

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

ANS：

1) RMSE:

	18 feature	1 feature
public+private	6.611151042	6.514489721

2) Feature 的影響:

由上述的結果不難看出，variance 的數目多並不代表訓練的結果可以最好，而是要看資訊本身是否適合做為訓練的 feature，如果做 feature selection 的話，可以有更好的結果。在這很明顯的前 9 小時的 P.M. 2.5 可能就是些很好的特徵。

2. (1%)將feature從抽前9小時改成抽前5小時，討論其變化

ANS：

1) RMSE:

	18 feature	1 feature
public+private	6.623875413	6.744909392

2) 變化:

很明顯的當天數減少時，RMSE 都有些許上升了，所以可以猜測 P.M. 2.5 的預測跟前面多個小時的值相當有關係。

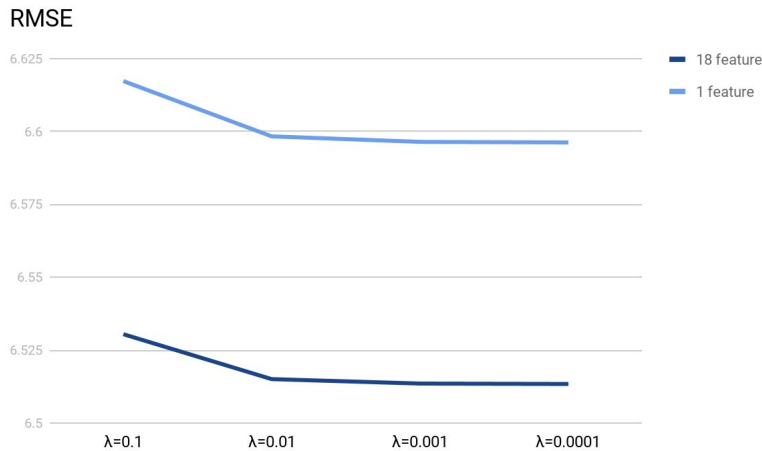
3. (1%)Regularization on all the weight with $\lambda=0.1$ 、 0.01 、 0.001 、 0.0001 ，並作圖

ANS：

1) RMSE:

	$\lambda=0.1$	$\lambda=0.01$	$\lambda=0.001$	$\lambda=0.0001$
18 feature	6.530476185	6.515069757	6.513548519	6.513398534
1 feature	6.617300235	6.598338885	6.596447767	6.596256964

2) RMSE變化圖



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) $(X^T X) X^T y$
- (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 : C

$$\text{minimize } \|y - Xw\| \Leftrightarrow Xw = \text{proj}_X y$$

$$\Leftrightarrow \langle y - Xw, b \rangle = 0, \forall b \in R(X)$$

$$\because b \in R(X), \text{ 令 } b = Xw'$$

$$\Leftrightarrow \langle y - Xw, Xw' \rangle = 0$$

$$\Leftrightarrow \langle y, Xw' \rangle - \langle Xw, Xw' \rangle = 0$$

$$\Leftrightarrow \langle y, Xw' \rangle = \langle Xw, Xw' \rangle$$

$$\Leftrightarrow (Xw')^T y = (Xw')^T (Xw)$$

$$\Leftrightarrow w'^T X^T y = w'^T X^T X w$$

$$\Leftrightarrow \langle X^T y, w' \rangle = \langle X^T X w, w' \rangle$$

$$\Leftrightarrow X^T y = X^T X w$$

$$\because (X^T X) \text{ is invertible}$$

$$\therefore w = (X^T X)^{-1} y$$