Exercise 1: ridge regression $\min_{w\in\mathbb{R}^d, b\in\mathbb{R}} \frac{1}{2n}\|Xw + b1 - y\|_2^2 + \lambda\|w\|_2^2 \quad X\in\mathbb{R}^{n\times d}, y\in\mathbb{R}^n, \lambda>0$

1.

$$\frac{\partial}{\partial w} = \frac{\partial}{\partial w}\left[\frac{1}{2n}\|Xw + b1 - y\|_2^2 + \lambda\|w\|_2^2\right]$$

$$= \frac{\partial}{\partial w}\left[\frac{1}{2n}\|Xw + b1 - y\|_2^2\right] + \frac{\partial}{\partial w}\left[\lambda\|w\|_2^2\right]$$

$$= \underline{\frac{1}{2n}\frac{\partial}{\partial w}\left[\|Xw + b1 - y\|_2^2\right] + \lambda\frac{\partial}{\partial w}\left[\|w\|_2^2\right]} \quad \text{equation } (*)$$

To compute $\frac{\partial}{\partial w}\left[\|Xw + b1 - y\|_2^2\right]$:

Let k be an integer $\in [1, d]$.

$$\frac{\partial}{\partial w_k}\left[\|X_k w_k + b1_k - y_k\|_2^2\right] = \frac{\partial}{\partial w_k}\left[\sum_{i=1}^{n}(X_{ki}w_k + b1_k - y_k)^2\right] \text{ by Def}^n \text{ of Norm}$$

$$= \sum_{i=1}^{n}\frac{\partial}{\partial w_k}\left[(X_{ik}w_k + b1_k - y_k)^2\right]$$

$$= \sum_{i=1}^{n} 2(X_{ik}w_k + b1_k - y_k)\cdot X_{ki}$$

$$= 2\sum_{i=1}^{n} X_{ik}(X_{ki}w_k + b1_k - y_k)$$

$$= 2X_k^T(X_k w_k + b1_k - y_k)$$

So, $\underline{\frac{\partial}{\partial w}\left[\|Xw + b1 - y\|_2^2\right] = 2X^T(Xw + b1 - y)}$ ①

To compute $\frac{\partial}{\partial w}\left[\|w\|_2^2\right] = \frac{\partial}{\partial w}[w^T w]$

Let k be an integer $\in [1, d]$

$$w^T w = \sum_{i=1}^{d} w_k^2$$

$$\frac{\partial}{\partial w_k}(w^T w) = \frac{\partial}{\partial w_k}\left(\sum_{i=1}^{d} w_i^2\right) = \frac{\partial}{\partial w_k}(w_1^2 + \cdots + w_k^2 + \cdots + w_d^2)$$

$$= \frac{\partial}{\partial w_k}(w_k^2) = 2w_k$$

So, $\underline{\frac{\partial}{\partial w}[w^T w] = 2w}$ ②

Substituting ① & ② into equation $(*)$, we get:

$$\frac{\partial}{\partial w} = \frac{1}{2n}\cdot 2X^T(Xw + b1 - y) + \lambda\cdot(2w)$$

$$= \frac{1}{n}X^T(Xw + b1 - y) + 2\lambda w \quad \blacksquare$$

$$\frac{\partial}{\partial b} = \frac{\partial}{\partial b}\left[\frac{1}{2n}\|Xw + b\mathbf{1} - y\|_2^2 + \lambda\|w\|_2^2\right]$$

$$= \frac{\partial}{\partial b}\left[\frac{1}{2n}\|Xw + b\mathbf{1} - y\|_2^2\right] + \frac{\partial}{\partial b}\left[\lambda\|w\|_2^2\right]$$

$$= \frac{1}{2n}\cdot\frac{\partial}{\partial b}\left[\|Xw + b\mathbf{1} - y\|_2^2\right] + 0$$ equation (**)

