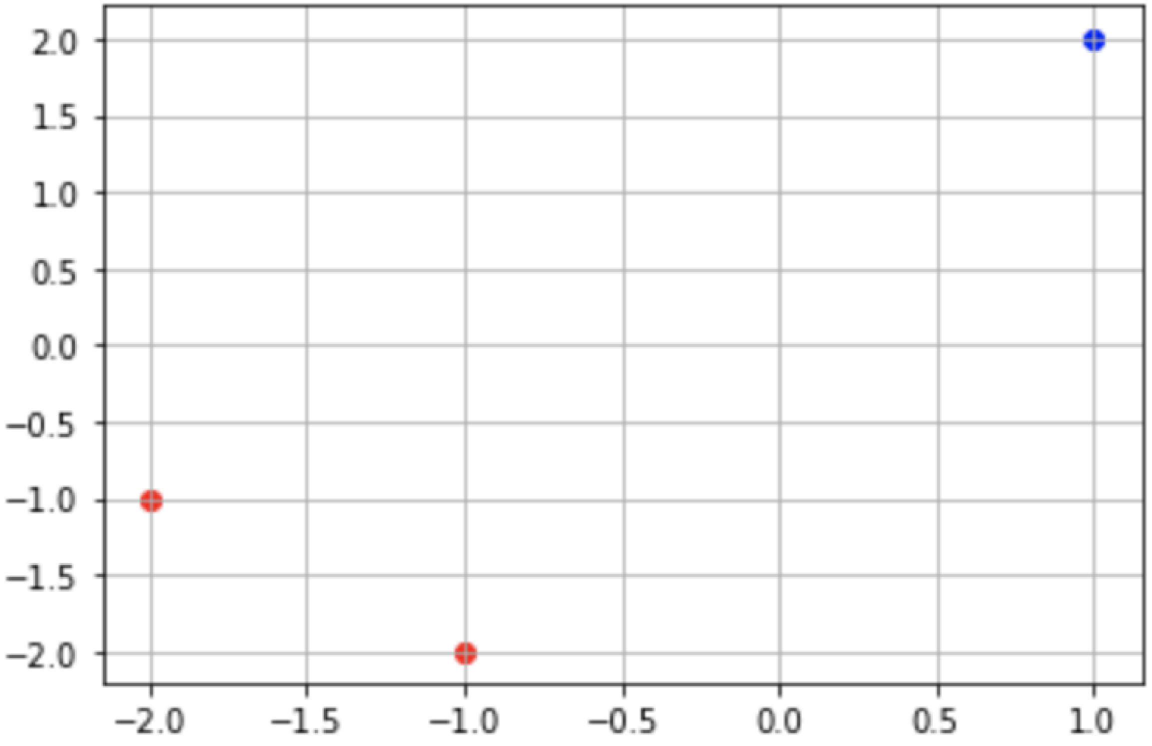
SUPPORT VECTOR MACHINES

EXERCISE 1

EXERCICE 1

Consider two classes: the class C_1: $\left\{\mathbf{x}^{(1)} = \begin{pmatrix}1\\2\end{pmatrix}\right\}$ with label 1 and C_2: $\left\{\mathbf{x}^{(2)} = \begin{pmatrix}-1\\-2\end{pmatrix}, \mathbf{x}^{(3)} = \begin{pmatrix}-2\\-1\end{pmatrix}\right\}$ with label -1

(a) Plot the points and write in parametric form the Support Vector Machine (SVM) classifier $g(\mathbf{x})$.

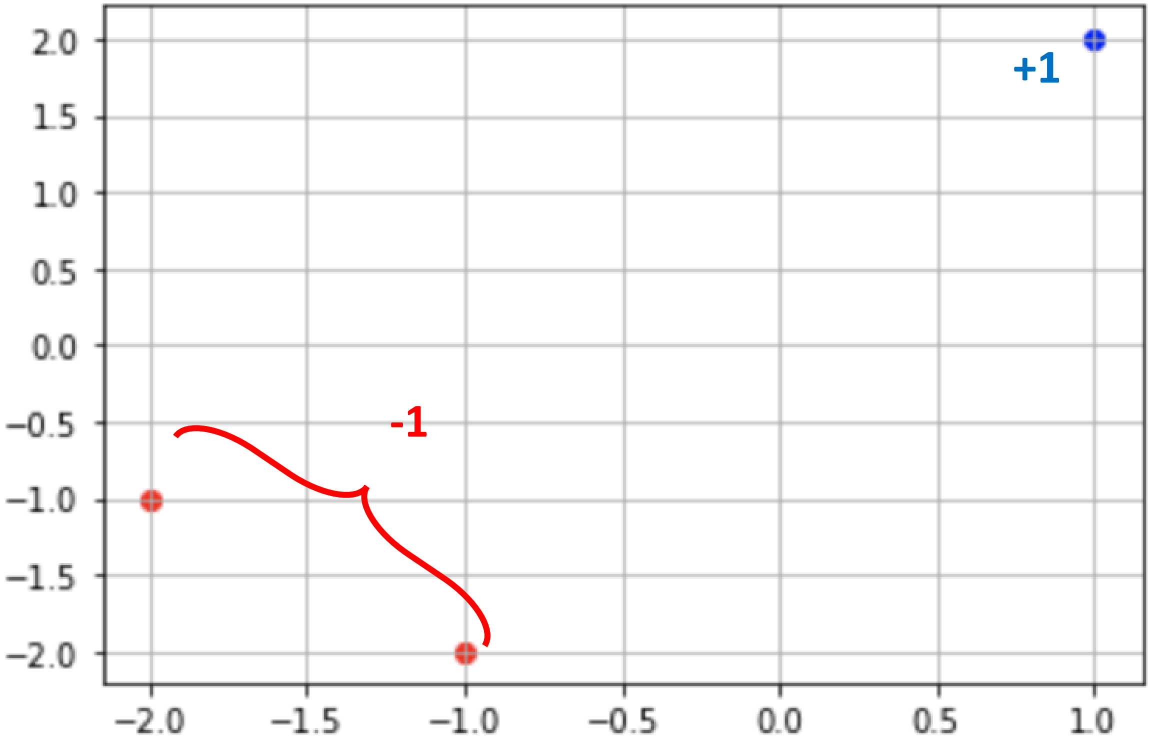


$g(\mathbf{x}) =$

EXERCICE 1

Consider two classes: the class C_1: $\left\{\mathbf{x}^{(1)} = \begin{pmatrix} 1 \\ 2 \end{pmatrix}\right\}$ with label 1 and C_2: $\left\{\mathbf{x}^{(2)} = \begin{pmatrix} -1 \\ -2 \end{pmatrix}, \mathbf{x}^{(3)} = \begin{pmatrix} -2 \\ -1 \end{pmatrix}\right\}$ with label -1

(a) Plot the points and write in parametric form the Support Vector Machine (SVM) classifier $g(\mathbf{x})$.



The generic expression of the classifier will be:

$$g\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = w_1 x_1 + w_2 x_2 + b$$

Let's train a SVM model!

(i.e. find the value of the parameters)

EXERCICE 1

Consider two classes: the class C_1: $\left\{\mathbf{x}^{(1)} = \begin{pmatrix}1\\2\end{pmatrix}\right\}$ with label 1 and C_2: $\left\{\mathbf{x}^{(2)} = \begin{pmatrix}-1\\-2\end{pmatrix}, \mathbf{x}^{(3)} = \begin{pmatrix}-2\\-1\end{pmatrix}\right\}$ with label -1

(b) Write the Dual Lagrangian $\mathcal{L}_D$ and derive the values of $\alpha_i, i = 1, 2, 3$

REMINDER Dual problem:

$$\mathcal{L}_D(\alpha) = \sum_{i=1}^{N} \alpha_i - \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N} \alpha_i \alpha_j y_i y_j {\mathbf{x}^{(i)}}^{\mathrm{T}} \mathbf{x}^{(j)}$$

Subject to $\alpha_i \geq 0 \quad \forall i$ and $\sum_{i=1}^{N} \alpha_i y_i = 0$

Then, the classifier is

$$g\begin{pmatrix}x_1\\x_2\end{pmatrix} = w_1 x_1 + w_2 x_2 + b$$

where $\boldsymbol{w} = \sum_{i=1}^{N} \alpha_i y_i \mathbf{x}^{(i)}$

EXERCICE 1

Consider two classes: the class C_1: $\left\{\mathbf{x}^{(1)} = \begin{pmatrix}1\\2\end{pmatrix}\right\}$ with label 1 and C_2: $\left\{\mathbf{x}^{(2)} = \begin{pmatrix}-1\\-2\end{pmatrix}, \mathbf{x}^{(3)} = \begin{pmatrix}-2\\-1\end{pmatrix}\right\}$ with label -1

