

學號：B04901061 系級：電機三 姓名：蔡忠紘

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**)

備註：

a. **NR** 請皆設為 0，其他的數值不要做任何更動

b. 所有 **advanced** 的 **gradient descent** 技術(如: **adam**, **adagrad** 等) 都是可以用的

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

feature	private	public	RMSE
(1)all	5.53562	7.46237	6.57001
(2)only PM2.5	5.62719	7.44013	6.59624

**private** 和 **public** 的結果大小，取決於評分者怎麼切 **test data**，因此主要以 **RMSE** 作討論。根據上述兩種 **feature** 模型的預測結果，當 **feature** 只取 9 小時內的 **PM2.5** 時，**RMSE** 較大，預測結果較不準確。推測原因有二：

(一) 考慮較少樣 **feature** 的模型，預測結果理應較不準確。

(二) **train.csv** 和 **test.csv** 中，**PM2.5** 的數據包含不合理的數字「-1」，可能是量測失敗造成。而我們在沒有另外處理 **train.csv** 和 **test.csv** 中「-1」的情況下 **train** 我們的 **model**，那 **feature** 只取 **PM2.5** 的模型相對能參考的數據較少，較容易被「-1」影響，則 **RMSE** 較大屬預料之中。

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

feature	private	public	RMSE
(1)all	5.44092	7.65925	6.64333
(2)only PM2.5	5.79187	7.57904	6.74491

和取 9 小時相比，兩種模型中只取 5 小時的 **RMSE** 都較大，而其中 **feature** 只取 **PM2.5** 的 **RMSE** 依然較大，較不準確。

推測因為天氣預測不論在時間、空間上規模都較大，取 9 小時的結果可能會比只取 5 小時來的好。

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

feature	lambda( $\lambda$ )	private	public	RMSE
(1)all	0.1	5.53259	7.46292	6.56905
	0.01	5.53531	7.46242	6.56991
	0.001	5.53559	7.46237	6.57000
	0.0001	5.53562	7.46237	6.57001
(2)only PM2.5	0.1	5.62742	7.44080	6.59672
	0.01	5.62721	7.44019	6.59628
	0.001	5.62719	7.44013	6.59624
	0.0001	5.62719	7.44013	6.59624

圖中資料顯示， $\lambda$  大小對於此次預測結果並無太大影響。

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  $\mathbf{x}^n$ ，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 $\sum_{n=1}^N (y^n - \mathbf{x}^n \cdot \mathbf{w})^2$ 。若將所有訓練資料的特徵值以矩陣  $\mathbf{X} = [\mathbf{x}^1 \mathbf{x}^2 \dots \mathbf{x}^N]^T$  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，請問如何以  $\mathbf{X}$  和  $\mathbf{y}$  表示可以最小化損失函數的向量  $\mathbf{w}$  ? 請寫下算式並選出正確答案。(其中  $\mathbf{X}^T \mathbf{X}$  為 invertible)

- (a)  $(\mathbf{X}^T \mathbf{X}) \mathbf{X}^T \mathbf{y}$
- (b)  $(\mathbf{X}^T \mathbf{X})^{-0} \mathbf{X}^T \mathbf{y}$
- (c)  $(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y}$
- (d)  $(\mathbf{X}^T \mathbf{X})^{-2} \mathbf{X}^T \mathbf{y}$

A:

$$\begin{aligned} \frac{dL(\mathbf{w})}{d\mathbf{w}} &= 0 \\ \mathbf{X}^T (\mathbf{X}\mathbf{w} - \mathbf{y}) &= 0 \\ \mathbf{X}^T \mathbf{X}\mathbf{w} &= \mathbf{X}^T \mathbf{y} \\ \mathbf{w} &= (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y} \end{aligned}$$

$\therefore$  the answer is (c)