To compute $\frac{\partial}{\partial b}\left[\|Xw + b\mathbf{1} - y\|_2^2\right]$:

$$\frac{\partial}{\partial b}\left[\|Xw + b\mathbf{1} - y\|_2^2\right] = \frac{\partial}{\partial b}\left[\sum_{i=1}^{n}\sum_{j=1}^{d}(x_{ij}w_j + b\mathbf{1}_j - y_i)^2\right]$$

$$= \frac{\partial}{\partial b}\left[\sum_{i=1}^{n}\sum_{j=1}^{d}(x_{ij}w_j + b - y_i)^2\right]$$

$$= \sum_{i=1}^{n}\sum_{j=1}^{d}\frac{\partial}{\partial b}\left[(x_{ij}w_j + b - y_i)^2\right]$$

$$= \sum_{i=1}^{n}\sum_{j=1}^{d} 2(x_{ij}w_j + b - y_i)$$

$$= 2\sum_{i=1}^{n}\sum_{j=1}^{d} 1\cdot(x_{ij}w_j + b - y_i)$$

$$= 2\mathbf{1}^T(Xw + b\mathbf{1} - y)$$ ③

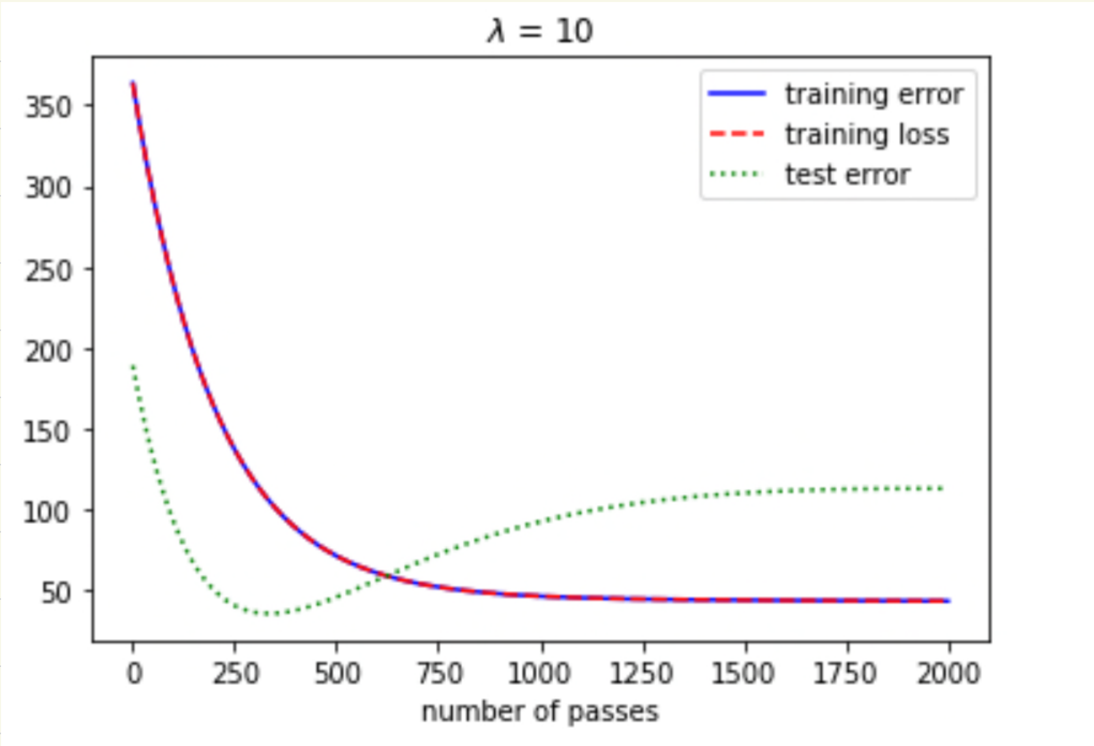
Substituting ③ into equation (**), we get