(b) Write the Dual Lagrangian $\mathcal{L}_D$ and derive the values of $\alpha_i, i = 1, 2, 3$

$$\mathcal{L}_D(\alpha) = \sum_{i=1}^{N} \alpha_i - \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N} \alpha_i \alpha_j y_i y_j \mathbf{x}^{(i)^{\mathrm{T}}} \mathbf{x}^{(j)} \qquad \text{subject to} \qquad \alpha_i \geq 0 \quad \forall i \qquad \text{and} \qquad \sum_{i=1}^{N} \alpha_i y_i = 0$$

$$\mathcal{L}_D(\alpha) = ?$$

Let’s write it in matricial form

EXERCICE 1

$$\mathcal{L}_D(\alpha) = \sum_{i=1}^{N} \alpha_i - \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} \alpha_i \alpha_j y_i y_j \mathbf{x}^{(i)^{\mathrm{T}}} \mathbf{x}^{(j)}$$

$$= \alpha_1 + \alpha_2 + \cdots + \alpha_N = \begin{pmatrix} 1 & 1 & \cdots & 1 \end{pmatrix} \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_N \end{pmatrix}$$

EXERCICE 1

$$\mathcal{L}_D(\alpha) = \sum_{i=1}^{N} \alpha_i - \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N} \alpha_i \alpha_j y_i y_j \mathbf{x}^{(i)^{\mathrm{T}}} \mathbf{x}^{(j)}$$

$$\alpha_i \alpha_j \underbrace{y_i y_j \mathbf{x}^{(i)^{\mathrm{T}}} \mathbf{x}^{(j)}}_{= a_{ij}} = \alpha_i \cdot a_{ij} \cdot \alpha_j \implies \sum_{i=1}^{N}\sum_{j=1}^{N} \alpha_i \cdot a_{ij} \cdot \alpha_j = \begin{pmatrix} \alpha_1 & \cdots & \alpha_N \end{pmatrix} \begin{pmatrix} a_{11} & \cdots & a_{1N} \\ \vdots & \ddots & \vdots \\ a_{N1} & \cdots & a_{NN} \end{pmatrix} \begin{pmatrix} \alpha_1 \\ \vdots \\ \alpha_N \end{pmatrix}$$

EXERCICE 1

$$\mathcal{L}_D(\alpha) = \sum_{i=1}^{N} \alpha_i - \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N} \alpha_i\alpha_j y_i y_j {\mathbf{x}^{(i)}}^{\mathrm{T}}\mathbf{x}^{(j)}$$

$$\alpha_i\alpha_j y_i y_j {\mathbf{x}^{(i)}}^{\mathrm{T}}\mathbf{x}^{(j)} = \alpha_i \cdot a_{ij} \cdot \alpha_j \quad \Longrightarrow \quad \sum_{i=1}^{N}\sum_{j=1}^{N} \alpha_i \cdot a_{ij} \cdot \alpha_j = \begin{pmatrix} \alpha_1 & \cdots & \alpha_N \end{pmatrix} \begin{pmatrix} a_{11} & \cdots & a_{1N} \\ \vdots & \ddots & \vdots \\ a_{N1} & \cdots & a_{NN} \end{pmatrix} \begin{pmatrix} \alpha_1 \\ \vdots \\ \alpha_N \end{pmatrix}$$

$$a_{ij} = y_i y_j {\mathbf{x}^{(i)}}^{\mathrm{T}}\mathbf{x}^{(j)} = \begin{pmatrix} - & y_i {\mathbf{x}^{(i)}}^{\mathrm{T}} & - \end{pmatrix} \begin{pmatrix} | \\ y_j \mathbf{x}^{(j)} \\ | \end{pmatrix}$$

$$\Longrightarrow \quad \mathbf{A} = \begin{pmatrix} a_{11} & \cdots & a_{1N} \\ \vdots & \ddots & \vdots \\ a_{N1} & \cdots & a_{NN} \end{pmatrix} = \begin{pmatrix} - & y_1 {\mathbf{x}^{(1)}}^{\mathrm{T}} & - \\ & \vdots & \\ - & y_N {\mathbf{x}^{(N)}}^{\mathrm{T}} & - \end{pmatrix} \begin{pmatrix} | & & | \\ y_1 \mathbf{x}^{(1)} & \cdots & y_n \mathbf{x}^{(N)} \\ | & & | \end{pmatrix}$$

EXERCICE 1

$$\mathcal{L}_D(\alpha) = \sum_{i=1}^{N} \alpha_i - \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} \alpha_i \alpha_j y_i y_j {\mathbf{x}^{(i)}}^{\mathrm{T}} \mathbf{x}^{(j)}$$

$$= \begin{pmatrix} \alpha_1 & \cdots & \alpha_N \end{pmatrix} \underbrace{\boxed{\begin{pmatrix} - & y_1 {\mathbf{x}^{(1)}}^{\mathrm{T}} & - \\ & \vdots & \\ - & y_N {\mathbf{x}^{(N)}}^{\mathrm{T}} & - \end{pmatrix} \begin{pmatrix} | & & | \\ y_1 \mathbf{x}^{(1)} & \cdots & y_n \mathbf{x}^{(N)} \\ | & & | \end{pmatrix}}}_{\mathbf{A}} \begin{pmatrix} \alpha_1 \\ \vdots \\ \alpha_N \end{pmatrix}$$

EXERCICE 1

$$\mathcal{L}_D(\alpha) = \sum_{i=1}^{N} \alpha_i - \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N} \alpha_i \alpha_j y_i y_j \mathbf{x}^{(i)^{\mathrm{T}}} \mathbf{x}^{(j)}$$

$$\mathcal{L}_D(\alpha) = \begin{pmatrix} 1 & 1 & \cdots & 1 \end{pmatrix} \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_N \end{pmatrix} - \frac{1}{2}\left[\begin{pmatrix} \alpha_1 & \cdots & \alpha_N \end{pmatrix} \underbrace{\begin{pmatrix} - & y_1 \mathbf{x}^{(1)^{\mathrm{T}}} & - \\ & \vdots & \\ - & y_N \mathbf{x}^{(N)^{\mathrm{T}}} & - \end{pmatrix} \begin{pmatrix} | & & | \\ y_1 \mathbf{x}^{(1)} & \cdots & y_n \mathbf{x}^{(N)} \\ | & & | \end{pmatrix}}_{\mathbf{A}} \begin{pmatrix} \alpha_1 \\ \vdots \\ \alpha_N \end{pmatrix} \right]$$

Two ways of computing $\mathbf{A}$:

- $\mathbf{A} = (\mathbf{Xy})^{\mathrm{T}}(\mathbf{Xy})$
- Computing $a_{ij} = y_i y_j \mathbf{x}^{(i)^{T}} \mathbf{x}^{(j)} \quad \forall\, i, j$

EXERCICE 1

$C_1 : \left\{ \mathbf{x}^{(1)} = \begin{pmatrix} 1 \\ 2 \end{pmatrix} \right\}$ with label 1

$C_2 : \left\{ \mathbf{x}^{(2)} = \begin{pmatrix} -1 \\ -2 \end{pmatrix}, \mathbf{x}^{(3)} = \begin{pmatrix} -2 \\ -1 \end{pmatrix} \right\}$ with label -1

(b) Write the Dual Lagrangian $\mathcal{L}_D$ and derive the values of $\alpha_i, i = 1, 2, 3$

$$\mathcal{L}_D(\alpha) = \sum_{i=1}^{N} \alpha_i - \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} \alpha_i \alpha_j y_i y_j \mathbf{x}^{(i)^{\mathrm{T}}} \mathbf{x}^{(j)} \qquad \text{subject to} \qquad \alpha_i \geq 0 \quad \forall i \qquad \text{and} \qquad \sum_{i=1}^{N} \alpha_i y_i = 0$$

$$= \mathbf{1}^{\mathrm{T}} \boldsymbol{\alpha} - \frac{1}{2} \boldsymbol{\alpha}^{\mathrm{T}} \mathbf{A} \boldsymbol{\alpha}$$

- $\mathbf{A} = (\mathbf{X}\mathbf{y})^{\mathrm{T}} (\mathbf{X}\mathbf{y})$

$$\mathbf{X}\mathbf{y} = \begin{pmatrix} | & & | \\ y_1 \mathbf{x}^{(1)} & \cdots & y_n \mathbf{x}^{(N)} \\ | & & | \end{pmatrix} = \begin{pmatrix} 1 & 1 & 2 \\ 2 & 2 & 1 \end{pmatrix}$$

$$\Rightarrow \mathbf{A} = \begin{pmatrix} 1 & 1 & 2 \\ 2 & 2 & 1 \end{pmatrix}^{\mathrm{T}} \begin{pmatrix} 1 & 1 & 2 \\ 2 & 2 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 2 \\ 1 & 2 \\ 2 & 1 \end{pmatrix} \begin{pmatrix} 1 & 1 & 2 \\ 2 & 2 & 1 \end{pmatrix} = \begin{pmatrix} 5 & 5 & 4 \\ 5 & 5 & 4 \\ 4 & 4 & 5 \end{pmatrix}$$

EXERCICE 1

$C_1 : \left\{ \mathbf{x}^{(1)} = \begin{pmatrix} 1 \\ 2 \end{pmatrix} \right\}$ with label 1

$C_2 : \left\{ \mathbf{x}^{(2)} = \begin{pmatrix} -1 \\ -2 \end{pmatrix}, \mathbf{x}^{(3)} = \begin{pmatrix} -2 \\ -1 \end{pmatrix} \right\}$ with label -1

