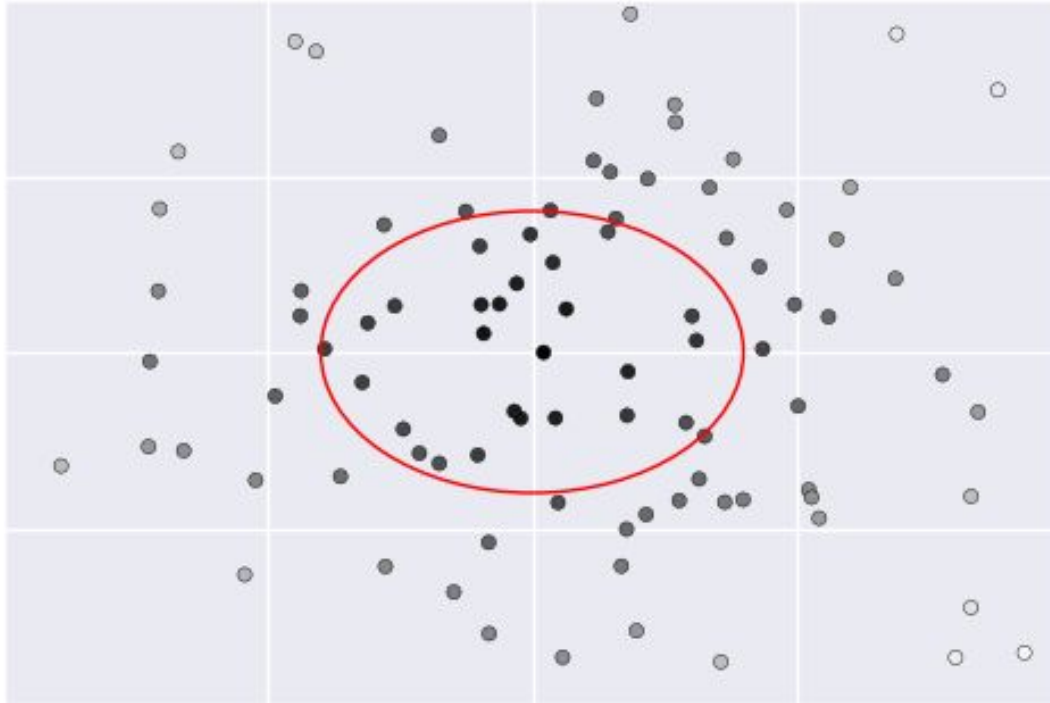# The Expected Improvement

$$EI_{\varepsilon*}(a) = \int max(\varepsilon* - \varepsilon, 0)p_M(\varepsilon|a)d\varepsilon$$