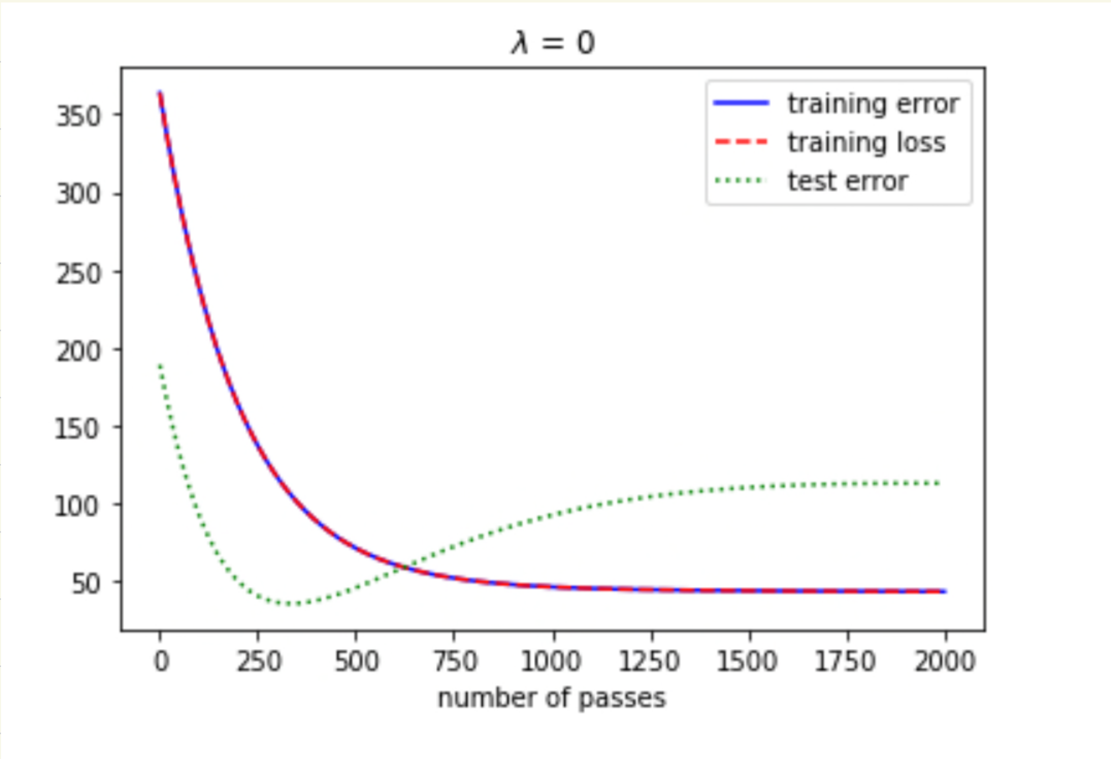
$$\frac{\partial}{\partial b} = \frac{1}{2n}\cdot 2\mathbf{1}^T(Xw + b\mathbf{1} - y) + 0$$

$$= \frac{1}{n}\mathbf{1}^T(Xw + b\mathbf{1} - y)$$ ☒

2. The implementations are shown in the "code_part.ipynb".

The result is shown as below:





3. Given w, solve $\frac{\partial}{\partial b} = \frac{1}{n}\mathbf{1}^T(Xw + b\mathbf{1} - y) = 0$ for b:

$$\Rightarrow \mathbf{1}^T(Xw + b\mathbf{1} - y) = 0 \quad \text{where } 0 \in \mathbb{R}$$

$$\Rightarrow \langle \mathbf{1}, Xw + b\mathbf{1} - y \rangle = 0 \quad \text{where } \mathbf{1} \in \mathbb{R}^d$$

$$\Rightarrow \langle \mathbf{1}, Xw \rangle + \langle \mathbf{1}, b\mathbf{1} \rangle - \langle \mathbf{1}, y \rangle = 0$$

$$\Rightarrow \langle \mathbf{1}, Xw \rangle + b\langle \mathbf{1}, \mathbf{1} \rangle - \langle \mathbf{1}, y \rangle = 0$$

$$\Rightarrow b\langle \mathbf{1}, \mathbf{1} \rangle = \langle \mathbf{1}, y \rangle - \langle \mathbf{1}, Xw \rangle$$

$$\Rightarrow b\,\|\mathbf{1}\|_2^2 = \langle \mathbf{1}, y \rangle - \langle \mathbf{1}, Xw \rangle \quad \text{since } \langle \mathbf{1}, \mathbf{1} \rangle = \mathbf{1}^T\mathbf{1} = \|\mathbf{1}\|_2^2$$

$$\Rightarrow b = \frac{\langle \mathbf{1}, y \rangle - \langle \mathbf{1}, Xw \rangle}{\|\mathbf{1}\|_2^2} = \frac{\langle \mathbf{1}, y - Xw \rangle}{\|\mathbf{1}\|_2^2}$$

The re-implementations are shown in the "code_part.ipynb".

The modification does not converge to the same solution.

