

請實做以下兩種不同 feature 的模型，回答第 (1) ~ (3) 題：

1. 抽全部 9 小時內的污染源 feature 的一次項(加 bias)
2. 抽全部 9 小時內 pm2.5 的一次項當作 feature(加 bias)

1. (2%)記錄誤差值 (RMSE)(根據 kaggle public+private 分數)，討論兩種 feature 的影響

- 抽全部 9 小時內的污染源 feature 的一次項(加 bias)

private score : 5.40280

public score : 7.68784

RMSE : 6.644288

submit_simple_all_feature.csv 2 minutes ago by Yao Lin 抽全部9小時內的污染源feature的一次項(加bias)	5.40280	7.68784	<input type="checkbox"/>
--	---------	---------	--------------------------

- 抽全部 9 小時內 pm2.5 的一次項當作 feature(加 bias)

private score : 5.64609

public score : 7.47123

RMSE : 6.621843

submit_simple_one.csv 2 hours ago by Yao Lin 抽全部9小時內pm2.5的一次項當作feature(加bias)	5.64609	7.47123	<input type="checkbox"/>
---	---------	---------	--------------------------

上面的結果顯示，當我使用較少 feature 所產生出來的 model (其他參數設定相同情況下)，雖然在 public score 上似乎有較好的表現，但 private score 的表現卻比多 feature 的差，由此可知當我擁有較少的資訊時，所產生出來的 model 較無法準確預測未知的資料。換句話說就是一個 machine learning 的 work 並不一定我們看到是好的就代表越準確。

2. (1%)將 feature 從抽前 9 小時改成抽前 5 小時，討論其變化

- 抽全部 5 小時內的污染源 feature 的一次項(加 bias)

private score : 5.35036

public score : 7.69952

RMSE : 6.629866

submit_simple_all_five.csv 2 hours ago by Yao Lin 抽全部5小時內的污染源feature的一次項(加bias)	5.35036	7.69952	<input type="checkbox"/>
---	---------	---------	--------------------------

- 抽全部 5 小時內 pm2.5 的一次項當作 feature(加 bias)

private score : 5.80073

public score : 7.59870

RMSE : 6.75976

[submit_simple_one_five.csv](#)

3 minutes ago by Yao Lin

5.80073

7.59870



抽全部5小時內pm2.5的一次項當作feature(加bias)

由此結果得知，當我選取五小時資料當 **training data**，效果會比取九小時的結果還要差，可以知道當我擁有更多的資訊，就比較能夠更準確的預測。

3. (1%)Regularization on all the weight with $\lambda=0.1$ 、 0.01 、 0.001 、 0.0001 ，並作圖

- 抽全部 9 小時內的污染源 **feature** 的一次項(加 bias) + regularization

1. $\lambda=0.1$

[submit_simple_all_1.csv](#)

5 minutes ago by Yao Lin

simple_all_re_1

5.40280

7.68784

2. $\lambda=0.01$

[submit_simple_all_2.csv](#)

2 minutes ago by Yao Lin

simple_all_re_2

5.40280

7.68784

3. $\lambda=0.001$

[submit_simple_all_3.csv](#)

a few seconds ago by Yao Lin

simple_all_re_3

5.40280

7.68784

4. $\lambda=0.0001$

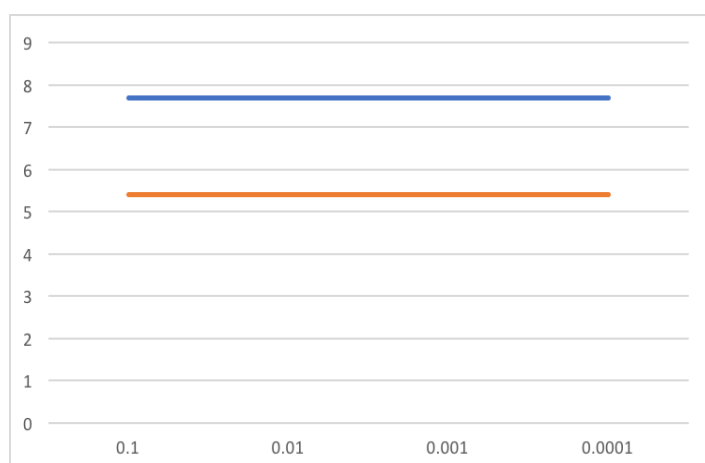
[submit_simple_all_4.csv](#)

just now by Yao Lin

simple_all_re_4

5.40280

7.68784



- 抽全部 9 小時內 pm2.5 的一次項當作 feature(加 bias) + regularization

1. $\lambda=0.1$

[submit_simple_one_re_1.csv](#)

