

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 https://docs.google.com/presentation/d/1qMxjDGkS6fVY0LtMh_VJFhI9QNk0GtiEJegxw3Eq3Ok/edit

- Kaggle 連結 <https://www.kaggle.com/account/login?ReturnUrl=%2Ft%2F5808de7e75cf4e509d28d014a9f36f7d>
 - 遲交表單（网络课程不提供）
 - 小老師申請表單（网络课程不提供）
 - Github <https://github.com/ntumlta/2017fall-ml-hw2/tree/master>
 - TA Hour Slide（网络课程不提供）
-

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](#).