(b) Write the Dual Lagrangian $\mathcal{L}_D$ and derive the values of $\alpha_i, i = 1, 2, 3$

$$\mathcal{L}_D(\alpha) = \sum_{i=1}^{N} \alpha_i - \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} \alpha_i \alpha_j y_i y_j {\mathbf{x}^{(i)}}^{\mathrm{T}} \mathbf{x}^{(j)} \qquad \text{subject to} \qquad \alpha_i \geq 0 \quad \forall i \qquad \text{and} \qquad \sum_{i=1}^{N} \alpha_i y_i = 0$$

$$= \mathbf{1}^{\mathrm{T}} \boldsymbol{\alpha} - \frac{1}{2} \boldsymbol{\alpha}^{\mathrm{T}} \mathbf{A} \boldsymbol{\alpha}$$

- Computing $a_{ij} = y_i y_j {\mathbf{x}^{(i)}}^{T} \mathbf{x}^{(j)} \quad \forall\, i, j$

$$\mathbf{A} = \begin{pmatrix} \times & \times & \times \\ \times & \times & \times \\ \times & \times & \times \end{pmatrix}$$

$$\boxed{y_3 \cdot y_1 \cdot {\mathbf{x}^{(3)}}^{\mathrm{T}} \cdot \mathbf{x}^{(1)} = (-1) \cdot (+1) \cdot (-2 \quad -1) \cdot \begin{pmatrix} 1 \\ 2 \end{pmatrix} = 4}$$

$$\boxed{y_1 \cdot y_1 \cdot {\mathbf{x}^{(1)}}^{\mathrm{T}} \cdot \mathbf{x}^{(1)} = (+1) \cdot (+1) \cdot (1 \quad 2) \cdot \begin{pmatrix} 1 \\ 2 \end{pmatrix} = 5}$$

EXERCICE 1

$C_1: \left\{ \mathbf{x}^{(1)} = \begin{pmatrix} 1 \\ 2 \end{pmatrix} \right\}$ with label 1

$C_2: \left\{ \mathbf{x}^{(2)} = \begin{pmatrix} -1 \\ -2 \end{pmatrix}, \mathbf{x}^{(3)} = \begin{pmatrix} -2 \\ -1 \end{pmatrix} \right\}$ with label -1

(b) Write the Dual Lagrangian $\mathcal{L}_D$ and derive the values of $\alpha_i, i = 1, 2, 3$

$$\mathcal{L}_D(\alpha) = \sum_{i=1}^{N} \alpha_i - \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} \alpha_i \alpha_j y_i y_j \mathbf{x}^{(i)^{\mathrm{T}}} \mathbf{x}^{(j)} \qquad \text{subject to} \qquad \alpha_i \geq 0 \quad \forall i \qquad \text{and} \qquad \sum_{i=1}^{N} \alpha_i y_i = 0$$

$$= \mathbf{1}^{\mathrm{T}} \boldsymbol{\alpha} - \frac{1}{2} \boldsymbol{\alpha}^{\mathrm{T}} \mathbf{A} \boldsymbol{\alpha}$$

$$\mathcal{L}_D(\alpha) = (1 \quad 1 \quad 1) \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \end{pmatrix} - \frac{1}{2} \left[(\alpha_1 \quad \alpha_2 \quad \alpha_3) \begin{pmatrix} 5 & 5 & 4 \\ 5 & 5 & 4 \\ 4 & 4 & 5 \end{pmatrix} \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \end{pmatrix} \right]$$

EXERCICE 1

$C_1: \left\{\mathbf{x}^{(1)} = \begin{pmatrix}1\\2\end{pmatrix}\right\}$ with label 1

$C_2: \left\{\mathbf{x}^{(2)} = \begin{pmatrix}-1\\-2\end{pmatrix}, \mathbf{x}^{(3)} = \begin{pmatrix}-2\\-1\end{pmatrix}\right\}$ with label -1

(b) Write the Dual Lagrangian $\mathcal{L}_D$ and <u>derive the values of</u> $\alpha_i, i = 1, 2, 3$

$$\max_{\alpha_1,\alpha_2,\alpha_3} \mathcal{L}_D(\alpha) = (1 \quad 1 \quad 1)\begin{pmatrix}\alpha_1\\\alpha_2\\\alpha_3\end{pmatrix} - \frac{1}{2}\left[(\alpha_1 \quad \alpha_2 \quad \alpha_3)\begin{pmatrix}5 & 5 & 4\\5 & 5 & 4\\4 & 4 & 5\end{pmatrix}\begin{pmatrix}\alpha_1\\\alpha_2\\\alpha_3\end{pmatrix}\right] \xrightarrow{\frac{\partial \mathcal{L}_D(\boldsymbol{\alpha})}{\partial \boldsymbol{\alpha}} = 0} \alpha_1, \alpha_2, \alpha_3$$

$$= \mathbf{1}^{\mathrm{T}}\boldsymbol{\alpha} - \frac{1}{2}\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{A}\boldsymbol{\alpha}$$

$$\frac{\partial \mathcal{L}_D(\boldsymbol{\alpha})}{\partial \boldsymbol{\alpha}} = ?$$

REMINDERS $\quad \dfrac{\partial \mathbf{x}^{\mathrm{T}}\mathbf{b}}{\partial \mathbf{x}} = \dfrac{\partial \mathbf{b}^{\mathrm{T}}\mathbf{x}}{\partial \mathbf{x}} = \mathbf{b} \qquad \dfrac{\partial \mathbf{x}^{\mathrm{T}}\mathbf{A}\mathbf{x}}{\partial \mathbf{x}} = 2\mathbf{A}\mathbf{x}$

EXERCICE 1

$C_1 : \left\{ \mathbf{x}^{(1)} = \begin{pmatrix} 1 \\ 2 \end{pmatrix} \right\}$ with label 1

$C_2 : \left\{ \mathbf{x}^{(2)} = \begin{pmatrix} -1 \\ -2 \end{pmatrix}, \mathbf{x}^{(3)} = \begin{pmatrix} -2 \\ -1 \end{pmatrix} \right\}$ with label -1

(b) Write the Dual Lagrangian $\mathcal{L}_D$ and derive the values of $\alpha_i, i = 1, 2, 3$

$$\max_{\alpha_1,\alpha_2,\alpha_3} \mathcal{L}_D(\alpha) = (1 \quad 1 \quad 1) \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \end{pmatrix} - \frac{1}{2} \left[(\alpha_1 \quad \alpha_2 \quad \alpha_3) \begin{pmatrix} 5 & 5 & 4 \\ 5 & 5 & 4 \\ 4 & 4 & 5 \end{pmatrix} \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \end{pmatrix} \right] \xrightarrow{\frac{\partial \mathcal{L}_D(\boldsymbol{\alpha})}{\partial \boldsymbol{\alpha}} = 0} \alpha_1, \alpha_2, \alpha_3$$

$$= \mathbf{1}^{\mathrm{T}} \boldsymbol{\alpha} - \frac{1}{2} \boldsymbol{\alpha}^{\mathrm{T}} \mathbf{A} \boldsymbol{\alpha}$$

$$\frac{\partial \mathcal{L}_D(\boldsymbol{\alpha})}{\partial \boldsymbol{\alpha}} = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} - \left[\begin{pmatrix} 5 & 5 & 4 \\ 5 & 5 & 4 \\ 4 & 4 & 5 \end{pmatrix} \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \end{pmatrix} \right] = 0 \Rightarrow$$

$$\boxed{\begin{aligned} 1 &= 5\alpha_1 + 5\alpha_2 + 4\alpha_3 \\ 1 &= 5\alpha_1 + 5\alpha_2 + 4\alpha_3 \\ 1 &= 4\alpha_1 + 4\alpha_2 + 5\alpha_3 \end{aligned}}$$

Are the same eq.
$\Longrightarrow$ **dependent system**

EXERCICE 1

$C_1 : \left\{ \mathbf{x}^{(1)} = \begin{pmatrix} 1 \\ 2 \end{pmatrix} \right\}$ with label 1

$C_2 : \left\{ \mathbf{x}^{(2)} = \begin{pmatrix} -1 \\ -2 \end{pmatrix}, \mathbf{x}^{(3)} = \begin{pmatrix} -2 \\ -1 \end{pmatrix} \right\}$ with label -1

(b) Write the Dual Lagrangian $\mathcal{L}_D$ and <u>derive the values of</u> $\alpha_i, i = 1, 2, 3$

$$\frac{\partial \mathcal{L}_D(\boldsymbol{\alpha})}{\partial \boldsymbol{\alpha}} = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} - \left[\begin{pmatrix} 5 & 5 & 4 \\ 5 & 5 & 4 \\ 4 & 4 & 5 \end{pmatrix} \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \end{pmatrix} \right] = 0 \quad \Rightarrow \quad \boxed{\begin{aligned} 1 &= 5\alpha_1 + 5\alpha_2 + 4\alpha_3 \\ 1 &= 5\alpha_1 + 5\alpha_2 + 4\alpha_3 \\ 1 &= 4\alpha_1 + 4\alpha_2 + 5\alpha_3 \end{aligned}}$$

