

Assignment 2 – Income Prediction

Let's do the BINARY CLASSIFICATION !!!

FAQ

- Q1.請問我可以在github上傳自己的weight並在shell script中使用嗎?
可以，要是同學擔心自己在hw2_best.sh中使用的方法，training的時間可能會超過時限，可以上傳weight，我們到時候clone下來只要確定同學的code可以infer就好。
- Q2.請問Report.pdf名稱與HW1規定不一致，是大小寫都可以用嗎？
請同學依照HW2中投影片的名稱繳交。

Announcement

- Kaggle Deadline: 2017/10/26 11:59 P.M. (GMT+8)
- Github Deadline: 2017/10/27 11:59 P.M. (GMT+8)
- TA會於10/20釋出範例程式碼，亦為超過 Kaggle simple baseline 的加分截止日期
- (10/13) 整理同學課堂問題，請見最上面的FAQ
- (10/17) 新增hw2_best.sh可以使用的套件，請見下方Rules
- (10/20) 上傳samplecode，請見Requirements裡的連結，或是Link的Github連結。
- (10/27) 上傳answer，請見Dataset部分

Link

- 投影片連結 
 - Kaggle 連結 
 - 遲交表單 
 - 小老師申請表單 
 - Github 
 - TA Hour Slide 
-

Requirements

In this assignment, you are asked to implement the following two models.

1. Logistic Regression

Handcrafted Gradient Descent Optimizer to solve logistic regression.

2. Probabilistic Generative Model

Implement probabilistic generative model to do binary classification.

Dataset : ADULT Dataset

Raw Data

- [train.csv](#)
- [test.csv](#)
- [Ref](#)

Feature

- [X_train](#)
 - 106 dims
 - one-hot encoding
- [Y_train](#)
 - label = 0 表示小於等於50K, label = 1 表示大於50K
- [X_test](#)

Ans

- [Answer](#)
-

Rules

- Basic : No extra dataset.
- hw2_logistic.sh, hw2_generative.sh : Only toolkits below are allowed to use.
 - Python 3.5+
 - Numpy
 - Pandas
 - Python Standard Lib
- hw2_best.sh : 除了上述套件, 可以使用以下套件, 若有其他想用的請再來信詢問TA是否可以使用。

- Tensorflow 1.3
- Keras 2.0.8
- Pytorch 0.2.0
- Scikit-learn 0.19.0
- XGBoost 0.6
- h5py 2.7.0



🔄 Posted by: [ntumlta](#)

✉ Contact information: ntu.mlta@gmail.com .

🎓 Course information: [Machine Learning \(2017, Fall\) @ National Taiwan University](#).