

=

1. 請簡明扼要地闡述你如何抽取模型的輸入特徵 (**feature**)

答：

抽取前九個小時的 PM2.5 作為 feature

2. 請作圖比較不同訓練資料量對於 **PM2.5** 預測準確率的影響

答：

在抽取前九個小時的 PM2.5 作為 feature 的情況下

train 100 個連續 10 小時

-> kaggle 分數 6.00183(public) 7.39316(private)

train 200 個連續 10 小時

-> kaggle 分數 5.84801(public) 7.38341(private)

train 471 個連續 10 小時

-> kaggle 分數 5.79969(public) 7.30647(private)

一般來說，train 越多資料會越準確

3. 請比較不同複雜度的模型對於 **PM2.5** 預測準確率的影響

答：

抽取前九個小時的 PM2.5

-> kaggle 分數 5.79969(public) 7.30647(private)

抽取前九個小時的 PM2.5 以及 $(PM2.5)^2$

-> kaggle 分數 5.80742(public) 6.58957(private)

就結果而言多了 $(PM2.5)^2$ 以後(變得複雜)比較能夠準確預測

4. 請討論正規化(**regularization**)對於 **PM2.5** 預測準確率的影響

答：

我認為我的 model 很簡單(只抽取 PM2.5 作為 feature)，不太會有 overfit 的現象，因此用 regularization 來調整 model 也不會造成什麼影響。

5. 在線性回歸問題中，假設有 N 筆訓練資料，每筆訓練資料的特徵 (feature) 為一向量 \mathbf{x}^n ，其標註(label)為一存量 y^n ，模型參數為一向量 \mathbf{w} (此處忽略偏權值 b)，則線性回歸的損失函數(loss function)為 $\sum_{n=1}^N (\hat{y}^n - \mathbf{x}^n \cdot \mathbf{w})^2$ 。若將所有訓練資料的特徵值以矩陣 $\mathbf{X} = [\mathbf{x}^1 \mathbf{x}^2 \dots \mathbf{x}^N]$ 表示，所有訓練資料的標註以向量 $\mathbf{y} = [y^1 y^2 \dots y^N]^T$ 表示，請以 \mathbf{X} 和 \mathbf{y} 表示可以最小化損失函數的向量 \mathbf{w} 。

答：

$((\mathbf{X}^T \mathbf{X})^{-1}) * \mathbf{X}^T * \mathbf{y}$ (\mathbf{X}^T 為 \mathbf{X} 之轉置矩陣)

參考資料: [https://en.wikipedia.org/wiki/Linear_least_squares_\(mathematics\)](https://en.wikipedia.org/wiki/Linear_least_squares_(mathematics))