Are the same eq.
$\Longrightarrow$ **dependent system**

BUT REMEMBER Dual problem:

$$\mathcal{L}_D(\alpha) = \sum_{i=1}^{N} \alpha_i - \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} \alpha_i \alpha_j y_i y_j {\mathbf{x}^{(i)}}^{\mathrm{T}} \mathbf{x}^{(j)}$$

Subject to $\alpha_i \geq 0 \quad \forall i$ and $\sum_{i=1}^{N} \alpha_i y_i = 0$

Considering also this eq.
$\Longrightarrow$ **independent system**

EXERCICE 1

$C_1 : \left\{ \mathbf{x}^{(1)} = \begin{pmatrix} 1 \\ 2 \end{pmatrix} \right\}$ with label 1

$C_2 : \left\{ \mathbf{x}^{(2)} = \begin{pmatrix} -1 \\ -2 \end{pmatrix}, \mathbf{x}^{(3)} = \begin{pmatrix} -2 \\ -1 \end{pmatrix} \right\}$ with label -1

(b) Write the Dual Lagrangian $\mathcal{L}_D$ and <u>derive the values of</u> $\alpha_i, i = 1, 2, 3$

$$\cancel{1 = 5\alpha_1 + 5\alpha_2 + 4\alpha_3}$$

$$1 = 5\alpha_1 + 5\alpha_2 + 4\alpha_3$$

$$1 = 4\alpha_1 + 4\alpha_2 + 5\alpha_3$$

$$\sum_{i=1}^{N} \alpha_i y_i = 0 \longrightarrow \alpha_1 - \alpha_2 - \alpha_3 = 0$$

$$\boxed{\alpha_1 = \frac{1}{9} \quad \alpha_2 = 0 \quad \alpha_3 = \frac{1}{9}}$$

EXERCICE 1

$C_1 : \left\{ \mathbf{x}^{(1)} = \begin{pmatrix} 1 \\ 2 \end{pmatrix} \right\}$ with label 1

$C_2 : \left\{ \mathbf{x}^{(2)} = \begin{pmatrix} -1 \\ -2 \end{pmatrix}, \mathbf{x}^{(3)} = \begin{pmatrix} -2 \\ -1 \end{pmatrix} \right\}$ with label -1

(c) Write the final classifier $g(\mathbf{x})$ and draw the decision hyperplane.

REMINDER Dual problem:

$$\mathcal{L}_D(\alpha) = \sum_{i=1}^{N} \alpha_i - \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} \alpha_i \alpha_j y_i y_j \mathbf{x}^{(i)^{\mathrm{T}}} \mathbf{x}^{(j)}$$

Subject to $\quad \alpha_i \geq 0 \quad \forall i \quad$ and $\quad \sum_{i=1}^{N} \alpha_i y_i = 0$

$$\boxed{\begin{aligned} \alpha_1 &= \frac{1}{9} \\ \alpha_2 &= 0 \\ \alpha_3 &= \frac{1}{9} \end{aligned}}$$

Then, the classifier is $\quad g\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = w_1 x_1 + w_2 x_2 + b$

where $\quad \boldsymbol{w} = \sum_{i=1}^{N} \alpha_i y_i \mathbf{x}^{(i)}$

EXERCICE 1

$C_1 : \left\{ \mathbf{x}^{(1)} = \begin{pmatrix} 1 \\ 2 \end{pmatrix} \right\}$ with label 1

$C_2 : \left\{ \mathbf{x}^{(2)} = \begin{pmatrix} -1 \\ -2 \end{pmatrix}, \mathbf{x}^{(3)} = \begin{pmatrix} -2 \\ -1 \end{pmatrix} \right\}$ with label -1

(c) Write the final classifier $g(\mathbf{x})$ and draw the decision hyperplane.

Then, the classifier is $g\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = w_1 x_1 + w_2 x_2 + b$ where $\boldsymbol{w} = \sum_{i=1}^{N} \alpha_i y_i \mathbf{x}^{(i)}$

$$\boxed{\begin{aligned} \alpha_1 &= \frac{1}{9} \\ \alpha_2 &= 0 \\ \alpha_3 &= \frac{1}{9} \end{aligned}}$$

$\boldsymbol{w} =$

EXERCICE 1

$\mathrm{C}_1 : \left\{ \mathbf{x}^{(1)} = \begin{pmatrix} 1 \\ 2 \end{pmatrix} \right\}$ with label 1

$\mathrm{C}_2 : \left\{ \mathbf{x}^{(2)} = \begin{pmatrix} -1 \\ -2 \end{pmatrix}, \mathbf{x}^{(3)} = \begin{pmatrix} -2 \\ -1 \end{pmatrix} \right\}$ with label -1

(c) Write the final classifier $g(\mathbf{x})$ and draw the decision hyperplane.

Then, the classifier is $g\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = w_1 x_1 + w_2 x_2 + b$ where $\boldsymbol{w} = \sum_{i=1}^{N} \alpha_i y_i \mathbf{x}^{(i)}$

$$\boxed{\begin{aligned} \alpha_1 &= \frac{1}{9} \\ \alpha_2 &= 0 \\ \alpha_3 &= \frac{1}{9} \end{aligned}}$$

$$\boldsymbol{w} = \alpha_1 \begin{pmatrix} 1 \\ 2 \end{pmatrix} - \alpha_2 \begin{pmatrix} -1 \\ -2 \end{pmatrix} - \alpha_3 \begin{pmatrix} -2 \\ -1 \end{pmatrix} = \frac{1}{9}\begin{pmatrix} 1 \\ 2 \end{pmatrix} - 0\begin{pmatrix} -1 \\ -2 \end{pmatrix} - \frac{1}{9}\begin{pmatrix} -2 \\ -1 \end{pmatrix} = \begin{pmatrix} 1/3 \\ 1/3 \end{pmatrix} \longrightarrow \boxed{w_1 = \frac{1}{3}, w_2 = \frac{1}{3}}$$

EXERCICE 1

$C_1 : \left\{ \mathbf{x}^{(1)} = \begin{pmatrix} 1 \\ 2 \end{pmatrix} \right\}$ with label 1

$C_2 : \left\{ \mathbf{x}^{(2)} = \begin{pmatrix} -1 \\ -2 \end{pmatrix}, \mathbf{x}^{(3)} = \begin{pmatrix} -2 \\ -1 \end{pmatrix} \right\}$ with label -1

(c) Write the final classifier $g(\mathbf{x})$ and draw the decision hyperplane.

Then, the classifier is $g\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \frac{1}{3} x_1 + \frac{1}{3} x_2 + b$?

EXERCICE 1

$C_1 : \left\{ \mathbf{x}^{(1)} = \begin{pmatrix} 1 \\ 2 \end{pmatrix} \right\}$ with label 1

$C_2 : \left\{ \mathbf{x}^{(2)} = \begin{pmatrix} -1 \\ -2 \end{pmatrix}, \mathbf{x}^{(3)} = \begin{pmatrix} -2 \\ -1 \end{pmatrix} \right\}$ with label -1

(c) Write the final classifier $g(\mathbf{x})$ and draw the decision hyperplane.

Then, the classifier is $$g\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \frac{1}{3}x_1 + \frac{1}{3}x_2 + b$$

To calculate b, we impose the conditions for the support vectors. In particular:

For $\mathbf{x} = \mathbf{x}^{(1)}$, $g\begin{pmatrix} 1 \\ 2 \end{pmatrix} = +1$

For $\mathbf{x} = \mathbf{x}^{(2)}$, $g\begin{pmatrix} -1 \\ -2 \end{pmatrix} = -1$

For $\mathbf{x} = \mathbf{x}^{(3)}$, $g\begin{pmatrix} -2 \\ -1 \end{pmatrix} = -1$

EXERCICE 1

$C_1: \left\{ \mathbf{x}^{(1)} = \begin{pmatrix} 1 \\ 2 \end{pmatrix} \right\}$ with label 1

$C_2: \left\{ \mathbf{x}^{(2)} = \begin{pmatrix} -1 \\ -2 \end{pmatrix}, \mathbf{x}^{(3)} = \begin{pmatrix} -2 \\ -1 \end{pmatrix} \right\}$ with label -1

(c) Write the final classifier $g(\mathbf{x})$ and draw the decision hyperplane.

