Machine Learning 2019 Fall
HW1 Report

b05902105 資工四 余友竹

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

	PUBLIC	PRIVATE
PM2.5	5.98437	5.89846
All	5.60406	5.57277

本次實驗單獨探討使用的feature數對結果的影響，兩個model的有做data preprocessing，並採用助教提供的minibatch求出最佳解。

根據上表，可以看出，使用全部18個資料，比單獨使用PM2.5做預測來的更精準，似乎可以解釋，PM2.5不僅跟歷史紀錄有關，也受到當下其他的環境數據的影響。

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。

由於Linear Regression 的 Error Function (RMSE)是一個Convex Function，收斂到最小值(0)時的w, b理論上一定是固定的。

因此，本次作業要得到良好預測的關鍵其實在於Data Preprocessing，而非Gradient Descent的實作。

Note: 避免浪費Kaggle上傳次數，底下將以：

- Training Error (E_{in})
- 10-fold cross validation error (E_{val})

做為比較依據。

首先，列出完全沒做Data Preprocessing 的Error，作為Baseline。

	E_{in}	E_{val}
PM2.5	38.69072	38.22786
ALL	37.89876	40.86500

(1). 使用前、後一小時的數據補足缺失項

在本次測驗的Training Data中，有大量的缺失項，因此如何補齊這些缺失項是一大哉問。

考慮到此次作業提供的數據都屬於時序性的數據，因此，我的直覺是使用前後一小時的資料來補齊缺失項。

	E_{in}	E_{val}
PM2.5	38.65485	37.93288
ALL	37.74840	41.10642

由表可知，Training Error 略有降低，但老實說，這是一個非常糟糕的想法。

在實際上傳到Kaggle後，我發現Training Error雖然如此之高，最後的Testing Error卻僅有7.11左右。

會造成Training Error 與Testing Error如此大相逕庭的原因在於，助教已經事先清過Testing Data了，而在Training Data上，許多資料都是由腦補的方式取得，相當的不精準。

(2). 將所有的缺失項全部以0填補，並將PM2.5項怪異的值去除

有缺失的項就代表當天的統計資料其實是有缺漏、或者是不精準的。

考慮到我們的Training Data其實不算少，其實直接將不精準的數據直接剔除才是最好的設計。

因此我將缺失項以0填補，而PM2.5值不論再低，都不該為過低；同時，PM2.5值不論再高，都不應該超過100，所以在Data Preprocessing的部分，我只保留所有$2 < \text{PM2.5} \leq 100$ 的數據。

	E_{in}	E_{val}
PM2.5	4.72659	4.72840
ALL	4.45382	4.56207

可以看到顯著地進步，由此判斷，Training Data上的確有許多不精準的數據。

值得注意的是，在測試時，我另外測試了僅考慮 > 0的值，卻得到跟baseline差不多的結果，反而是加上≤ 100的限制後，才得到如此好的結果。

(3). 考慮(2)的結果，並剔除所有≤ 0的值(final result)

由上表的進步幅度可以得知，問題的確出在不精準的資料，因此我將限制範圍擴大到所有的data上。

除了降雨量外，所有的環境資料的量測都不該為0，因此我將所有≤ 0的data全部剔除，不予以採用。

同時，延續上個Trial的結果，對PM2.5的值做限制，並套用到相同屬性的數據PM10上。

	E_{in}	E_{val}
PM2.5	4.44609	4.45383
ALL	4.12393	4.21948

3. Math Problem(見下頁)

ML 2019 Fall HW1 Hand-Written Assignment

b05902105 余友竹

1. Closed-Form Linear Regression Solution

· 考慮簡化版 linear regression (without bias).

$\Rightarrow \min \sum_{i=1}^{n} (\vec{w}^T \vec{x}_i - y_i)^2$, $\vec{w}, \vec{x}_i \in \mathbb{R}^{k\times 1}$, $y_i \in \mathbb{R}$, $\forall i = 1, \cdots, n$

