1.
(1)的 kaggle 分數分別為 5.51158, 5.90694
(2)的 kaggle 分數分別為 6.13449, 6.22446
前為 private 後為 public, 可以發現使用全部汙染源當作 feature 的模型有較好的表現, 不過只用 PM2.5 的表現也不差, 代表大部分的解釋力可能來自 PM2.5

2.
首先檢查有沒有不符合資料型態的資料 EX:* # NR
將其調整為適當的型態然後再來填補缺失值(全部填 0)
此時的 RMSE 約為 6.81
檢查後發現存在離群值, 將超過 1.5 倍標準差的離群值給去掉
此時的 RMSE 約為 5.69

1.(a)

$\frac{\partial L}{\partial b} = 2\sum(y_i - wx_i - b) = 0$

$\frac{\partial L}{\partial w} = 2\sum(y_i - wx_i - b)x_i = 0$

$$\Rightarrow \begin{cases} 5b + w\sum x_i = \sum y_i \\ b\sum x_i + w\sum x_i^2 = \sum x_i y_i \end{cases}$$

$$\Rightarrow \begin{bmatrix} 5 & \sum x_i \\ \sum x_i & \sum x_i^2 \end{bmatrix} \begin{bmatrix} b \\ w \end{bmatrix} = \begin{bmatrix} \sum y_i \\ \sum x_i y_i \end{bmatrix}$$

by cramer's rule

$$w = \frac{\begin{vmatrix} 5 & \sum y_i \\ \sum x_i & \sum x_i y_i \end{vmatrix}}{\begin{vmatrix} 5 & \sum x_i \\ \sum x_i & \sum x_i^2 \end{vmatrix}} = \frac{\sum(x_i - \overline{x})(y_i - \overline{y})}{\sum(x_i - \overline{x})^2}$$

$b = \overline{y} - w\overline{x}$

$w = 1.05, b = 0.21$

1.(b)

$w^T = [w_1 w_2 w_3 ... w_k]$

$x_i^T = [x_{1i} x_{2i} x_{3i} ... x_{ki}]$

$let \quad \beta = [b \quad w]^T$

$\Rightarrow L = \frac{1}{2N}(\sum y_i^2 - 2\beta \sum x_i y_i + \beta^T \sum x_i x_i^T \beta)$

$\Rightarrow \frac{\partial L}{\partial \beta} = -2\sum x_i y_i + 2\sum x_i x_i^T \beta = 0$

$\Rightarrow \beta = (\sum x_i x_i^T)^{-1} \sum x_i y_i$

1.(c)

$Let\ w^T = (w_1\ w_2\ ....\ w_k\ b),\ x_i^T = (x_{1i}\ x_{2i}...\ x_{ki}\ 1)$

$L = \frac{1}{2N}\sum(y_i - w^T x_i)^2 + \frac{\lambda}{2}\|w\|^2$

$\frac{L}{\partial w^T} = \frac{1}{N}\sum(y_i - w^T x_i)x_i^T + \lambda w^T - \lambda = 0 \Rightarrow \frac{1}{N}\sum w^T x_i x_i^T + \lambda w^T = \frac{1}{N}\sum y_i x_i^T + \lambda$

$\Rightarrow w^T = (\sum y_i x_i^T)(\sum x_i x_i^T + \lambda)^{-1}$

2.

$L = \frac{1}{2N}E\left[\sum(f(x_i + \eta_i)^2 - 2\sum f(x_i + \eta_i)y_i + \sum {y_i}^2\right]$

$$E(f(x_i+\eta_i)^2) = w^T E[\ (x_i+\eta_i)(x_i+\eta_i)^T\ ]w + b$$

$$= w^T E[\ x_i x_i^T + x_i \eta_i^T + \eta_i x_i^T + \eta_i \eta_i^T\ ]w + b$$

$$= w^T (x_i x_i^T + \sigma^2) w + b$$

$$= w^T x_i x_i^T w + b + \sigma \left\| w \right\|^2$$

$$= f(x_i)^2 + \sigma \left\| w \right\|^2$$

$$E[\ f(x_i+\eta_i) y_i\ ] = E[\ w^T (x_i+\eta_i) y_i\ ]$$

$$= f(x_i) y_i$$

$\Rightarrow L = \frac{1}{2N} \sum (f(x_i) - y_i)^2 + \frac{\sigma}{2} \|w\|^2$

3.(a)

$$e_k = \frac{1}{N} \sum (g_k(x_i) - y_i)^2$$

$$= \frac{1}{N} \sum (g_k(x_i)^2 - 2 g_k(x_i) y_i + y_i^2)$$

$$= s_k + e_0 - \frac{2}{N} \sum (g_k(x_i) y_i)$$

$\Rightarrow \sum g_k(x_i)y_i = \frac{2}{N}(s_k + e_0 - e_k)$

3.(b)

$L = \frac{1}{N}\sum_i(\alpha^T g - y_i)^2$

$\frac{\partial L}{\partial \alpha} = \frac{2}{N}\sum_i(\alpha^T g - y_i)g^T = 0$

$\Rightarrow \sum_i(\alpha^T g g^T) = \sum_i y_i g^T$

$\Rightarrow \alpha^T s = (s + e_0 - e)^T$

$\Rightarrow \alpha^T = (s + e_0 - e)^T s^{-1}$