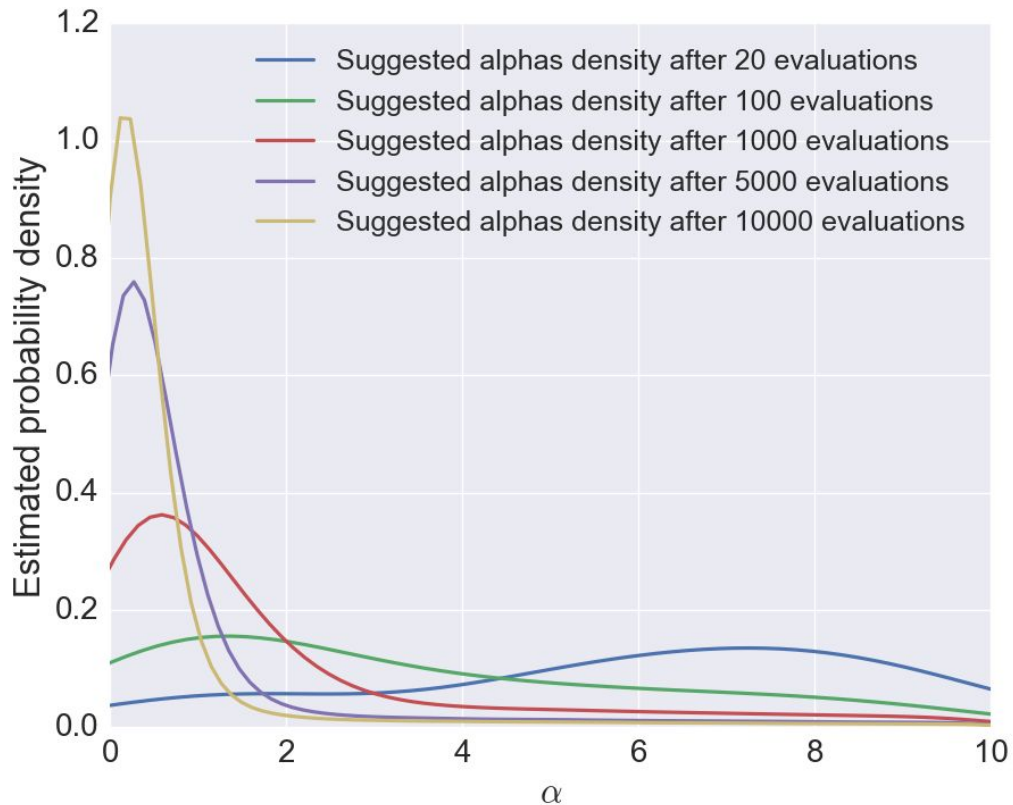
How to Optimize the EI ? (1)
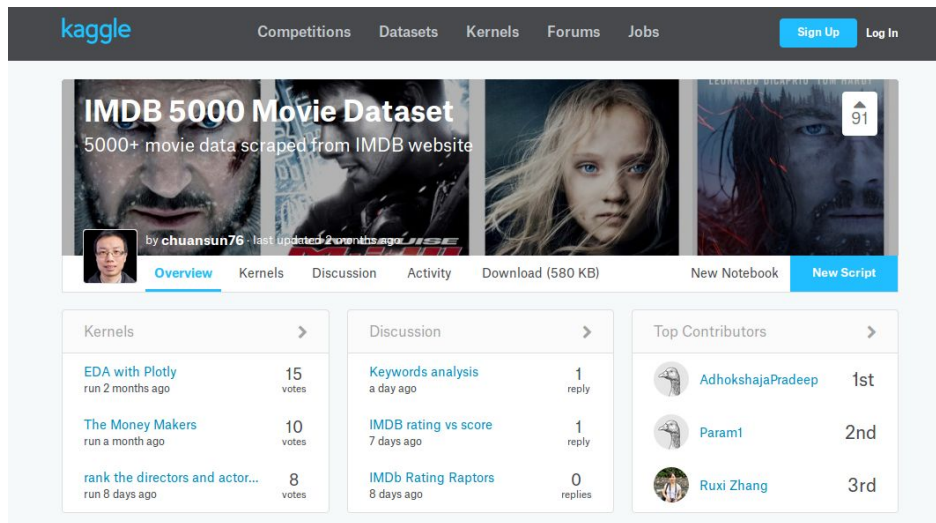
How to Optimize the EI ? (2)