$\because \vec{x}_i, \vec{w} \in \mathbb{R}^{k\times 1}$, $\therefore \vec{w}^T \vec{x}_i = \vec{x}_i^T \vec{w} = \langle \vec{w}, \vec{x}_i \rangle$, $\forall i = 1, \cdots, n$

$$\Rightarrow \sum_{i=1}^{n} (\vec{w}^T \vec{x}_i - y_i)^2 = \sum_{i=1}^{n} (w^T \vec{x}_i - y_i)(w^T \vec{x}_i - y_i)$$

$$= \begin{bmatrix} \vec{x}_1^T \vec{w} - y_1 \\ \vdots \\ \vec{x}_n^T \vec{w} - y_n \end{bmatrix}^T \begin{bmatrix} \vec{x}_1^T \vec{w} - y_1 \\ \vdots \\ \vec{x}_n^T \vec{w} - y_n \end{bmatrix} = \left\langle \begin{bmatrix} \vec{x}_1^T \vec{w} - y_1 \\ \vdots \\ \vec{x}_n^T \vec{w} - y_n \end{bmatrix}, \begin{bmatrix} \vec{x}_1^T \vec{w} - y_1 \\ \vdots \\ \vec{x}_n^T \vec{w} - y_n \end{bmatrix} \right\rangle$$

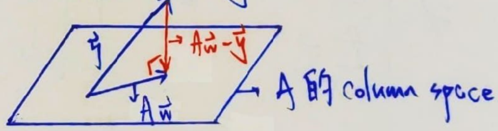
$$= \left\| \begin{bmatrix} \vec{x}_1^T \vec{w} - y_1 \\ \vdots \\ \vec{x}_n^T \vec{w} - y_n \end{bmatrix} \right\|^2 = \left\| \begin{bmatrix} \vec{x}_1^T \\ \vdots \\ \vec{x}_n^T \end{bmatrix} \vec{w} - \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} \right\|^2$$

令 $\begin{bmatrix} \vec{x}_1^T \\ \vdots \\ \vec{x}_n^T \end{bmatrix} = A \in \mathbb{R}^{n\times k}$, $\begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} = \vec{y} \in \mathbb{R}^{n\times 1}$.

則 $\underset{\vec{w}}{\operatorname{argmin}} \sum_{i=1}^{n} (\vec{w}^T \vec{x}_i - y_i)^2 \Leftrightarrow \underset{\vec{w}}{\operatorname{argmin}} \|A\vec{w} - \vec{y}\|^2 \Leftrightarrow \underset{\vec{w}}{\operatorname{argmin}} \|A\vec{w} - \vec{y}\|$

claim: $\underset{\vec{w}}{\operatorname{argmin}} \|A\vec{w} - \vec{y}\| \Leftrightarrow$ 求解 $A^T A \vec{w} = A^T \vec{y}$

pf:



$\min \|A\vec{w} - \vec{y}\|$

$\Leftrightarrow A\vec{w} - \vec{y} \perp CS(A)$, $CS(A) = \{ A\vec{b} \mid \vec{b} \in \mathbb{R}^{k\times 1} \}$

$\Leftrightarrow A\vec{w} - \vec{y} \perp A\vec{b}$, $\forall \vec{b} \in \mathbb{R}^{k\times 1}$

$\Leftrightarrow \langle A\vec{w} - \vec{y}, A\vec{b} \rangle = 0$, $\forall \vec{b} \in \mathbb{R}^{k\times 1}$

$\Leftrightarrow \vec{b}^T A^T (A\vec{w} - \vec{y}) = 0$ 〃

$\Leftrightarrow \vec{b}^T (A^T A\vec{w} - A^T \vec{y}) = 0$ 〃

$\Leftrightarrow \langle A^T A\vec{w} - A^T \vec{y}, \vec{b} \rangle = 0$, $\forall \vec{b} \in \mathbb{R}^{k\times 1}$

$\Leftrightarrow A^T A\vec{w} - A^T \vec{y} = \vec{0}$

$\Leftrightarrow A^T A\vec{w} = A^T \vec{y}$ ✱

$\therefore \underset{\vec{w}}{\operatorname{argmin}} \sum_{i=1}^{n} (\vec{w}^T \vec{x}_i - y_i)^2 \Leftrightarrow$ 求解 $A^T A\vec{w} = A^T \vec{y}$

a. $\underset{w,b}{\operatorname{argmin}} \sum_{i=1}^{5} (w \cdot x_i + b - y_i)^2$，令 $A = \begin{bmatrix} 1 & 1 \\ 2 & 1 \\ 3 & 1 \\ 4 & 1 \\ 5 & 1 \end{bmatrix}$，$\vec{x} = \begin{bmatrix} w \\ b \end{bmatrix}$，$\vec{y} = \begin{bmatrix} 1.2 \\ 2.4 \\ 3.5 \\ 4.1 \\ 5.6 \end{bmatrix}$.