4. Given :

$$\frac{\partial}{\partial w} = \frac{1}{n} X^T(Xw + b\mathbf{1} - y) + 2\lambda w$$

$$\frac{\partial}{\partial b} = \frac{1}{n} \mathbf{1}^T(Xw + b\mathbf{1} - y)$$

and $X^T\mathbf{1} = 0$, $\mathbf{1}^T y = 0$

Let $\frac{\partial}{\partial b} = \frac{1}{n} \mathbf{1}^T(Xw + b\mathbf{1} - y) = 0.$

$\Rightarrow \quad \mathbf{1}^T(Xw + b\mathbf{1} - y) = 0$

$\Rightarrow \quad \mathbf{1}^T X w + \mathbf{1}^T b\mathbf{1} - \mathbf{1}^T y = 0$

$\Rightarrow \quad X^T\mathbf{1} w + \mathbf{1}^T b\mathbf{1} - \mathbf{1}^T y = 0$ since $\mathbf{1}^T X = \langle X, \mathbf{1} \rangle = \langle \mathbf{1}, X \rangle = X^T\mathbf{1}$

$\Rightarrow \quad \mathbf{1}^T b \mathbf{1} = 0$ since $X^T\mathbf{1} = 0$, $\mathbf{1}^T y = 0$

$\Rightarrow \quad b\, \mathbf{1}^T\mathbf{1} = 0$ since $b \in \mathbb{R}$

$\Rightarrow \quad b \|\mathbf{1}\|_2^2 = 0$ since $\mathbf{1}^T\mathbf{1} = \|\mathbf{1}\|_2^2$

$\Rightarrow \quad b\, n = 0$ where $n \in \mathbb{R}$, $n > 0$

$\Rightarrow \quad b = 0$

Therefore, $b=0$ is the optimum value.

I verify the results by running the code in "code-part.ipynb".

Exercise 2 :

1. Hyperplane $\partial H_{w,b} = \{ x \in \mathbb{R}^d : \langle w, x \rangle + b = 0 \}$ $\quad x \in \mathbb{R}^d, \ b \in \mathbb{R}$

$$= \{ x \in \mathbb{R}^d : w^T x + b = 0 \}$$

To solve the problem $\min_{x \in \partial H_{w,b}} \frac{1}{2} \| x - z \|_2^2$ where $z \in \mathbb{R}^d$:

① Convert the problem to : $\min_{x \in \mathbb{R}^d} \frac{1}{2} \| x - z \|_2^2$ such that $w^T x + b = 0$

② Lagrangial dual is : $\max_{\lambda \in \mathbb{R}} \min_{x \in \mathbb{R}^d} L(x, \lambda)$

where $L(x, \lambda) = \frac{1}{2} \| x - z \|_2^2 + \lambda (w^T x + b)$

③ Solve the inner minimization problem: $f(\lambda) = \arg\min_{x \in \mathbb{R}^d} L(x, \lambda)$

$$\nabla L(x, \lambda) = \frac{1}{2} \cdot 2 (x - z) + \lambda (w^T)^T$$

$$= (x - z) + \lambda w$$

Applying Fermat's Condition :

Let $\nabla L(x, \lambda) = (x - z) + \lambda w = 0$

$$\underline{x^* = z - \lambda w}$$

So, $f(\lambda) = L(x^*, \lambda)$

$$= \frac{1}{2} \| x^* - z \|_2^2 + \lambda (w^T x^* + b)$$

$$= \frac{1}{2} \| z - \lambda w - z \|_2^2 + \lambda (w^T (z - \lambda w) + b)$$

$$= \frac{1}{2} \| \lambda w \|_2^2 + \lambda (w^T z - w^T \lambda w + b)$$

$$= \frac{1}{2} \lambda^2 \| w \|_2^2 + \lambda w^T z - \lambda^2 w^T w + \lambda b \quad \text{since } \lambda \in \mathbb{R}$$

$$= \frac{1}{2} \lambda^2 \| w \|_2^2 + \lambda w^T z - \lambda^2 \| w \|_2^2 + \lambda b \quad \text{since } w \in \mathbb{R}^d$$

$$\Rightarrow \| w \|_2^2 = w^T w$$

④ Solve the outer maximization problem: $\max_{\alpha \in \mathbb{R}} f(\alpha)$

$$\nabla f(\alpha) = \frac{1}{2}\|w\|_2^2 \cdot 2\alpha + w^T z - \|w\|_2^2 \cdot 2\alpha + b$$
$$= \alpha \|w\|_2^2 + w^T z - 2\alpha \|w\|_2^2 + b$$
$$= w^T z + b - \alpha \|w\|_2^2$$

Applying Fermat's Condition:

Let $\nabla f(\alpha) = w^T z + b - \alpha \|w\|_2^2 = 0$

$$w^T z + b = \alpha \|w\|_2^2$$

$$\alpha^* = \frac{w^T z + b}{\|w\|_2^2} = \frac{\langle w, z\rangle + b}{\|w\|_2^2}$$ since $\langle w, z\rangle = w^T z$