- Lasso model on the Boston Housing Dataset
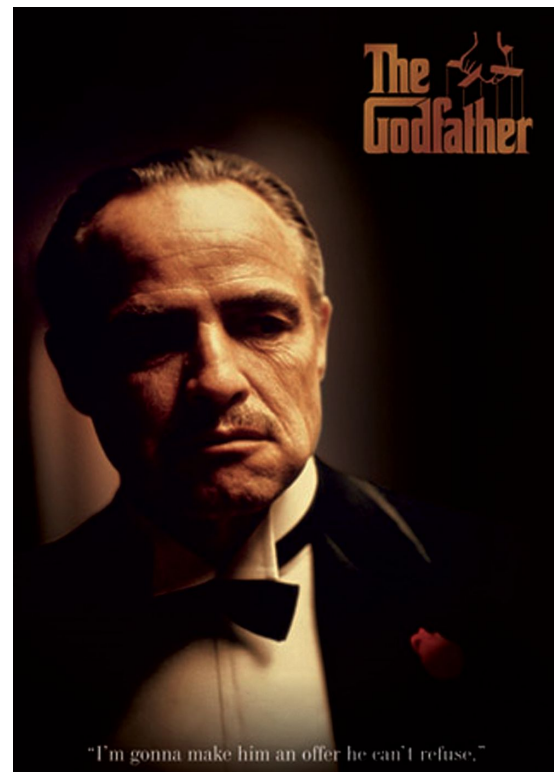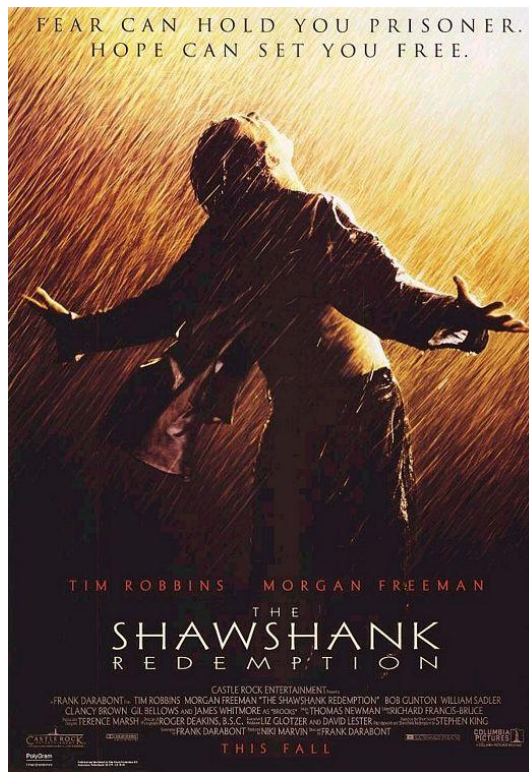- Distribution of the suggested αs

4. Live-coding Example

Description of the dataset

- IMDb dataset
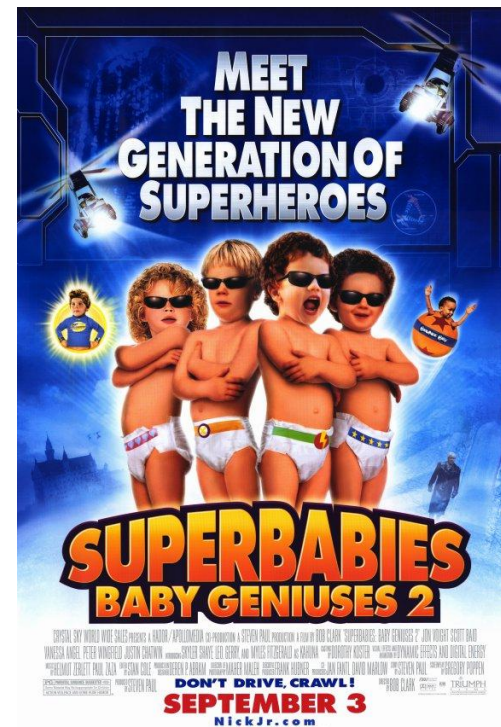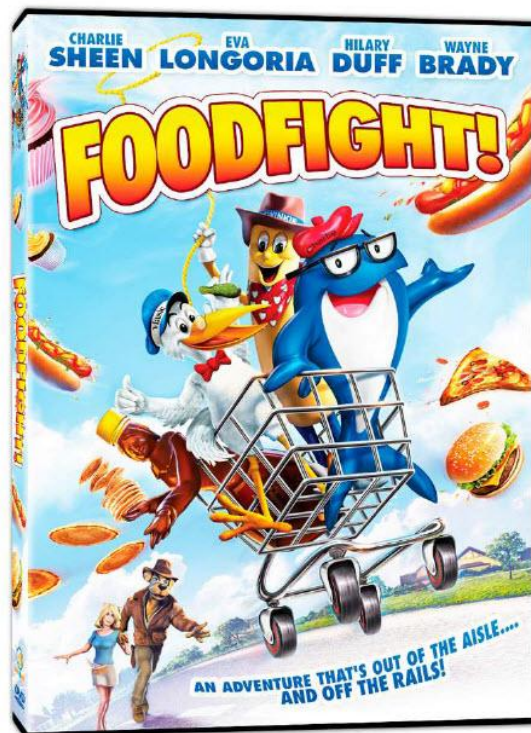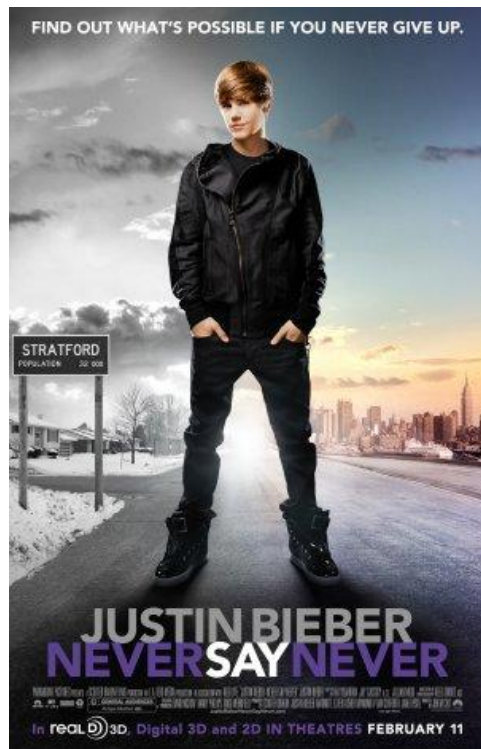- Dataset publicly available (from Kaggle)
- Contains 5043 samples