所求 $\Leftrightarrow$ 求解 $A^T A \vec{x} = A^T \vec{y}$.

$\Rightarrow \vec{x} = (A^T A)^{-1} A^T \vec{y} = \begin{bmatrix} 55 & 15 \\ 15 & 5 \end{bmatrix}^{-1} \begin{bmatrix} 1 & 2 & 3 & 4 & 5 \\ 1 & 1 & 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 1.2 \\ 2.4 \\ 3.5 \\ 4.1 \\ 5.6 \end{bmatrix}$.

$= \cdots = \begin{bmatrix} 1.05 \\ 0.21 \end{bmatrix}$ (計算略).

$\Rightarrow w = 1.05$，$b = 0.21$ #.

b. 令 $X' = \begin{bmatrix} \vec{x}_1^T & 1 \\ \vec{x}_2^T & 1 \\ \vdots & \vdots \\ \vec{x}_n^T & 1 \end{bmatrix} \in \mathbb{R}^{n \times (k+1)}$，$\vec{w}' = \begin{bmatrix} \vec{w} \\ b \end{bmatrix} \in \mathbb{R}^{(k+1) \times 1}$

則所求 $\Leftrightarrow$ 求解 $(X'^T X') \vec{w}' = X'^T \vec{y}$

$\Rightarrow \left(\begin{bmatrix} \vec{x}_1 & \cdots & \vec{x}_n \\ 1 & \cdots & 1 \end{bmatrix} \begin{bmatrix} \vec{x}_1^T & 1 \\ \vdots & \vdots \\ \vec{x}_n^T & 1 \end{bmatrix} \right) \begin{bmatrix} \vec{w} \\ b \end{bmatrix} = \begin{bmatrix} \vec{x}_1 & \cdots & \vec{x}_n \\ 1 & \cdots & 1 \end{bmatrix} \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}$.

$\Rightarrow \begin{bmatrix} \sum_{i=1}^{n} \vec{x}_i \vec{x}_i^T & \sum_{i=1}^{n} \vec{x}_i \\ \sum_{i=1}^{n} \vec{x}_i^T & n \end{bmatrix} \begin{bmatrix} \vec{w} \\ b \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^{n} \vec{x}_i \cdot y_i \\ \sum_{i=1}^{n} y_i \end{bmatrix}$.

$\Rightarrow \begin{cases} \sum_{i=1}^{n} \vec{x}_i \vec{x}_i^T \cdot \vec{w} + \sum_{i=1}^{n} \vec{x}_i \cdot b = \sum_{i=1}^{n} \vec{x}_i \cdot y_i & \text{—— ①} \\ \sum_{i=1}^{n} \vec{x}_i^T \cdot \vec{w} + nb = \sum_{i=1}^{n} y_i & \text{—— ②} \end{cases}$

由②得：

$b = \frac{1}{n} \sum_{i=1}^{n} y_i - \frac{1}{n} \sum_{i=1}^{n} \vec{x}_i^T \cdot \vec{w}$

令 $\frac{1}{n} \sum_{i=1}^{n} y_i = \bar{y}$，$\frac{1}{n} \sum_{i=1}^{n} \vec{x}_i = \bar{X}$，$\frac{1}{n} \sum_{i=1}^{n} \vec{x}_i^T = \overline{X^T} = \bar{X}^T$

則 $b = \bar{y} - \bar{X}^T \cdot \vec{w}$ 代回 ①

$\Rightarrow \sum_{i=1}^{n} \vec{x}_i \vec{x}_i^T \cdot \vec{w} + (\bar{y} - \bar{X}^T \vec{w}) \cdot \sum_{i=1}^{n} \vec{x}_i = \sum_{i=1}^{n} \vec{x}_i \cdot y_i$

$\Rightarrow \sum_{i=1}^{n} (\vec{x}_i \vec{x}_i^T \vec{w} - \vec{x}_i \cdot \bar{X}^T \vec{w}) = \sum_{i=1}^{n} \vec{x}_i \cdot y_i - \bar{y} \cdot \sum_{i=1}^{n} \vec{x}_i$