⑤ Plugging α^* into $x^* = z - \alpha w$:

$$x^* = z - \alpha^* w$$
$$= z - \frac{\langle w, z\rangle + b}{\|w\|_2^2} w$$

⑥ $$\|x^* - z\|_2 = \left\| z - \frac{\langle w, z\rangle + b}{\|w\|_2^2} w - z \right\|_2$$

$$= \left\| \frac{(\langle w, z\rangle + b) w}{\|w\|_2^2} \right\|_2$$

$$= \left\| \frac{(\langle w, z\rangle + b) w}{w^T w} \right\|_2$$ since $w \in \mathbb{R}^d$, $\|w\|_2^2 = w^T w$

$$= \sqrt{\left(\frac{\langle w, z\rangle + b}{w^T w} w\right)^T \left(\frac{\langle w, z\rangle + b}{w^T w} w\right)}$$

$$= \sqrt{\frac{\langle w, z\rangle + b}{w^T w} w^T \frac{\langle w, z\rangle + b}{w^T w} w}$$

$$= \sqrt{\frac{(\langle w, z \rangle + b)^2 \; w^T w}{w^T w \; w^T w}} \quad \text{since } \langle w, z \rangle + b \in \mathbb{R}$$

$$= \sqrt{\frac{(\langle w, z \rangle + b)^2 \; w^T w \; (w^T w)^{-1}}{w^T w \; w^T w \; (w^T w)^{-1}}}$$

$$= \sqrt{\frac{(\langle w, z \rangle + b)^2}{w^T w}}$$

$$= \sqrt{\frac{(\langle w, z \rangle + b)^2}{\| w \|_2^2}}$$

$$= \sqrt{\left(\frac{\langle w, z \rangle + b}{\| w \|_2} \right)^2}$$

$$= \frac{| \langle w, z \rangle + b |}{\| w \|_2}$$

Therefore, the distance from the given point $z \in \mathbb{R}^d$ to the hyperplane is

$$\frac{| \langle w, z \rangle + b |}{\| w \|_2}$$

2. Halfplane: $H_{w,b} = \{x \in \mathbb{R}^d : \langle w, x \rangle + b \le 0\}$

$= \{x \in \mathbb{R}^d : w^T x + b \le 0\}$

To solve the problem $\min_{x \in H_{w,b}} \frac{1}{2}\|x - z\|_2^2$ where $z \in \mathbb{R}^d$:

① Convert the problem to : $\min_{x \in \mathbb{R}^d} \frac{1}{2}\|x - z\|_2^2$ such that $w^T x + b \le 0$

② Lagrangial dual is : $\max_{\lambda \in \mathbb{R}_+} \min_{x \in \mathbb{R}^d} L(x, \lambda)$

where $L(x,\lambda) = \frac{1}{2}\|x - z\|_2^2 + \lambda(w^T x + b)$

③ Solve the inner minimization problem: $f(\lambda) = \arg\min_{x \in \mathbb{R}^d} L(x, \lambda)$

$$\nabla L(x,\lambda) = \frac{1}{2} \cdot 2(x - z) + \lambda (w^T)^T$$

$$= (x - z) + \lambda w$$

Applying Fermat's Condition:

Let $\nabla L(x,\lambda) = (x - z) + \lambda w = 0$

$$x^* = z - \lambda w$$

So, $f(\lambda) = L(x^*, \lambda)$

$$= \frac{1}{2}\|x^* - z\|_2^2 + \lambda(w^T x^* + b)$$

$$= \frac{1}{2}\| z - \lambda w - z\|_2^2 + \lambda(w^T(z - \lambda w) + b)$$

$$= \frac{1}{2}\|\lambda w\|_2^2 + \lambda(w^T z - w^T \lambda w + b)$$

$$= \frac{1}{2}\lambda^2 \|w\|_2^2 + \lambda w^T z - \lambda^2 w^T w + \lambda b \quad \text{since } \lambda \in \mathbb{R}$$

$$= \frac{1}{2}\lambda^2 \|w\|_2^2 + \lambda w^T z - \lambda^2 \|w\|_2^2 + \lambda b \quad \text{since } w \in \mathbb{R}^d$$

$$\Rightarrow \|w\|_2^2 = w^T w$$

④ Solve the outer maximization problem: $\max_{\alpha \in \mathbb{R}_+} f(\alpha)$

$$\nabla f(\alpha) = \frac{1}{2}\|w\|_2^2 \cdot 2\alpha + w^T z - \|w\|_2^2 \cdot 2\alpha + b$$
$$= \alpha\|w\|_2^2 + w^T z - 2\alpha\|w\|_2^2 + b$$
$$= w^T z + b - \alpha\|w\|_2^2$$