Credits: screenshot, 24/10/2016, https://www.kaggle.com/deepmatrix/imdb-5000-movie-dataset

Movies having the best score

Movies having the worst score

Task

- Predict the IMDB movie score

- Gradient Boosting algorithm (XGBoost package)

- 3 hyperparameters optimization strategies
  - A *naive* grid search
  - An *expert* grid search (*)
  - The TPE algorithm (hyperopt package)

(*) http://blog.kaggle.com/2016/07/21/approaching-almost-any-machine-learning-problem-abhishek-thakur/
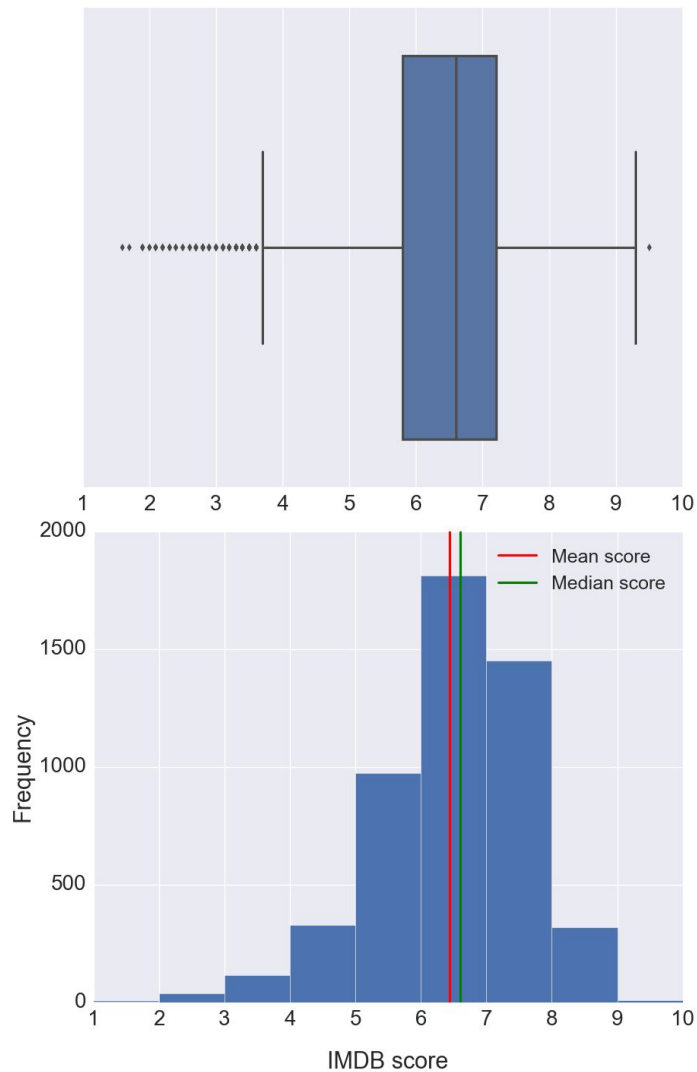
Features description