$\Rightarrow \left[\sum_{i=1}^{n} \vec{x}_i (\vec{x}_i^T - \bar{X}^T) \right] \vec{w} = \sum_{i=1}^{n} \vec{x}_i (y_i - \bar{y})$

$\Rightarrow \begin{cases} \vec{w} = \left[\sum_{i=1}^{n} \vec{x}_i (\vec{x}_i^T - \bar{X}^T) \right]^{-1} \left(\sum_{i=1}^{n} \vec{x}_i (y_i - \bar{y}) \right) \\ b = \bar{y} - \bar{X}^T \vec{w} \text{ (由上式 } \vec{w} \text{ 代回即可求解)} \end{cases}$

P2. 再做化簡：

令 $X = \begin{bmatrix} \vec{x}_1^T \\ \vdots \\ \vec{x}_n^T \end{bmatrix} \in \mathbb{R}^{n \times k}$，

則 $\begin{cases} \vec{w} = [X^TX - n \cdot \bar{x}\,\bar{x}^T]^{-1}(X^Ty - n\bar{y}\,\bar{x}) \\ b = \bar{y} - \bar{x}^T\vec{w} \quad (\vec{w} \text{ 由上式代回}) \end{cases}$ #

c. $\underset{w,b}{\operatorname{argmin}} \frac{1}{2N}\sum_{i=1}^{N}(y_i - (\vec{w}^T\vec{x}_i + b))^2 + \frac{\lambda}{2}\|\vec{w}\|^2$

因 $L_{reg}(w,b)$ 為 convex function，求最小值時的 w, b.

$\Rightarrow$ 求解 $\nabla_w(L_{reg}(w,b)) = \vec{0}$，$\nabla_b(L_{reg}(w,b)) = 0$.

$$\frac{\partial L}{\partial w} = \frac{1}{N}\sum_{i=1}^{N}(y_i - (\vec{w}^T\vec{x}_i + b)) \cdot (-\vec{x}_i) + \lambda\vec{w}$$

$\Rightarrow \frac{1}{N}\left(\sum_{i=1}^{N}\vec{x}_i \cdot \vec{x}_i^T\vec{w}\right) + \lambda\vec{w} = \frac{1}{N}\sum_{i=1}^{N}(\vec{x}_i \cdot y_i - b \cdot \vec{x}_i)$

$\Rightarrow \frac{1}{N}X^TX\vec{w} + \lambda\vec{w} = \frac{1}{N}(X^Ty - Nb\bar{x})$.

$\Rightarrow (\frac{1}{N}X^TX + \lambda I)\vec{w} = \frac{1}{N}(X^Ty - Nb\bar{x})$

$\Rightarrow \vec{w} = (\frac{1}{N}X^TX + \lambda I)^{-1} \cdot \frac{1}{N}(X^Ty - Nb\bar{x})$.

$$\frac{\partial L}{\partial b} = \frac{1}{N}\sum_{i=1}^{n}(y_i - (\vec{w}^T\vec{x}_i + b)) \cdot (-1)$$

$\Rightarrow \sum_{i=1}^{n}(\vec{w}^T\vec{x}_i + b) = \sum_{i=1}^{n}y_i$

$\Rightarrow b = \frac{1}{n}\sum_{i=1}^{n}y_i - \frac{1}{n}\sum_{i=1}^{n}\vec{x}_i^T\vec{w} = \bar{y} - \bar{x}^T\vec{w}$

$\Rightarrow \begin{cases} \vec{w} = (\frac{1}{N}X^TX + \lambda I)^{-1} \cdot \frac{1}{N}(X^Ty - Nb\bar{x}) & \text{— ③} \\ b = \bar{y} - \bar{x}^T\vec{w} & \text{— ④} \end{cases}$

④ 代入 ③.

$\Rightarrow (\frac{1}{N}X^TX + \lambda I)\vec{w} = \frac{1}{N}X^Ty - (\bar{y} - \bar{x}^T\vec{w})\bar{x}$

$\Rightarrow (\frac{1}{N}X^TX + \lambda I)\vec{w} = \frac{1}{N}X^Ty - \bar{y}\bar{x} + \bar{x}\bar{x}^T\vec{w}$

$\Rightarrow (\frac{1}{N}X^TX + \lambda I - \bar{x}\bar{x}^T)\vec{w} = \frac{1}{N}X^Ty - \bar{y}\bar{x}$

$\Rightarrow \begin{cases} \vec{w} = (\frac{1}{N}X^TX - \bar{x}\bar{x}^T + \lambda I)^{-1}(\frac{1}{N}X^Ty - \bar{y}\bar{x}) \\ b = \bar{y} - \bar{x}^T\vec{w} \quad (\vec{w} \text{ 由上式代回即可.}) \end{cases}$ #.