Applying Fermat's Condition:

Let $\nabla f(\alpha) = w^T z + b - \alpha\|w\|_2^2 = 0$

$$w^T z + b = \alpha\|w\|_2^2$$
$$\alpha^* = \frac{w^T z + b}{\|w\|_2^2} = \frac{\langle w, z\rangle + b}{\|w\|_2^2}$$

⑤ Solve $\alpha^* \geq 0$

Since $\alpha \in \mathbb{R}_+$, $\alpha \geq 0$.

We need $\alpha^* = \dfrac{\langle w, z\rangle + b}{\|w\|_2^2} \geq 0$.

$\Rightarrow \langle w, z\rangle + b \geq 0$ since $\|w\|_2^2 \geq 0$

Case 1 : If $\langle w, z\rangle + b \geq 0$, we know $\alpha^* \geq 0$
and plugging $\alpha^* = \dfrac{\langle w, z\rangle + b}{\|w\|_2^2}$ into x^* and
then find the distance is $\dfrac{|\langle w, z\rangle + b|}{\|w\|_2}$
(shown in part 1 step ⑤ & ⑥).

Case 2 : If $\langle w, z\rangle + b < 0$, z is on the halfplane $H_{w,b}$
by the definition of halfplane.
So, the distance from z to halfplane is 0.

Therefore, the distance from the given point $z \in \mathbb{R}^d$ to the halfplane $H_{w,b}$ is

$$\begin{cases} 0 & \text{if } z \text{ is in } H_{w,b} \quad (\langle w, z\rangle + b < 0) \\ \dfrac{|\langle w, z\rangle + b|}{\|w\|_2^2} = \dfrac{\langle w, z\rangle + b}{\|w\|_2^2} & \text{if } z \text{ is not in } H_{w,b} \quad (\langle w, z\rangle + b \geq 0) \end{cases}$$

3. To solve this question, we need to solve $\min_{y_i(\langle w, x_i\rangle + b) \geq 1} \frac{1}{2}\|w\|_2^2$

⓪ Constraint $y_i(\langle w_i, x_i\rangle + b) \geq 1 \Rightarrow 1 - y_i(\langle w, x_i\rangle + b) \leq 0$

① Lagrangial dual is :

$$\max_{\alpha_1, \alpha_2 \in \mathbb{R}_+} \min_{\substack{w \in \mathbb{R}^d \\ b \in \mathbb{R}}} L(w, b, \alpha_1, \alpha_2) \quad \text{where}$$

$$L(w, b, \alpha_1, \alpha_2) = \frac{1}{2}\|w\|_2^2 + \alpha_1[1 - y_1(\langle w, x_1\rangle + b)] + \alpha_2[1 - y_2(\langle w, x_2\rangle + b)]$$
$$= \frac{1}{2}\|w\|_2^2 + \alpha_1[1 - y_1(w^T x_1 + b)] + \alpha_2[1 - y_2(w^T x_2 + b)]$$

② Solve the minimization Problem: $w^T x_1$ $w^T x_1$

$w^T x_1$

$$\frac{\partial L}{\partial w} = \frac{1}{2} \cdot 2w + \frac{\partial}{\partial w}[\alpha_1 - \alpha_1 y_1 w^T x_1 - \alpha_1 y_1 b]$$
$$+ \frac{\partial}{\partial w}[\alpha_2 - \alpha_2 y_2 w^T x_2 - \alpha_2 y_2 b]$$

$$= w + (-\alpha_1 y_1 x_1^T)^T + (-\alpha_2 y_2 x_2^T)^T$$
$$= w - \alpha_1 y_1 x_1 - \alpha_2 y_2 x_2 \quad \text{since } \alpha_1, \alpha_2, y_1, y_2 \in \mathbb{R}$$

$$\frac{\partial L}{\partial b} = \frac{\partial}{\partial b}[\alpha_1 - \alpha_1 y_1 w^T x_1 - \alpha_1 y_1 b] + \frac{\partial}{\partial b}[\alpha_2 - \alpha_2 y_2 w^T x_2 - \alpha_2 y_2 b]$$
$$= -\alpha_1 y_1 - \alpha_2 y_2$$