- 28 features:
  - 14 movie-related
  - 4 review-related
  - 10 cast-related
- 16 kept:
  - 11 numerical
  - 5 categorical
- 12 removed
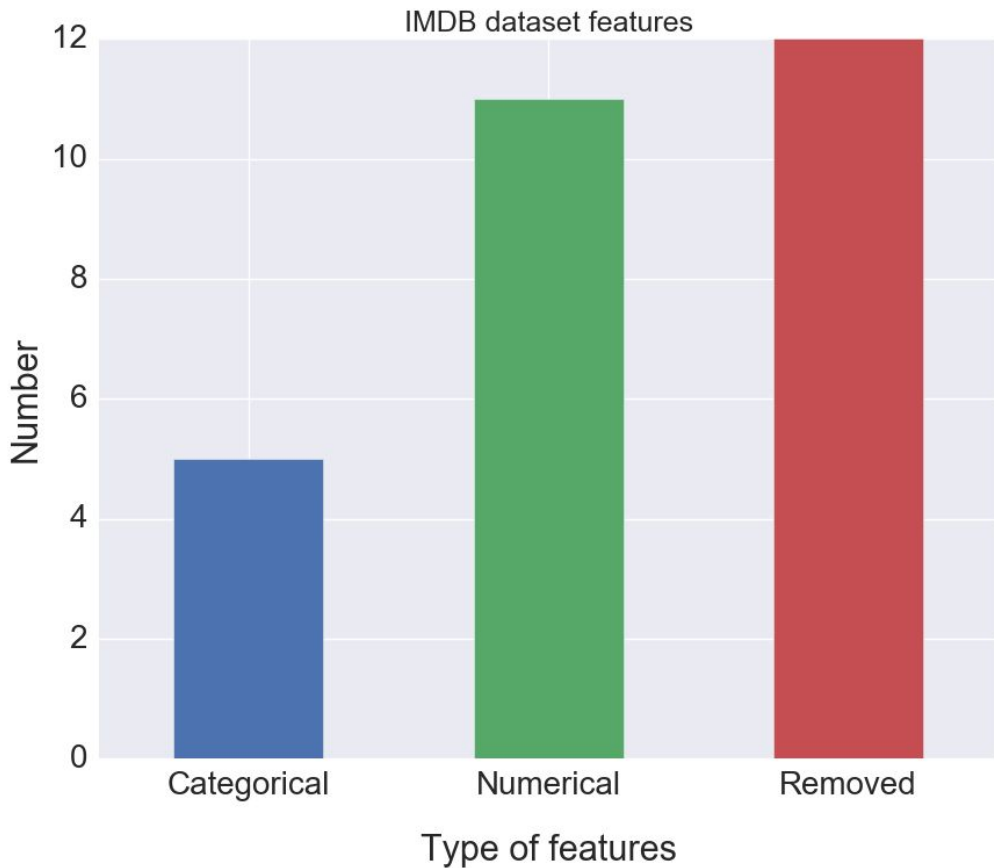
Live demo

Our code is available here:
https://github.com/yassineAlouini/
hyperparameters-optimization-talk

Conclusion

- Outperforms the standard methods in most cases
- Search space matters
- Other Python libraries: Spearmint, BayesOpt, Scikit-Optimize
- Distributed optimization (using MongoDB)

Thanks for your attention.
Question time

Qucit is hiring!

References

- https://papers.nips.cc/paper/4443-algorithms-for-hyper-parameter-optimization.pdf
- https://conference.scipy.org/proceedings/scipy2013/pdfs/bergstra_hyperopt.pdf
- https://github.com/scikit-optimize
- http://jaberg.github.io/hyperopt/
- https://github.com/JasperSnoek/spearmint
- https://github.com/fmfn/BayesianOptimization
- http://xgboost.readthedocs.io/en/latest/
- http://www.cs.ubc.ca/~hutter/papers/13-BayesOpt_EmpiricalFoundation.pdf