2. $\tilde{L}_{ssq} = \mathbb{E}\left[\frac{1}{2N}\sum_{i=1}^{N}\left(f_{\vec{w},b}(\vec{x}_i+\vec{\eta}_i)-y_i\right)^2\right]$

$$= \mathbb{E}\left[\frac{1}{2N}\sum_{i=1}^{N}\left(\vec{w}^T(\vec{x}_i+\vec{\eta}_i)+b-y_i\right)^2\right]$$

$$= \mathbb{E}\left[\frac{1}{2N}\left(\sum_{i=1}^{N}(\vec{w}^T\vec{x}_i+b-y_i)^2+2\cdot\sum_{i=1}^{N}(\vec{w}^T\vec{x}_i+b-y_i)(\vec{w}^T\vec{\eta}_i)+\sum_{i=1}^{N}(\vec{w}^T\vec{\eta}_i)^2\right)\right].$$

$$= \frac{1}{2N}\left(\sum_{i=1}^{N}\mathbb{E}\left[(f_{\vec{w},b}(\vec{x}_i)-y_i)^2\right]+2\sum_{i=1}^{N}\mathbb{E}\left[(\vec{w}^T\vec{x}_i+b-y_i)(\vec{w}^T\vec{\eta}_i)\right]+\sum_{i=1}^{N}\mathbb{E}\left[(\vec{w}^T\vec{\eta}_i)^2\right]\right)$$

$$= \frac{1}{2N}\left(\sum_{i=1}^{N}(f_{\vec{w},b}(\vec{x}_i)-y_i)^2+2\sum_{i=1}^{N}(\vec{w}^T\vec{x}_i+b-y_i)(\vec{w}^T\underbrace{\mathbb{E}(\vec{\eta}_i)}_{=\vec{0}})+\sum_{i=1}^{N}\mathbb{E}\left[(\vec{w}^T\vec{\eta}_i)^2\right]\right)$$

$$= \frac{1}{2N}\left(\sum_{i=1}^{N}(f_{\vec{w},b}(\vec{x}_i)-y_i)^2+0+\sum_{i=1}^{N}\mathbb{E}\left[(\vec{w}^T\vec{\eta}_i)^2\right]\right)$$

其中, $\sum_{i=1}^{N}\mathbb{E}\left[(\vec{w}^T\vec{\eta}_i)^2\right]$

$$= \sum_{i=1}^{N}\mathbb{E}\left[\left(\sum_{j=1}^{K}w_j\eta_{ij}\right)^2\right]$$

$$= \sum_{i=1}^{N}\mathbb{E}\left[\sum_{j=1}^{K}\sum_{j'=1}^{K}w_jw_{j'}\eta_{ij}\eta_{ij'}\right].$$

$$= \sum_{i=1}^{N}\left(\sum_{j=1}^{K}\sum_{j'=1}^{K}w_jw_{j'}\,\mathbb{E}[\eta_{ij}\eta_{ij'}]\right).$$

$$= \sum_{i=1}^{N}\left(\sum_{j=1}^{K}\sum_{j'=1}^{K}w_jw_{j'}\delta_{ii}\delta_{jj'}\sigma^2\right), \text{其中}, \delta_{ii}=1, \ \delta_{jj'}=\begin{cases}1 & \text{if } j=j'\\ 0 & \text{otherwise}\end{cases}$$

$$= \sum_{i=1}^{N}\left(\sum_{j=1}^{K}w_jw_j\sigma^2\right)$$

$$= \sum_{i=1}^{N}\sigma^2\left(\sum_{j=1}^{K}w_j^2\right) = \sum_{i=1}^{N}\sigma^2\|\vec{w}\|^2 = \sigma^2\|\vec{w}\|^2\cdot N$$

$$\Rightarrow \tilde{L}_{ssq} = \frac{1}{2N}\left(\sum_{i=1}^{N}(f_{\vec{w},b}(x_i)-y_i)^2+N\cdot\sigma^2\|\vec{w}\|^2\right)$$

$$= \frac{1}{2N}\sum_{i=1}^{N}(f_{\vec{w},b}(x_i)-y_i)^2+\frac{\sigma^2}{2}\|\vec{w}\|^2 \ \#$$