Then, the classifier is $g\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \frac{1}{3}x_1 + \frac{1}{3}x_2 + b$

To calculate b, we impose the conditions for the support vectors. In particular:

For $\mathbf{x} = \mathbf{x}^{(1)}$, $g\begin{pmatrix} 1 \\ 2 \end{pmatrix} = +1 \longrightarrow +1 = g\begin{pmatrix} 1 \\ 2 \end{pmatrix} = \frac{1}{3} + \frac{1}{3} \cdot 2 + b \longrightarrow \boxed{b = 0}$

For $\mathbf{x} = \mathbf{x}^{(2)}$, $g\begin{pmatrix} -1 \\ -2 \end{pmatrix} = -1 \longrightarrow -1 = g\begin{pmatrix} -1 \\ -2 \end{pmatrix} = \frac{1}{3}(-1) + \frac{1}{3}(-2) + b \longrightarrow \boxed{b = 0}$

For $\mathbf{x} = \mathbf{x}^{(3)}$, $g\begin{pmatrix} -2 \\ -1 \end{pmatrix} = -1 \longrightarrow -1 = g\begin{pmatrix} -2 \\ -1 \end{pmatrix} = \frac{1}{3}(-2) + \frac{1}{3}(-1) + b \longrightarrow \boxed{b = 0}$

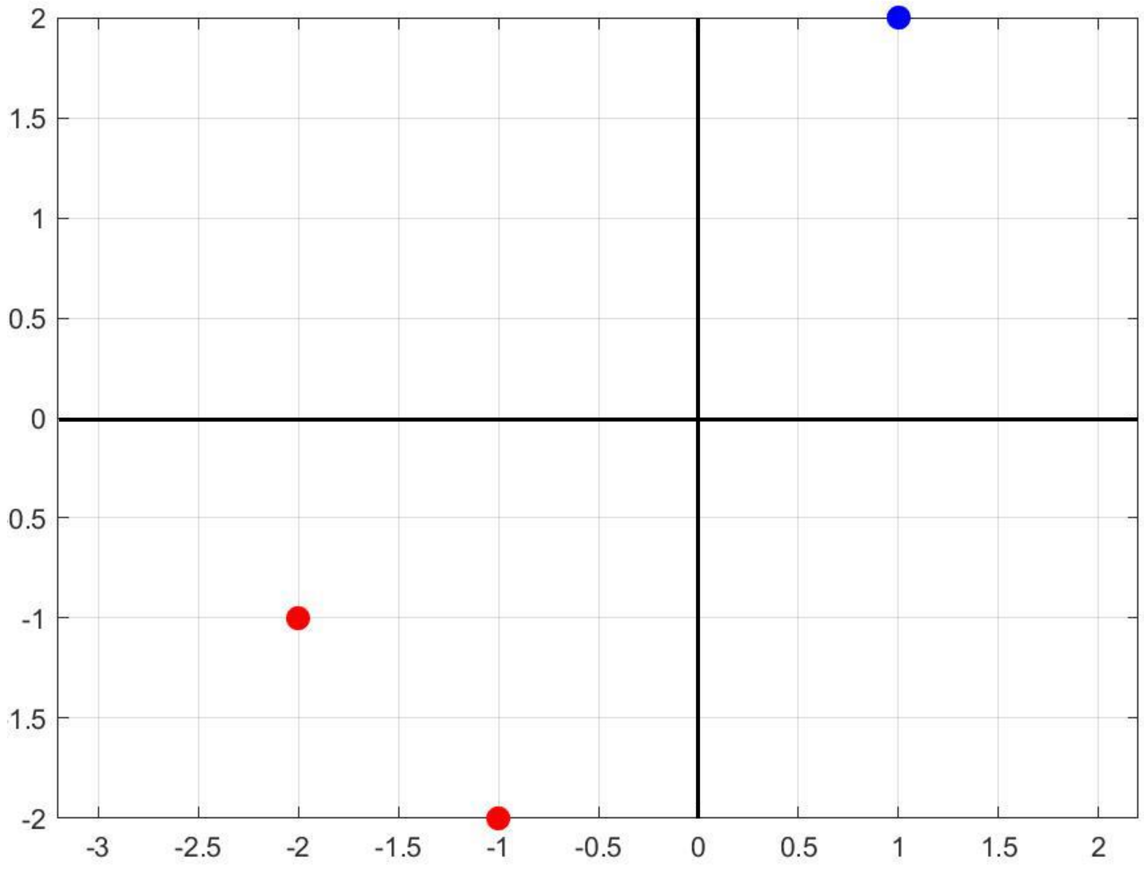
EXERCICE 1

$$C_1 : \left\{ \mathbf{x}^{(1)} = \begin{pmatrix} 1 \\ 2 \end{pmatrix} \right\} \text{ with label } 1$$

$$C_2 : \left\{ \mathbf{x}^{(2)} = \begin{pmatrix} -1 \\ -2 \end{pmatrix}, \mathbf{x}^{(3)} = \begin{pmatrix} -2 \\ -1 \end{pmatrix} \right\} \text{ with label } -1$$

(c) Write the final classifier $g(\mathbf{x})$ and <u>draw the decision hyperplane</u>.

$$g\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \frac{1}{3}x_1 + \frac{1}{3}x_2$$



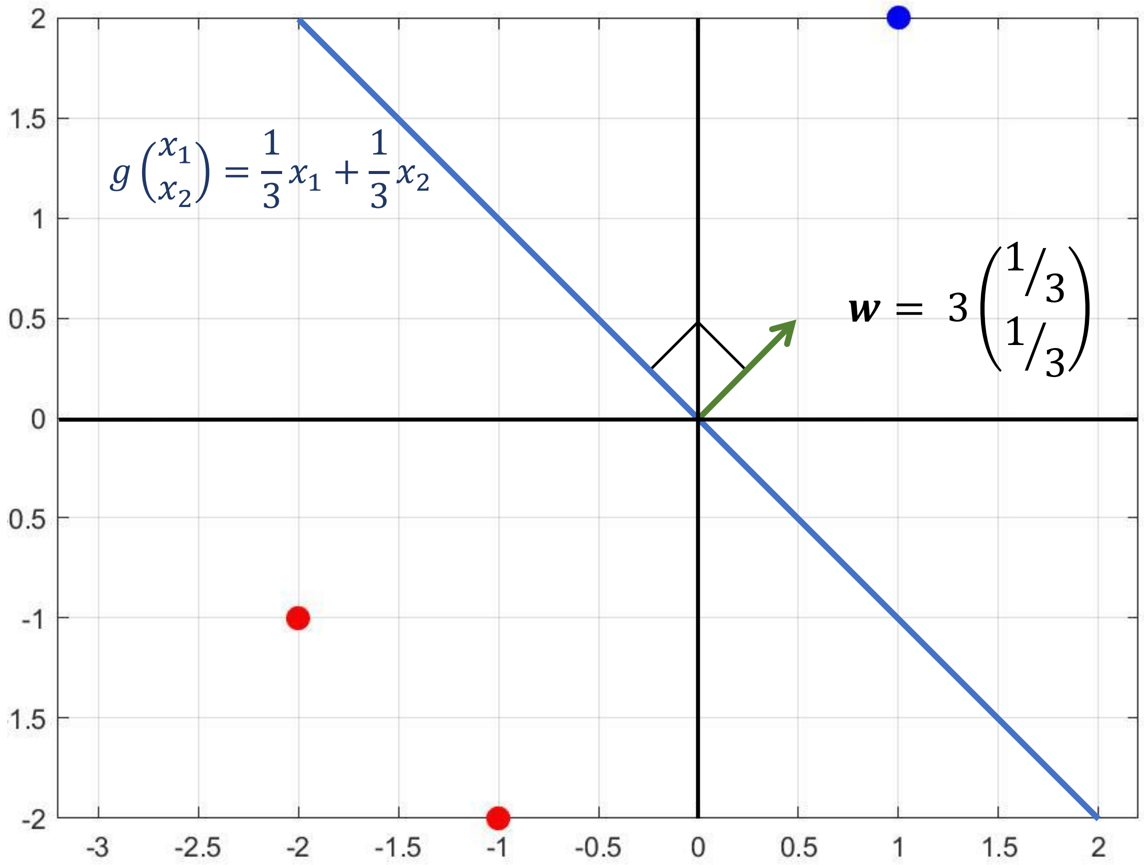
EXERCICE 1

$C_1 : \left\{ \mathbf{x}^{(1)} = \begin{pmatrix} 1 \\ 2 \end{pmatrix} \right\}$ with label 1

$C_2 : \left\{ \mathbf{x}^{(2)} = \begin{pmatrix} -1 \\ -2 \end{pmatrix}, \mathbf{x}^{(3)} = \begin{pmatrix} -2 \\ -1 \end{pmatrix} \right\}$ with label -1

(c) Write the final classifier $g(\mathbf{x})$ and <u>draw the decision hyperplane</u>.

$$g\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \frac{1}{3}x_1 + \frac{1}{3}x_2$$



EXERCICE 1

$C_1: \left\{ \mathbf{x}^{(1)} = \begin{pmatrix} 1 \\ 2 \end{pmatrix} \right\}$ with label 1

$C_2: \left\{ \mathbf{x}^{(2)} = \begin{pmatrix} -1 \\ -2 \end{pmatrix}, \mathbf{x}^{(3)} = \begin{pmatrix} -2 \\ -1 \end{pmatrix} \right\}$ with label -1

(d) What is the <u>margin value</u> of the obtained classifier?

