學號：R07521603 系級： 土木所碩二 姓名：蔡松霖

請實做以下兩種不同feature的模型，回答第 (1) ~ (2) 題：

(1) 抽全部9小時內的污染源feature當作一次項(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，其他的非數值(特殊字元)可以自己判斷

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

c. 第1-2題請都以題目給訂的兩種model來回答

d. 同學可以先把model訓練好，kaggle死線之後便可以無限上傳。

e. 根據助教時間的公式表示，(1) 代表 p = 9x18+1 而(2) 代表 p = 9*1+1

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

Adam, learning_rate = 0.003, early stop step = 30 (20% training data for validation)

RMSE result:

Features	Public	Private
抽全部9小時內的污染源feature當作一次項(加bias)	5.49596	5.48863
抽全部9小時內PM2.5的一次項當作feature(加bias)	5.99751	5.89860

可以看到使用全部features跟只使用PM2.5比較在成績上有較好的表現。不過單是PM2.5就有一定程度的預測能力，可見PM2.5是相當關鍵的一個feature，但仍有其他的features也同樣有一定的影響力不容忽視。

2. (1%)解釋什麼樣的data preprocessing 可以improve你的training/testing accuracy，ex. 你怎麼挑掉你覺得不適合的data points。請提供數據(RMSE)以佐證你的想法。

在training data的處理部分，假如有包含特殊符號或是NaN的資料部分，都直接過濾掉不使用，主要是因為含有特殊符號的部分有些數值遠遠高於平均值很多，影響訓練收斂，光是在validation set上的結果就有很顯著的差異。

RMSE result:

Data process	Public	Private
僅濾掉NaN的資料	8.78112	6.47980
濾掉所有含特殊符號或是NaN的資料	5.49596	5.48863

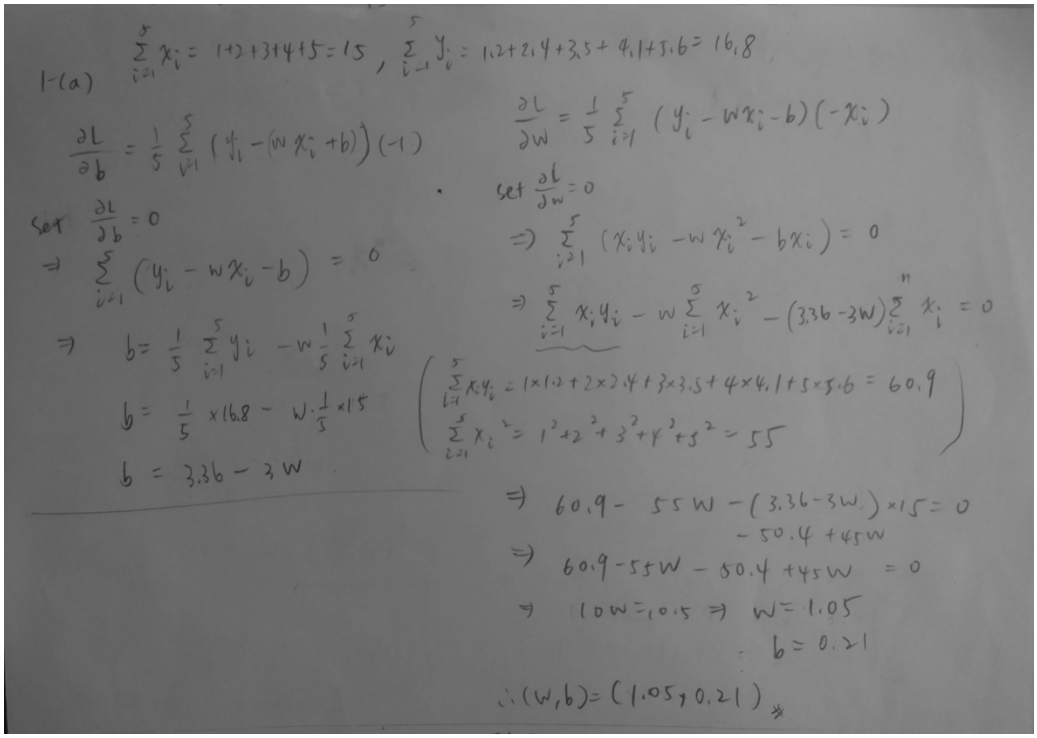
3.(3%) Refer to math problem
https://hackmd.io/RFiu1FsYR5uQTrrpdxUvlw?view