3.

a. $e_k = \frac{1}{N}\sum_{i=1}^{N}(g_k(\vec{x}_i)-y_i)^2$

$$= \frac{1}{N}\sum_{i=1}^{N}\left(g_k(\vec{x}_i)^2-2g_k(\vec{x}_i)\cdot y_i+y_i^2\right)$$

$$= \frac{1}{N}\sum_{i=1}^{N}(g_k(\vec{x}_i))^2-\frac{2}{N}\sum_{i=1}^{N}g_k(\vec{x}_i)\cdot y_i+\frac{1}{N}\sum_{i=1}^{N}y_i^2$$

$$= s_k-\frac{2}{N}\sum_{i=1}^{N}g_k(\vec{x}_i)\cdot y_i+e_0$$

$$\Rightarrow \sum_{i=1}^{N}g_k(\vec{x}_i)\cdot y_i = \frac{N}{2}(s_k-e_k+e_0) \ \#$$

b. 令 $G(\vec{x}_i) = \sum_{k=1}^{K}\alpha_k g_k(\vec{x}_i) = [g_1(\vec{x}_i)\ \cdots\ g_K(\vec{x}_i)]\begin{bmatrix}\alpha_1\\ \vdots\\ \alpha_K\end{bmatrix}$.

令 $\vec{G}_i = \begin{bmatrix}g_1(\vec{x}_i)\\ \vdots\\ g_K(\vec{x}_i)\end{bmatrix}\in\mathbb{R}^{K\times1}$, $\vec{\alpha} = \begin{bmatrix}\alpha_1\\ \vdots\\ \alpha_K\end{bmatrix}\in\mathbb{R}^{K\times1}$

则 $G(\vec{x}_i) = \vec{G}_i^T\vec{\alpha}$

$$\Rightarrow \underset{\alpha_1,\cdots,\alpha_K}{\operatorname{argmin}} \frac{1}{N}\sum_{i=1}^{N}\left(\sum_{k=1}^{K}\alpha_k g_k(\vec{x}_i) - y_i\right)^2$$

$$= \underset{\alpha_1,\cdots,\alpha_K}{\operatorname{argmin}} \frac{1}{N}\sum_{i=1}^{N}\left(G(x_i) - y_i\right)^2$$

$$= \underset{\alpha}{\operatorname{argmin}} \frac{1}{N}\sum_{i=1}^{N}\left(\vec{G}_i^T\vec{\alpha} - y_i\right)^2$$

$$= \underset{\alpha}{\operatorname{argmin}} \frac{1}{N}\sum_{i=1}^{N}(\vec{G}_i^T\vec{\alpha})^2 - \frac{2}{N}\sum_{i=1}^{N}(\vec{G}_i^T\vec{\alpha}\cdot y_i) + \underbrace{\frac{1}{N}\sum_{i=1}^{N} y_i^2}$$ 與 α 無關

$$= \underset{\alpha}{\operatorname{argmin}} \frac{1}{N}\sum_{i=1}^{N}(\vec{G}_i^T\vec{\alpha})^2 - \frac{2}{N}\sum_{i=1}^{N}(\vec{G}_i^T\vec{\alpha}\cdot y_i)$$

其中. $\sum_{i=1}^{N}(\vec{G}_i^T\vec{\alpha})^2 = \sum_{i=1}^{N}(\vec{G}_i^T\vec{\alpha})(\vec{G}_i^T\vec{\alpha})$

$$= \sum_{i=1}^{N}\vec{\alpha}^T\vec{G}_i\vec{G}_i^T\vec{\alpha}$$

$$= \vec{\alpha}^T\vec{G}_1\vec{G}_1^T\vec{\alpha} + \cdots + \vec{\alpha}^T\vec{G}_N\vec{G}_N^T\vec{\alpha}$$

$$= [\alpha_1 \cdots \alpha_K]\,[\vec{G}_1 \cdots \vec{G}_N]\begin{bmatrix}\vec{G}_1^T\\ \vdots\\ \vec{G}_N^T\end{bmatrix}\begin{bmatrix}\alpha_1\\ \vdots\\ \alpha_K\end{bmatrix}$$