REMINDER $\quad \text{margin} = \dfrac{2}{||\boldsymbol{w}||} =$

EXERCICE 1

$C_1 : \left\{ \mathbf{x}^{(1)} = \begin{pmatrix} 1 \\ 2 \end{pmatrix} \right\}$ with label 1

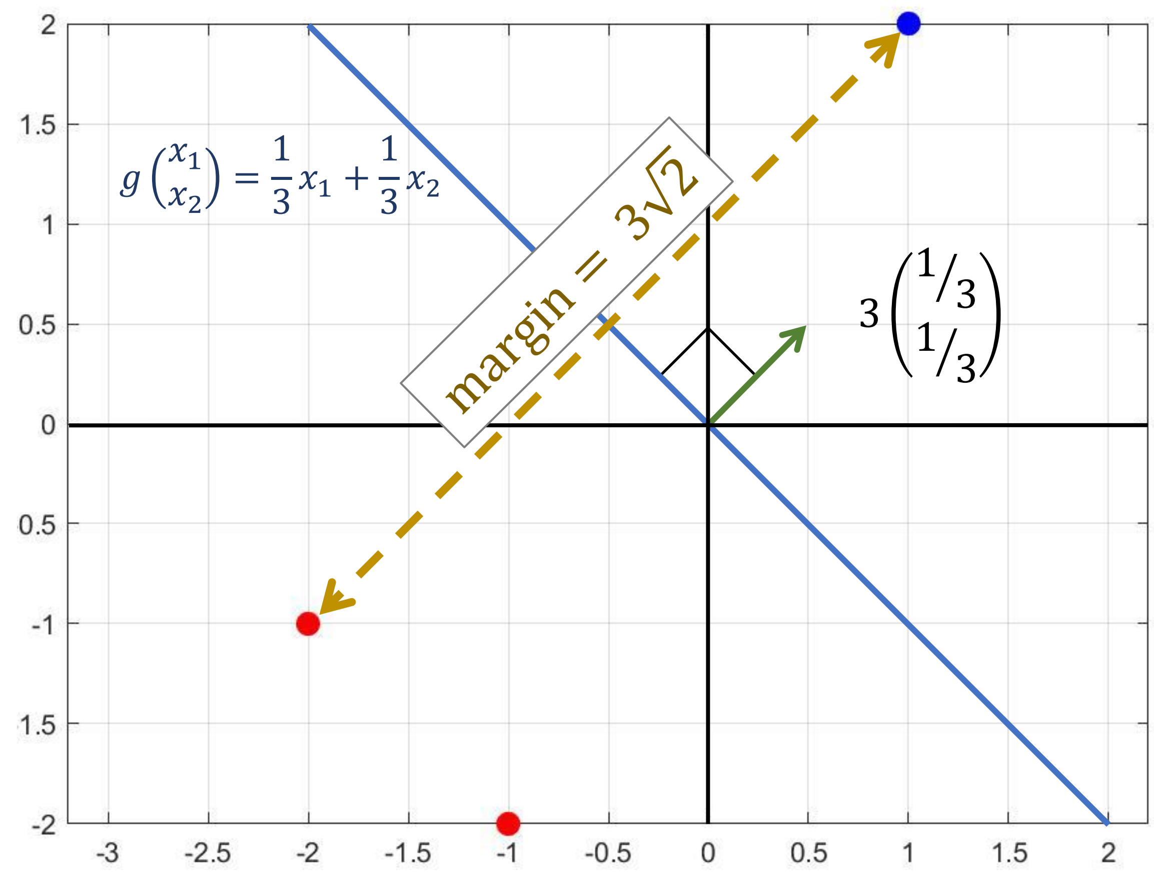
$C_2 : \left\{ \mathbf{x}^{(2)} = \begin{pmatrix} -1 \\ -2 \end{pmatrix}, \mathbf{x}^{(3)} = \begin{pmatrix} -2 \\ -1 \end{pmatrix} \right\}$ with label -1

(d) What is the margin value of the obtained classifier?

From previous exercise, we obtained: $\boldsymbol{w} = \frac{1}{3}\begin{pmatrix} 1 \\ 1 \end{pmatrix}$

Therefore, **margin of the SVM** is:

$$\boxed{\text{margin} = \frac{2}{||\boldsymbol{w}||} = \frac{2}{\frac{1}{3}\sqrt{2}} = 3\sqrt{2}}$$



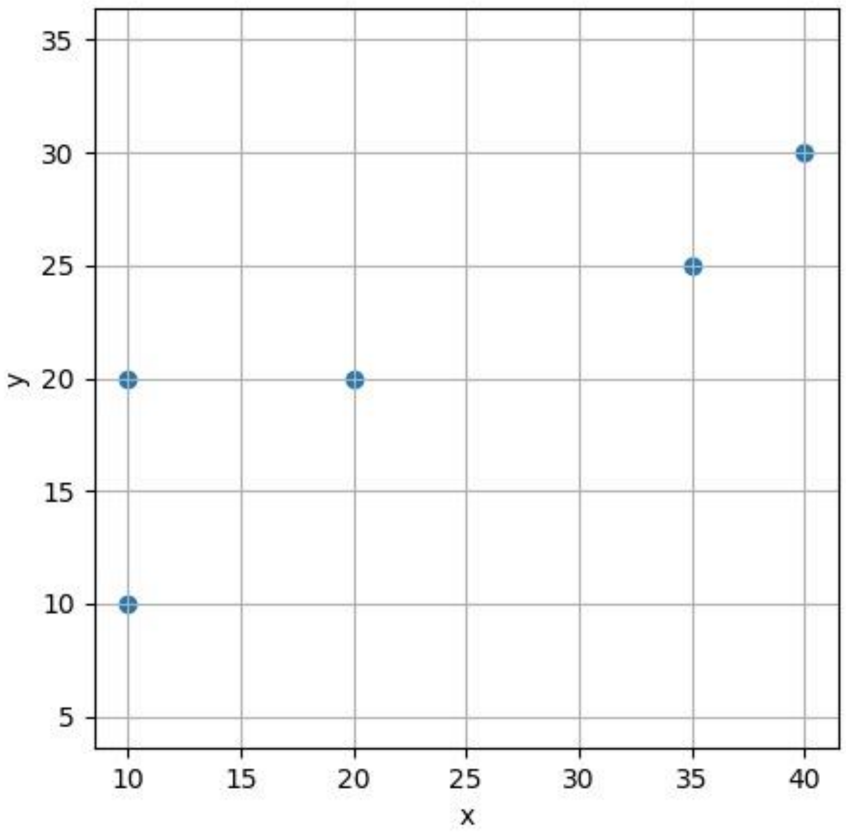
LINEAR REGRESSION AND REGULARISATION

EXERCISE 1

EXERCICE 1

Consider the training set with feature variables $\{x^{(1)} = 10, x^{(2)} = 10, x^{(3)} = 20, x^{(4)} = 35, x^{(5)} = 40\}$ and target variables $\{y^{(1)} = 10, y^{(2)} = 20, y^{(3)} = 20, y^{(4)} = 25, y^{(5)} = 30\}$.

Linear regression



A linear regression model is a function

$$y = g(\mathbf{x})$$

with g linear, i.e. $g(\mathbf{x}) = \mathbf{w}^{\mathrm{T}}\mathbf{x} + w_0 \Longrightarrow g(\mathbf{x}) = \mathbf{w}^{\mathrm{T}}\mathbf{x}$ with $\mathbf{x} = (1, \mathbf{x})^{\mathrm{T}}$

In our case (1D)

$$y = g(x) = w_0 + w_1 x$$

$g(x) = \mathbf{w}^{\mathrm{T}}\mathbf{x}$, where $\mathbf{w}^{\mathrm{T}} = (w_0, w_1)$ and $\mathbf{x} = \begin{pmatrix} 1 \\ x \end{pmatrix}$.

EXERCICE 1

$g(x) = w_0 + w_1 x$

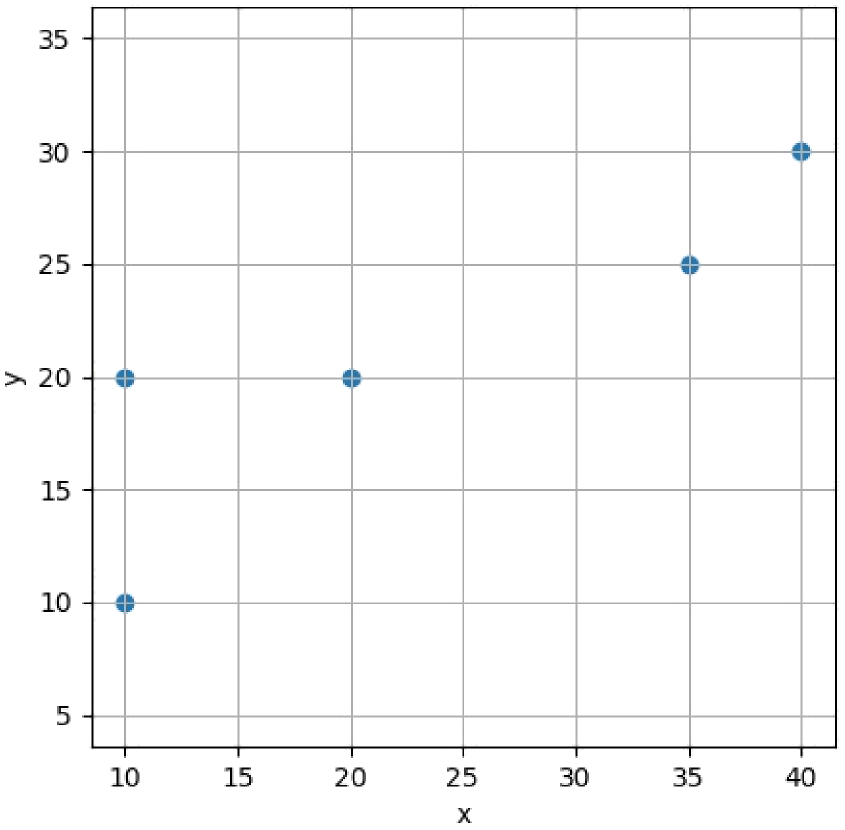
Feature variables $\{x^{(1)} = 10, x^{(2)} = 10, x^{(3)} = 20, x^{(4)} = 35, x^{(5)} = 40\}$