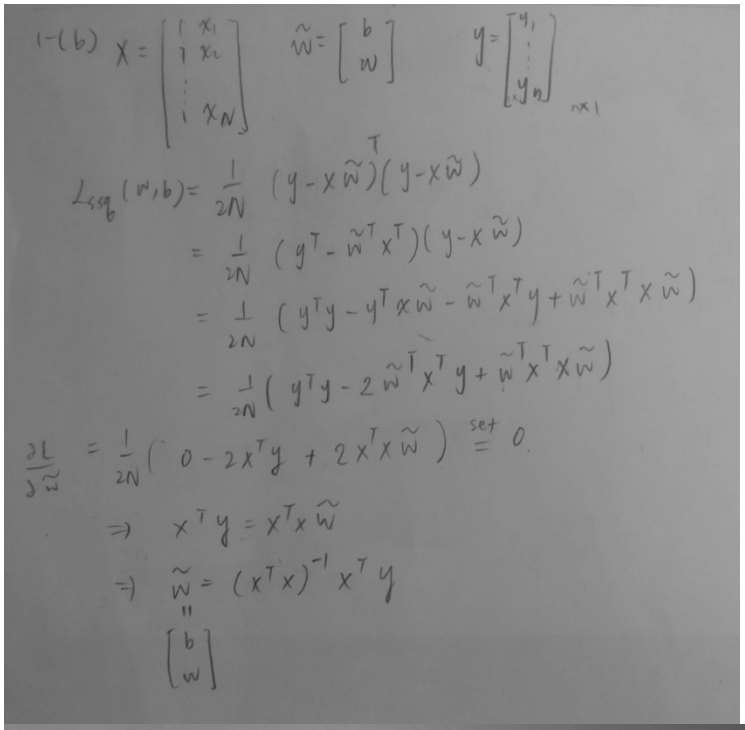
Math problem:

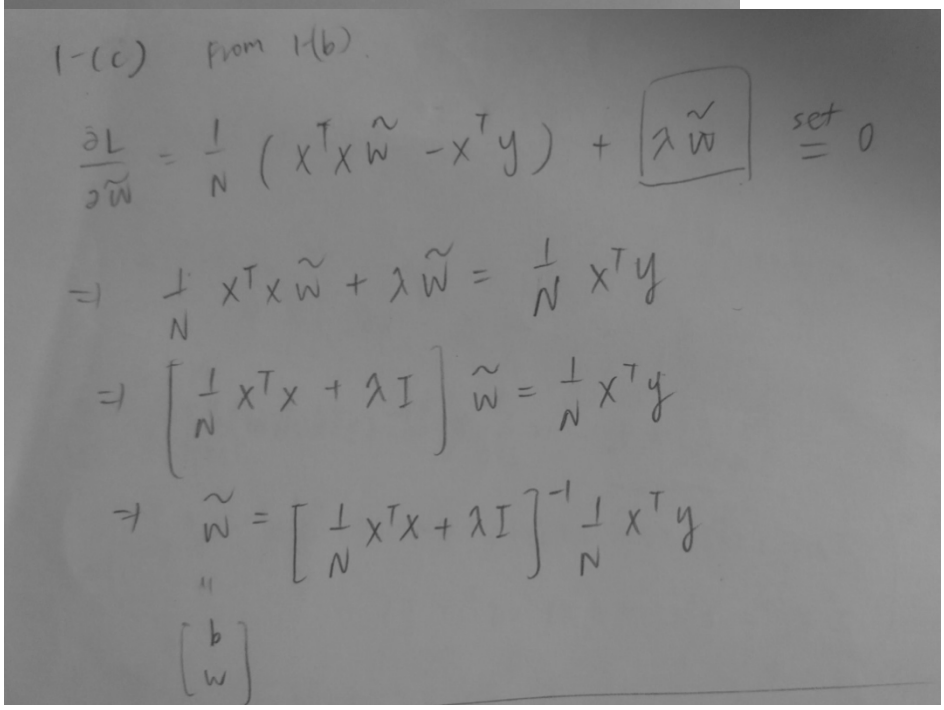
1.