Applying Fermat's Condition:

Let $\frac{\partial L}{\partial w} = 0$ and $\frac{\partial L}{\partial b} = 0$.

$$\frac{\partial L}{\partial w} = 0 \Rightarrow w - \alpha_1 y_1 x_1 - \alpha_2 y_2 x_2 = 0$$
$$\Rightarrow w^* = \alpha_1 y_1 x_1 + \alpha_2 y_2 x_2 \quad ①$$

$$\frac{\partial L}{\partial b} = 0 \Rightarrow -\alpha_1 y_1 - \alpha_2 y_2 = 0$$
$$\Rightarrow \alpha_1 y_1 + \alpha_2 y_2 = 0 \quad ②$$

③ Plugging ① into $L(L, b, \alpha_1, \alpha_2)$:

$$f(\alpha_1, \alpha_2) = L(w^*, b, \alpha_1, \alpha_2) = \frac{1}{2}\|w^*\|_2^2 + \alpha_1[1 - y_1(w^{*T}x_1 + b)] + \alpha_2[1 - y_2(w^{*T}x_2 + b)]$$

$$= \frac{1}{2}\|\alpha_1 y_1 x_1 + \alpha_2 y_2 x_2\|_2^2 + \alpha_1 - \alpha_1 y_1[(\alpha_1 y_1 x_1 + \alpha_2 y_2 x_2)^T x_1 + b] + \alpha_2 - \alpha_2 y_2[(\alpha_1 y_1 x_1 + \alpha_2 y_2 x_2)^T x_2 + b]$$

$$= \frac{1}{2}[\alpha_1^2 y_1^2 \|x_1\|_2^2 + \alpha_2^2 y_2^2 \|x_2\|_2^2] + \alpha_1 - \alpha_1 y_1[(\alpha_1 y_1 x_1^T x_1 + \alpha_2 y_2 x_2^T x_1) + b] + \alpha_2 - \alpha_2 y_2[(\alpha_1 y_1 x_1^T x_2 + \alpha_2 y_2 x_2^T x_2) + b]$$

$$= \frac{1}{2}\alpha_1^2 y_1^2 x_1^T x_1 + \frac{1}{2}\alpha_2^2 y_2^2 x_2^T x_2 + \alpha_1 - \alpha_1^2 y_1^2 x_1^T x_1 - \alpha_1 y_1 \alpha_2 y_2 x_2^T x_1 - \alpha_1 y_1 b + \alpha_2 - \alpha_2 y_2 \alpha_1 y_1 x_1^T x_2 - \alpha_2^2 y_2^2 x_2^T x_2 - \alpha_2 y_2 b$$

$$= -\frac{1}{2}\alpha_1^2 y_1^2 x_1^T x_1 - \frac{1}{2}\alpha_2^2 y_2^2 x_2^T x_2 - \alpha_1 y_1 \alpha_2 y_2 x_2^T x_1 - \alpha_1 y_1 \alpha_2 y_2 x_1^T x_2 - (\alpha_1 y_1 + \alpha_2 y_2) b + \alpha_1 + \alpha_2$$

$$= -\frac{1}{2}\alpha_1^2 y_1^2 x_1^T x_1 - \frac{1}{2}\alpha_2^2 y_2^2 x_2^T x_2 - \alpha_1 y_1 \alpha_2 y_2 x_2^T x_1 - \alpha_1 y_1 \alpha_2 y_2 x_1^T x_2 + \alpha_1 + \alpha_2$$ since equation ②

$= (*)$

Since $\alpha_1 y_1 + \alpha_2 y_2 = 0$, $\alpha_1 y_1 = -\alpha_2 y_2$. and $\alpha_2 = -\frac{\alpha_1 y_1}{y_2}$

Sub. into equation $(*)$:

$$f(\alpha_1, \alpha_2) = -\frac{1}{2}\alpha_1^2 y_1^2 x_1^T x_1 - \frac{1}{2}\alpha_1^2 y_1^2 x_2^T x_2 + \alpha_1^2 y_1^2 x_2^T x_1 + \alpha_1^2 y_1^2 x_1^T x_2 + \alpha_1 - \frac{\alpha_1 y_1}{y_2}$$

④ Solve the maximization problem: $\max\limits_{\alpha_1, \alpha_2 \geq 0} f(\alpha_1, \alpha_2)$.