Target variables $\{y^{(1)} = 10, y^{(2)} = 20, y^{(3)} = 20, y^{(4)} = 25, y^{(5)} = 30\}$

(a) *Closed-form solution*

(i) Estimate $\mathbf{w}$ without regularization.
Estimate $\mathbf{w}_{\text{reg}}$ with regularization (Ridge regression), with $\lambda = 1$.

(ii) Compute the error in the training set for both models and plot the training set and both models in the same figure.



EXERCICE 1

$g(x) = w_0 + w_1 x$

Feature variables $\{x^{(1)} = 10, x^{(2)} = 10, x^{(3)} = 20, x^{(4)} = 35, x^{(5)} = 40\}$

Target variables $\{y^{(1)} = 10, y^{(2)} = 20, y^{(3)} = 20, y^{(4)} = 25, y^{(5)} = 30\}$

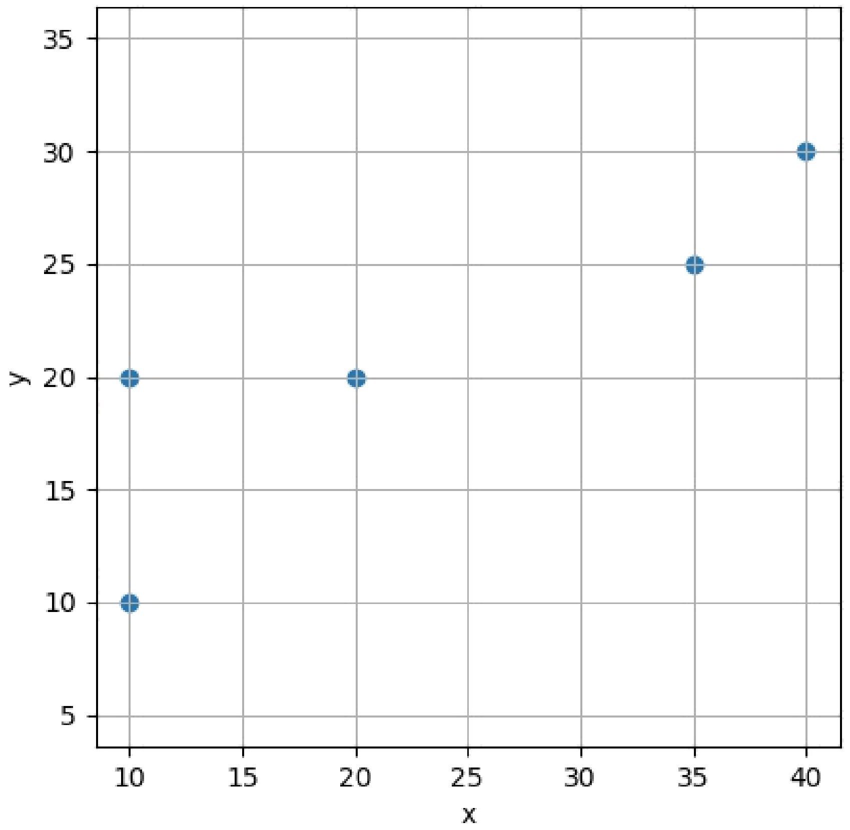
(a) *Closed-form solution*

(i) Estimate $\mathbf{w}$ without regularization.

$$\mathbf{w} = \left(\mathbf{X}^{\mathrm{T}}\mathbf{X}\right)^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{y}$$

Design matrix

Target vector



EXERCICE 1

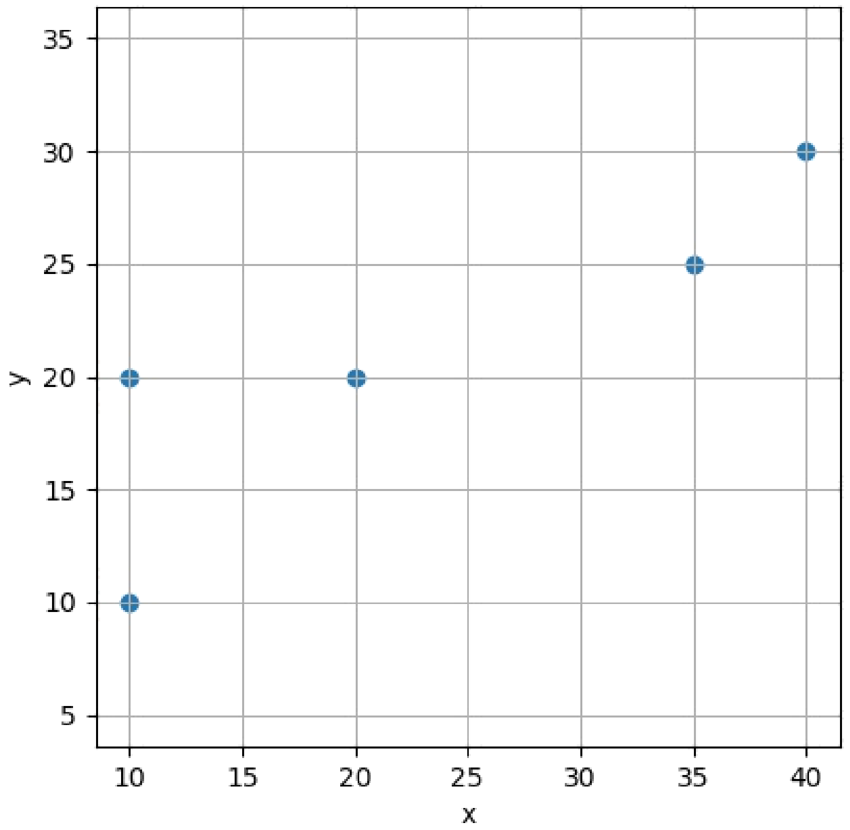
$g(x) = w_0 + w_1 x$

Feature variables $\{x^{(1)} = 10, x^{(2)} = 10, x^{(3)} = 20, x^{(4)} = 35, x^{(5)} = 40\}$

Target variables $\{y^{(1)} = 10, y^{(2)} = 20, y^{(3)} = 20, y^{(4)} = 25, y^{(5)} = 30\}$

(a) *Closed-form solution*

(i) Estimate $\mathbf{w}$ without regularization. $\mathbf{w} = \left(\mathbf{X}^{\mathrm{T}}\mathbf{X}\right)^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{y}$



$\boxed{\text{Design matrix}}$ $\mathbf{X} = ?$

$\boxed{\text{Target vector}}$ $\mathbf{y} = ?$

EXERCICE 1

$$g(x) = w_0 + w_1 x$$

Feature variables $\{x^{(1)} = 10, x^{(2)} = 10, x^{(3)} = 20, x^{(4)} = 35, x^{(5)} = 40\}$

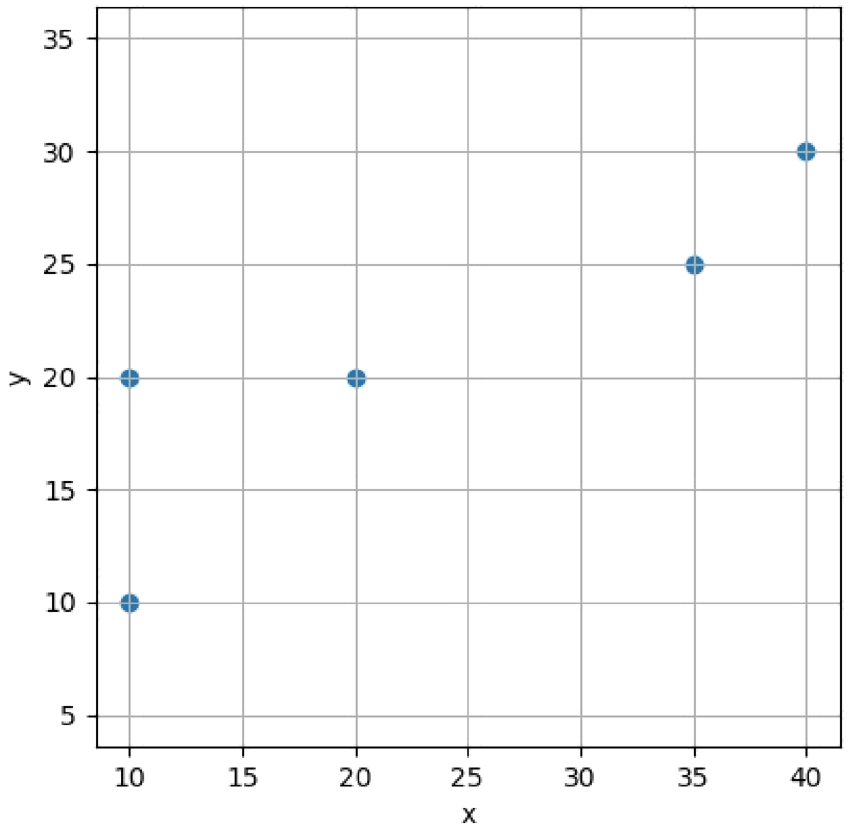
Target variables $\{y^{(1)} = 10, y^{(2)} = 20, y^{(3)} = 20, y^{(4)} = 25, y^{(5)} = 30\}$

