

Machine Learning HW1

林子雋
b04705003 資工三

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

如下表所示，只單單取PM2.5的數據明顯比多取其他污染物好，推測是因為PM2.5本身自己跟自己就是相關係數非常高，而其他污染物雖然在一些論文中說明某些污染物跟PM2.5也有關，但可能因為其他污染物的變異數太大，造成實際預測時受到干擾比較多。

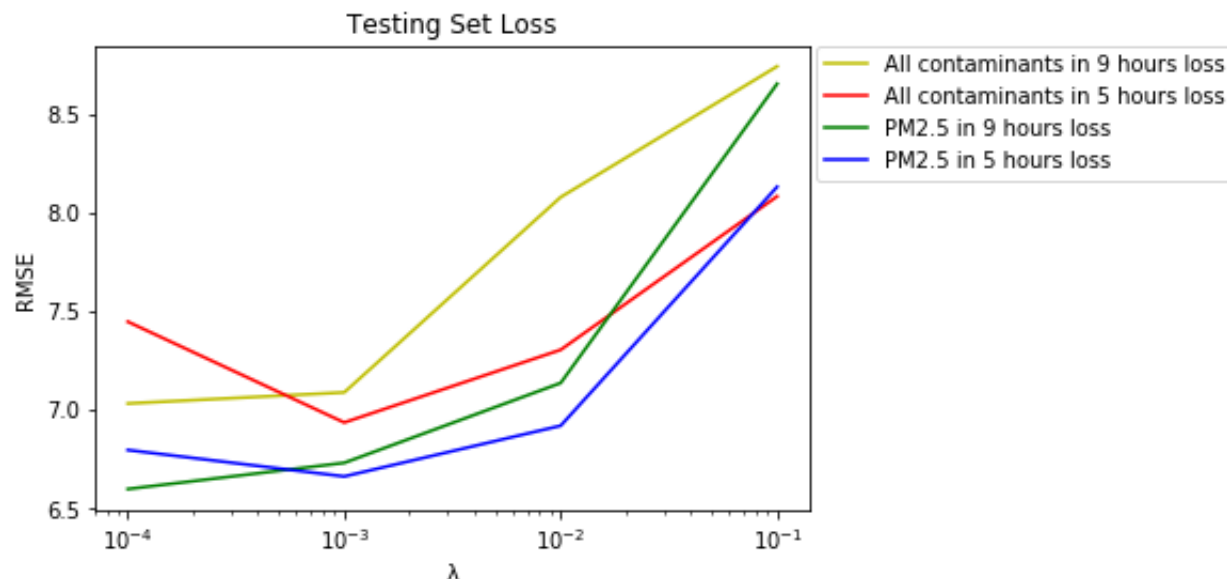
RMSE 表格				
λ	前九小時所有污染物	前五小時所有污染物	PM2.5前九小時	PM2.5前五小時
0.1	8.7398	8.0804	8.6515	8.1287
0.01	8.0763	7.3019	7.1336	6.9167
0.001	7.0857	6.9331	6.7284	6.6596
0.0001	7.0296	7.4446	6.5955	6.7935

Problem 2. 將feature從抽前9小時改成抽前5小時，討論其變化.

沿用上表，會發現抽前五小時來loss會變得比較小，推測是因為時間距離接近真正要被預測的那天，所以預測上就會比較準。

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

使用最上面的表格來作圖



Problem 4. 線性回歸問題,請寫下算式並選出正確答案.

(1) 方法一

這個問題可以被公式化成

$$Xw = y, \text{ where } w = [w_1, w_2, \dots, w_n]^T$$

令最佳解為 \hat{w} 目標要最小化 $\|y - X\hat{w}\|^2$, 已知若要最小化此方程式, $X\hat{w}$ 必須等於 y 在 X 的 column space 上投影, 因此問題可以被 formula 成

$$X\hat{w} = \text{proj}_{\text{column space of } X} y$$

又知 $x\hat{w} - y$ 的 row 向量與 X 中 column 向量正交, 因此

$$X^T(X\hat{w} - y) = 0$$

得知

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

答案為(c)

(2) 方法二 令想要最小化的函數為

$$J(w') = \|Xw' - y\|^2 = (Xw' - y)^T (Xw' - y)$$

取 w' 對 J 的偏微分

$$\frac{\partial J}{\partial w'} = \frac{\partial J}{\partial (Xw' - y)} \frac{\partial (Xw' - y)}{\partial w'} = 2X^T(Xw' - y)$$

已知最佳解落在偏微分值為零的地方

$$\frac{\partial J}{\partial w'} = 2X^T(Xw' - y)$$

令最佳解為 \hat{w} 則

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