Hillsboro Python Machine Learning Meetup

May/2018

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.

Today Presentation

“Advanced XGBoost Algorithm for Data Classification and Regression”

Ernest Bonat, Ph.D.

XGBoost - eXtreme Gradient Boosting

It's an implementation of gradient boosted decision trees designed for speed and performance. It is an implementation of gradient boosting machines created by Tianqi Chen ("Introduction to Boosted Trees" slides). Written in C++.

https://homes.cs.washington.edu/~tqchen/pdf/BoostedTree.pdf

It is capable of performing the three main forms of gradient boosting (Gradient Boosting (GB), Stochastic GB and Regularized GB) and it is robust enough to support fine tuning and addition of regularization parameters.

Documentation

http://xgboost.readthedocs.io/en/latest/

(Flexible, Portable, Multiple Languages, Battle-tested, Distributed on Cloud, Performance)

Excellent Blog Papers

1. “XGBoost, a Top Machine Learning Method on Kaggle, Explained”

(https://www.kdnuggets.com/2017/10/xgboost-top-machine-learning-method-kaggle-explained.html)

2. “A Simple XGBoost Tutorial Using the Iris Dataset”

(https://www.kdnuggets.com/2017/03/simple-xgboost-tutorial-iris-dataset.html)

GitHub Repo on Distributed Machine Learning Community (DMLC)

(https://github.com/dmlc/xgboost)

Introduction to Boosted Trees – Math Explanation (read more and understand it!)

(https://xgboost.readthedocs.io/en/latest/model.html)

XGBoost Windows 10 Installation

1. For 3.6.2 | Anaconda custom (64-bit), we need the wheel file from https://www.lfd.uci.edu/~gohlke/pythonlibs/
2. xgboost-0.71-cp36-cp36m-win_amd64.whl
3. PS cd C:\Users\Ernest\AppData\Local\Continuum\Scripts
4. .\pip install C:\xgboost-0.71-cp36-cp36m-win_amd64.whl
5. Results:
   - Installing collected packages: xgboost
   - Successfully installed xgboost-0.71

6. Check version:

```
import xgboost as xgb
print(xgb.__version__)
```

0.71

XGBoost Python Feature Walkthrough

(https://github.com/dmlc/xgboost/tree/master/demo/guide-python)

Complete Guide to Parameter Tuning in XGBoost (with codes in Python)

(https://www.analyticsvidhya.com/blog/2016/03/complete-guide-parameter-tuning-xgboost-with-codes-python/)

XGBoost Data Classification and Regression Examples

Get the source code from:
https://github.com/ebonat/hillsboro_machine_learning_05_2018

Trying Main Classification Algorithms:

1. Extreme Gradient Boosting (XGBoost)
2. Random Forests (RF)
3. Artificial Neural Network (ANN)