$$\frac{\partial f}{\partial \alpha_1} = -\frac{1}{2} \times 2\alpha_1 y_1^2 x_1^T x_1 - \frac{1}{2} \times 2\alpha_1 y_1^2 x_2^T x_2 + 2\alpha_1 y_1^2 x_2^T x_1 + 2\alpha_1 y_1^2 x_1^T x_2 + 1 - \frac{y_1}{y_2}$$

Sub. $y_1 = 1$ and $y_2 = -1$:

$$\frac{\partial f}{\partial \alpha_1} = -\alpha_1 x_1^T x_1 - \alpha_1 x_2^T x_2 + \alpha_1 x_2^T x_1 + \alpha_1 x_1^T x_2 + 1 + 1$$

Applying Fermat's Condition.

Let $\frac{\partial f}{\partial \alpha_1} = 0$.

$$-\alpha_1 (x_1^T x_1 + x_2^T x_2 - x_2^T x_1 - x_1^T x_2) + 2 = 0$$

$$\alpha_1 (x_1^T x_1 + x_2^T x_2 - x_2^T x_1 - x_1^T x_2) - 2 = 0$$

$$\alpha_1 (x_1^T x_1 + x_2^T x_2 - x_2^T x_1 - x_1^T x_2) = 2$$

$$\alpha_1 (x_1^T - x_2^T) (x_1 - x_2) = 2 \quad \text{since } (x_1^T - x_2^T) (x_1 - x_2) = x_1^T x_1 - x_2^T x_1 - x_1^T x_2 + x_2^T x_2$$

$$\alpha_1 (x_1 - x_2)^T (x_1 - x_2) = 2$$

$$\alpha_1 \| x_1 - x_2 \|_2^2 = 2 \quad \text{since } \| x_1 - x_2 \|_2^2 = (x_1 - x_2)^T (x_1 - x_2)$$

$$\alpha_1^* = \frac{2}{\| x_1 - x_2 \|_2^2}$$

$$\alpha_2^* = \frac{-\alpha_1^* y_1}{y_2} = \alpha_1^* = \frac{2}{\| x_1 - x_2 \|_2^2}$$

⑤ Sub. α_1^* and α_2^* into $w^* = \alpha_1 y_1 x_1 + \alpha_2 y_2 x_2$:

$$w^* = \alpha_1^* y_1 x_1 - \alpha_2^* y_2 x_2$$

$$= \frac{2}{\| x_1 - x_2 \|_2^2} y_1 x_1 + \frac{2}{\| x_1 - x_2 \|^2} y_2 x_2$$

$$= \frac{2}{\| x_1 - x_2 \|_2^2} x_1 - \frac{2}{\| x_1 - x_2 \|_2^2} x_2 \quad \text{since } y_1 = 1, y_2 = -1$$

⑥ Since $\alpha_1 [y_1 (w^T x_1 + b) - 1] = 0$

$$y_1 (w^T x_1 + b) - 1 = 0$$

$$y_1 w^T x_1 + y_1 b - 1 = 0$$

$$b^* = \frac{1 - y_1 w^{*T} x_1}{y_1} = 1 - w^{*T} x_1 = 1 - \langle w^*, x_1 \rangle$$

So, the hyperplane is : $\{ x \in \mathbb{R}^d : \langle w^*, x \rangle + b^* = 0 \}$

From part 1, the distance from the given point x_1 to the hyperplane is

$$\frac{|\langle w^*, x_1\rangle + b|}{\|w^*\|_2} = \frac{|\langle w^*, x_1\rangle + 1 - \langle w^*, x_1\rangle|}{\|w^*\|_2}$$

$$= \frac{1}{\|w^*\|_2}$$

$$= \frac{1}{\left\| \frac{2}{\|x_1 - x_2\|_2^2} x_1 - \frac{2}{\|x_1 - x_2\|_2^2} x_2 \right\|_2}$$

$$= \frac{1}{\left\| \frac{2(x_1 - x_2)}{\|x_1 - x_2\|_2^2} \right\|_2}$$

$$= \frac{1}{\frac{2}{\|x_1 - x_2\|_2^2} \|x_1 - x_2\|_2}$$

$$= \frac{1}{\frac{2}{\|x_1 - x_2\|_2}}$$

$$= \frac{\|x_1 - x_2\|_2}{2}$$

Therefore, the margin of the binary dataset

= the distance from x_1 to the hyperplane

$$= \frac{\|x_1 - x_2\|_2}{2}$$