

Hillsboro Python Machine Learning Meetup

Mar/2019

Ernest Bonat, Ph.D.

Senior Software Engineer

Senior Data Scientist

DAT Wi-Fi

Username: **DAT Guest**

Password: **beaverton dat**

- 6:00 – 6:40 pm: Pizza, **water only** and networking.
- 6:40 – 6:45 pm: Welcome message by Ernest Bonat, Ph.D.
- 6:45 – 8:00 pm: Presentation and open discussions.
- 8.00 pm – 9.00 pm: Coding and learning session. Bring your Python development laptop!

Why did I create this meetup?

1. Bad traffic to Portland downtown.
2. Very hard to find a parking lot.
3. Bad Python presentation code and old used Python tools.

Our Meetup Mission:

1. *“Come, Listen, Code and Learn”.*
2. Finding and presenting best practices of Machine Learning using Python Data Ecosystem.
3. Create great networking place for Hillsboro-Beaverton Data Scientists.

Breast Cancer Image Processing – TODO Project!

E:\Visual WWW\Python\07 PYTHON USER GROUPS
MEETUPS\Hillsboro Python Machine Learning Meetup\Breast Cancer
Image Processing Meetup Project

Today Presentation

“Fast Machine Learning Hyperparameters Model
Optimization using Hyperopt”

Hyperopt (**Distributed Asynchronous Hyper-parameter Optimization**) - is a Python library for optimizing over awkward search spaces with real-valued, discrete, and conditional dimensions.

GitHub Repository: <https://github.com/hyperopt/hyperopt>

Documentation: <http://hyperopt.github.io/hyperopt>

Why Hyperopt?

CSV data file with 20,376 rows.

```
# define xgb regressor model
ml_model = XGBRegressor()

# define hyper parameter candidates
hyper_parameter_candidates = {"nthread":[4], "objective":["reg:linear"],
                              "learning_rate":[.01, 0.5, .08], "max_depth":[5, 6, 7, 10],
                              "min_child_weight":[1, 5, 10], "silent":[1],
                              "subsample":[0.7], "colsample_bytree":[0.5, 0.6, 0.7, 0.8, 0.9, 1],
                              "n_estimators":[100, 500, 1000], "gamma":[0.1, 0.5, 0.5, 1, 2, 5]}
scoring_parameter = "r2"
cv_folds = KFold(n_splits=10, shuffle=True, random_state=1)
```

Best Score: 0.918 (very good score!)

Best Parameters: {'colsample_bytree': 0.6, 'gamma': 1, 'learning_rate': 0.08, 'max_depth': 10, 'min_child_weight': 1, 'n_estimators': 1000, 'nthread': 4, 'objective': 'reg:linear', 'silent': 1, 'subsample': 0.7}

Program Runtime: 33 hours!

Good blogs to read:

1. “On Using Hyperopt: Advanced Machine Learning”

<https://blog.goodaudience.com/on-using-hyperopt-advanced-machine-learning-a2dde2ccec7>

2. “Parameter Tuning with Hyperopt”

<https://medium.com/district-data-labs/parameter-tuning-with-hyperopt-faa86acdfdce>

3. “hyperopt-sklearn - Hyper-parameter optimization for scikit-learn”

<https://github.com/hyperopt/hyperopt-sklearn>

NeuPy - Neural Networks in Python

<http://neupy.com/pages/home.html>

Tree-structured Parzen Estimators (TPE)

http://neupy.com/2016/12/17/hyperparameter_optimization_for_neural_networks.html#id24

http://neupy.com/2016/12/17/hyperparameter_optimization_for_neural_networks.html#tree-structured-parzen-estimators-tpe

Basic Tutorial and Functions Explanation

<https://github.com/jaberg/hyperopt/wiki/FMin>