而 $\sum_{i=1}^{N}(\vec{G}_i^T\vec{\alpha}\cdot y_i)$

$$= \vec{G}_1^T\vec{\alpha}\cdot y_1 + \cdots + \vec{G}_N^T\vec{\alpha}\cdot y_N$$

$$= \vec{\alpha}^T\vec{G}_1 y_1 + \cdots + \vec{\alpha}^T\vec{G}_N\cdot y_N$$

$$= [\alpha_1 \cdots \alpha_K]\,[\vec{G}_1 \cdots \vec{G}_N]\begin{bmatrix}y_1\\ \vdots\\ y_N\end{bmatrix}.$$

令 $X = \begin{bmatrix}\vec{G}_1^T\\ \vdots\\ \vec{G}_N^T\end{bmatrix} \in R^{N\times K}$, $\vec{y} = \begin{bmatrix}y_1\\ \vdots\\ y_N\end{bmatrix} \in R^{K\times 1}$

$$\Rightarrow \underset{\alpha}{\operatorname{argmin}} \frac{1}{N}\sum_{i=1}^{N}(\vec{G}_i^T\vec{\alpha})^2 - \frac{2}{N}\sum_{i=1}^{N}(\vec{G}_i^T\vec{\alpha}\cdot y_i)$$

$$= \underset{\alpha}{\operatorname{argmin}} \frac{1}{N}(\vec{\alpha}^T X^T X\vec{\alpha}) - \frac{2}{N}\vec{\alpha}^T X^T\vec{y}$$

將原式對 α 做偏微分

$$\frac{\partial}{\partial\alpha}\left(\frac{1}{N}(\vec{\alpha}^T X^T X\vec{\alpha}) - \frac{2}{N}\vec{\alpha}^T X^T\vec{y}\right).$$

$$= \frac{\partial}{\partial\alpha}\left(\frac{1}{N}\|X\vec{\alpha}\|^2 - \frac{2}{N}\vec{y}^T X\vec{\alpha}\right)$$

$$= \frac{2}{N}X^T X\vec{\alpha} - \frac{2}{N}X^T\vec{y}$$

求解 $\nabla_\alpha L_{test}\left(\sum_{k=1}^{K}\alpha_k g_k\right) = 0.$

$$\Rightarrow \frac{2}{N} X^T X \vec{\alpha} - \frac{2}{N} X^T \vec{y} = 0$$

$$\Rightarrow X^T X \vec{\alpha} = X^T \vec{y}.$$

$$\Rightarrow \vec{\alpha} = (X^T X)^{-1} X^T \vec{y}$$

$$\Rightarrow \begin{bmatrix} \alpha_1 \\ \vdots \\ \alpha_k \end{bmatrix} = \left(\begin{bmatrix} g_1(\vec{x}_1) & \cdots & g_1(\vec{x}_N) \\ \vdots & \ddots & \vdots \\ g_k(\vec{x}_1) & \cdots & g_k(\vec{x}_N) \end{bmatrix} \begin{bmatrix} g_1(\vec{x}_1) & \cdots & g_k(\vec{x}_1) \\ \vdots & \ddots & \vdots \\ g_1(\vec{x}_N) & \cdots & g_k(\vec{x}_N) \end{bmatrix} \right)^{-1} \begin{bmatrix} g_1(\vec{x}_1) & \cdots & g_1(\vec{x}_N) \\ \vdots & \ddots & \vdots \\ g_k(\vec{x}_1) & \cdots & g_k(\vec{x}_N) \end{bmatrix} \begin{bmatrix} y_1 \\ \vdots \\ y_N \end{bmatrix}$$

$$= \left(\begin{bmatrix} g_1(\vec{x}_1) & \cdots & g_1(\vec{x}_N) \\ \vdots & \ddots & \vdots \\ g_k(\vec{x}_1) & \cdots & g_k(\vec{x}_N) \end{bmatrix} \begin{bmatrix} g_1(\vec{x}_1) & \cdots & g_k(\vec{x}_1) \\ \vdots & & \vdots \\ g_1(\vec{x}_N) & \cdots & g_k(\vec{x}_N) \end{bmatrix} \right)^{-1} \cdot \frac{N}{2} \begin{bmatrix} s_1 - e_1 + e_0 \\ \vdots \\ s_k - e_k + e_0 \end{bmatrix} \#$$