

Assignment 1 - Predicting PM2.5

In this assignment, you will practice using Gradient Descent to predict PM2.5.

Announcement

10/13

- report 中 public + private 分數的意思是：testing 中 240 筆的 RMSE，也就是 $\text{square root}\{[(\text{public})^2 + (\text{private})^2]/2\}$
- [testing 答案](#) 釋出！
- hw1_best.sh: 選擇的 2 筆 kaggle 分數中，private 較好的那一個 model (註：不用完全一模一樣重現，可接受誤差+0.2)
 - ex. (1) public:6.5/private:5.5 (2) public:6.1/private:5.7
 - 滿足 hw1_best.sh 的 private 分數 $< 5.5 + 0.2$ 即可
- hw1.sh: public 分數 $>$ public simple baseline 即可
- kaggle 的成績：只要 2 筆中任何 1 筆 $>$ 某 baseline，即得該 baseline 成績
- github 死線：今晚午夜

10/6

- 10/5 有通過 public simple baseline 名單 <https://docs.google.com/spreadsheets/d/1rrGrkrRp0tt2p7xukT4wqz8Jr-0qkJeqk1zy6z6N32E/edit#gid=0>
 - 因 **kaggle** 名稱...等等原因沒有被登記到的同學，請填寫（不提供给网络课程）

重要連結

- 投影片連結 https://docs.google.com/presentation/d/1JAaVzT8UTShojE385a-wGzRe8WsJpQE0FhWQGc_Q-L0/edit
- 老師講解投影片 [http://speech.ee.ntu.edu.tw/~tlkagk/courses/ML_2017_2/Lecture/HW1%20\(v2\).pdf](http://speech.ee.ntu.edu.tw/~tlkagk/courses/ML_2017_2/Lecture/HW1%20(v2).pdf)
- Kaggle 連結 <https://www.kaggle.com/c/ml-2017fall-hw1>
- Github Repo 表單（不提供给网络课程）
- report template（不提供给网络课程）
- 遲交表單（不提供给网络课程）

- 10/6(五)上課時間，助教會釋出 hw1 的 [Sample Code](#) 以及 [Supplementary Slide](#)，同時會邀請小老師幫大家解決程式問題。
- (已截止)小老師教學表單

The requirements of this assignment are as follows:

- hw1.sh
 - **Python3.5+** required
 - Only (1)numpy (2)scipy (3)pandas are allowed
 - numpy.linalg.lstsq is forbidden.
 - Please handcraft "linear regression" using **Gradient Descent**
 - beat public simple baseline
 - For those who wish to load model instead of running whole training process:
 - please upload your training code named **train.py**
 - as long as there are Gradient Descent Code in **train.py**, it's fine
- hw1_best.sh
 - **Python3.5+** required
 - any library is allowed
 - meet the higher score you choose in kaggle

Data 簡介

- [下載 train.csv](#)：每個月前 20 天每個小時的氣象資料(每小時有 18 種測資)。共 12 個月。
- [下載 test.csv](#)：排除 train.csv 中剩餘的資料，取連續 9 小時的資料當 feature，預測第 10 小時的 PM2.5 值。總共取 240 筆不重複的 test data。
- [下載 sampleSubmission.csv](#)

作業修正&講解

- report 第五題題目修正： $X = [x^1 \ x^2 \ \dots \ x^N]$ 改為 $X = [x^1 \ x^2 \ \dots \ x^N]^T$
- 第 1-3 題請都以題目給訂的兩種 model 來回答

FAQ

Q1. 為了回答 report(1)-(3)是不是要上傳 kaggle 8 次,這樣會浪費 kaggle 上傳的 coda

Ans. 同學可以先把 model 訓練好的答案做好，等到 kaggle 死線之後便可以無限上傳看 error 了

Q2. 如果預先對 training data 做 normalization，那我可以上傳 train.csv 然後在 hw1.sh 中自己讀進來嗎

Ans. 可以

Posted by: [ntumlta](#)

Contact information: ntu.mlta@gmail.com.

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