discuss with r08521604 王鈞平

ref: http://web.mit.edu/zoya/www/linearRegression.pdf

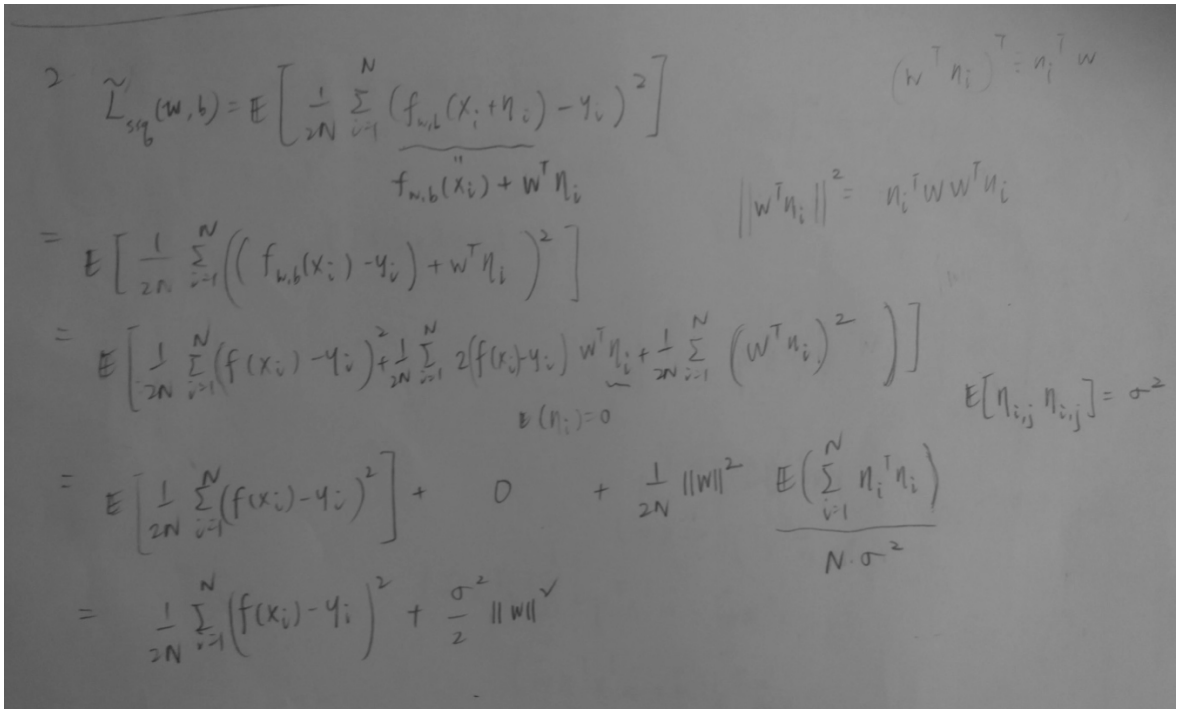






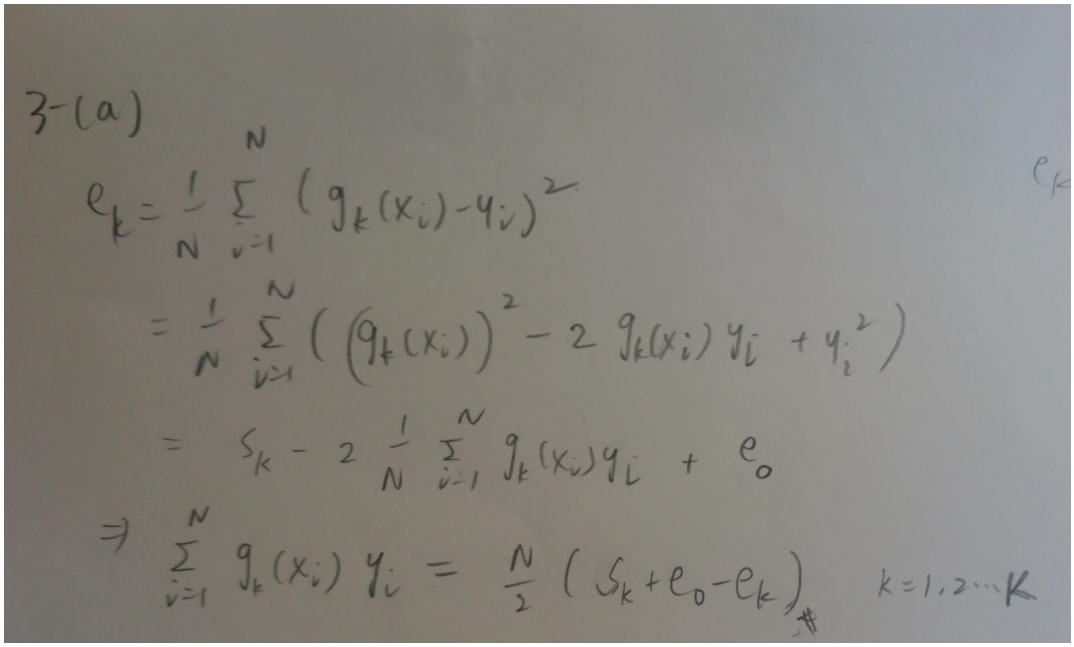
2.

ref: https://medium.com/autonomous-agents/mathematical-foundation-for-noise-bias-and-variance-in-neuralnetworks-4f79ee801850



3.

ref: https://arxiv.org/pdf/1802.04947.pdf



3-(b) $\frac{1}{N}\sum_{i=1}^{N}\left(\sum_{k=1}^{K}\alpha_k g_k(x_i)-y_i\right)^2$

$= \frac{1}{N}\|G\alpha - y\|^2$

由 1-(b) $\alpha = (G^TG)^{-1}G^Ty$ ★

$\alpha = \begin{bmatrix}\alpha_1\\ \vdots \\ \alpha_K\end{bmatrix} \in \mathbb{R}^K$

$G = \begin{bmatrix} g_1(x_1) & g_2(x_1) & \cdots & g_K(x_1) \\ \vdots & & & \\ & & & \end{bmatrix} \in \mathbb{R}^{N\times K}$ $(i=1\sim N)$

$Y = \begin{pmatrix} y_1 \\ \vdots \\ y_N \end{pmatrix}$ $N\times 1$

$\left(G^Ty = \left[\frac{N}{2}(S_k+e_0-e_k)\right]_{k=1,2\cdots K} \right)$