

1. How do you select features for your model input, and what preprocessing did you perform?

- a. 我選擇 PM2.5 當 features，且嘗試了各種 9 小時的組合，選擇 RMSE 最高的一組。
- b. 我先將 data 中的 #、\*、x、A 都改成 0，接著將 data 都轉成 float 型別，最後抓取 PM2.5 當 features。

2. Compare the impact of different amounts of training data on the PM2.5 prediction accuracy. Visualize the results and explain them.

a. 240 組 training data : RMSE: 3.9336589291887023

b. 120 組 training data : RMSE: 3.934248720311966

c. 60 組 training data : RMSE: 4.10772432113455

當 training data 組數下降，模型對 data 的擬合程度也會跟著下降，可能會發生欠擬合的狀況，也較容易受到 noise 或極端值的影響。

3. Discuss the impact of regularization on PM2.5 prediction accuracy.

a. 240 組 training data : RMSE: 3.933658929188702

b. 120 組 training data : RMSE: 3.934248720311966

c. 60 組 training data : RMSE: 4.107724321134549

在 linear regression 中，正規化對準確度不會產生很大的影響，因為線性回歸模型的損失函數是平方損失函數，且該函數對參數的微分是線性的。而正規化在線性回歸中的主要作用是防止過擬合，代表我的模型並沒有過擬合的現象。