2 minutes ago by Yao Lin

simple one re 1

5.64609

7.47123

2. $\lambda=0.01$

[submit_simple_one_re_2.csv](#)

a minute ago by Yao Lin

[add submission details](#)

5.64609

7.47123

3. $\lambda=0.001$

[submit_simple_one_re_3.csv](#)

3 minutes ago by Yao Lin

[add submission details](#)

5.64609

7.47123

4. $\lambda=0.0001$

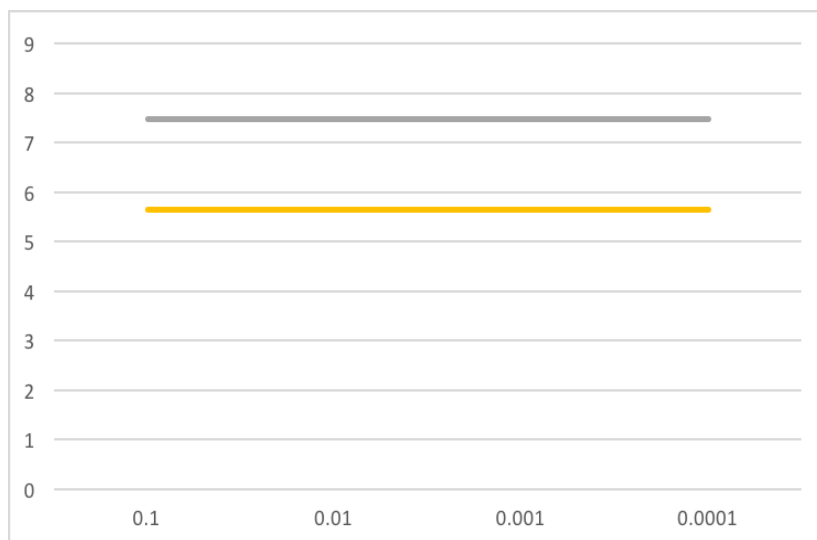
[submit_simple_one_re_4.csv](#)

a few seconds ago by Yao Lin

[add submission details](#)

5.64609

7.47123



4. (1%)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 x^n ，其標註(label)為一存量 y^n ，模型參數為一向量 w (此處忽略偏權值 b)，則線性回歸的損失函數(loss function)為 $\sum_{n=1}^N (y^n - x^n \cdot w)^2$ 。若將所有訓練資料的特徵值以矩陣 $X = [x^1 x^2 \dots x^N]^T$ 表示，所有訓練資料的標註以向量 $y = [y^1 y^2 \dots y^N]^T$ 表示，請問如何以 X 和 y 表示可以最小化損失函數的向量 w ？請寫下算式並選出正確答案。(其中 $X^T X$ 為invertible)

- A. $(X^T X)^{-1} X^T y$
- B. $(X^T X)^{-0} X^T y$
- C. $(X^T X)^{-1} X^T y$
- D. $(X^T X)^{-2} X^T y$

(ANS)

最小化 $Xw = y$

- I. 左右兩邊同乘 $X^T \Rightarrow X^T X w = X^T y$
- II. 左右兩邊同乘 $(X^T X)^{-1} \Rightarrow (X^T X)^{-1} X^T X w = (X^T X)^{-1} X^T y$
- III. 因 $(X^T X)^{-1} X^T X = I$ 故 上式 $\Rightarrow w = (X^T X)^{-1} X^T y$

\Rightarrow 答案為 C. $(X^T X)^{-1} X^T y$