(a) *Closed-form solution*

(i) Estimate $\mathbf{w}$ without regularization. $$\mathbf{w} = \left(\mathbf{X}^{\mathrm{T}}\mathbf{X}\right)^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{y}$$

Design matrix $$\mathbf{X} = \begin{pmatrix} 1 & x^{(1)} \\ 1 & x^{(2)} \\ 1 & x^{(3)} \\ 1 & x^{(4)} \\ 1 & x^{(5)} \end{pmatrix} = \begin{pmatrix} 1 & 10 \\ 1 & 10 \\ 1 & 20 \\ 1 & 35 \\ 1 & 40 \end{pmatrix}$$

Target vector $$\mathbf{y} = \begin{pmatrix} y^{(1)} \\ y^{(2)} \\ y^{(3)} \\ y^{(4)} \\ y^{(5)} \end{pmatrix} = \begin{pmatrix} 10 \\ 20 \\ 20 \\ 25 \\ 30 \end{pmatrix}$$



EXERCICE 1

$$g(x) = w_0 + w_1 x$$

Feature variables $\{x^{(1)} = 10, x^{(2)} = 10, x^{(3)} = 20, x^{(4)} = 35, x^{(5)} = 40\}$

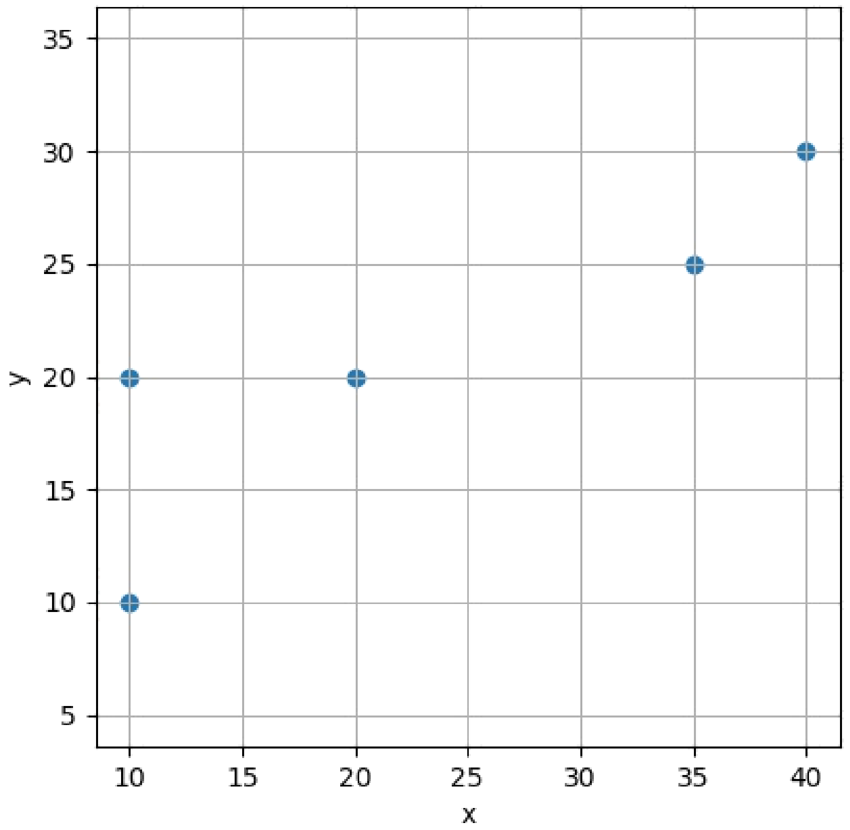
Target variables $\{y^{(1)} = 10, y^{(2)} = 20, y^{(3)} = 20, y^{(4)} = 25, y^{(5)} = 30\}$

(a) *Closed-form solution*

(i) Estimate $\mathbf{w}$ without regularization.

$$\mathbf{w} = \left(\mathbf{X}^\mathrm{T}\mathbf{X}\right)^{-1}\mathbf{X}^\mathrm{T}\mathbf{y} = \cdots = \frac{1}{13}\begin{pmatrix}45\\2\end{pmatrix} \approx \begin{pmatrix}10.38\\0.46\end{pmatrix}$$

$$\mathbf{X} = \begin{pmatrix}1 & 10\\1 & 10\\1 & 20\\1 & 35\\1 & 40\end{pmatrix} \qquad \mathbf{y} = \begin{pmatrix}10\\20\\20\\25\\30\end{pmatrix}$$



EXERCICE 1

$g(x) = w_0 + w_1 x$

Feature variables $\{x^{(1)} = 10, x^{(2)} = 10, x^{(3)} = 20, x^{(4)} = 35, x^{(5)} = 40\}$

Target variables $\{y^{(1)} = 10, y^{(2)} = 20, y^{(3)} = 20, y^{(4)} = 25, y^{(5)} = 30\}$

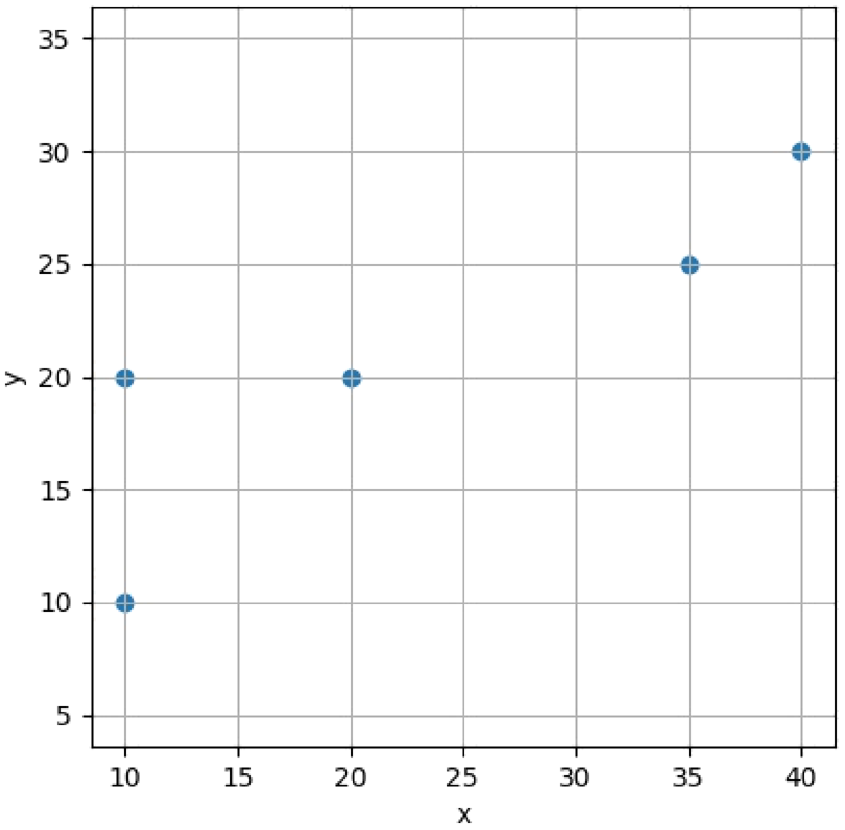
(a) *Closed-form solution*

(i) Estimate $\mathbf{w}$ without regularization.
Estimate $\mathbf{w}_{\text{reg}}$ with regularization (Ridge regression), with $\lambda = 1$.

$$\mathbf{w}_{\text{reg}} = \left(\mathbf{X}^{\mathrm{T}}\mathbf{X} + \boxed{\lambda \mathbf{Id}}\right)^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{y} = \cdots \approx \begin{pmatrix} 5.54 \\ 0.62 \end{pmatrix}$$

Regularisation term

$$\mathbf{X} = \begin{pmatrix} 1 & 10 \\ 1 & 10 \\ 1 & 20 \\ 1 & 35 \\ 1 & 40 \end{pmatrix} \qquad \mathbf{y} = \begin{pmatrix} 10 \\ 20 \\ 20 \\ 25 \\ 30 \end{pmatrix}$$



EXERCICE 1

$$g(x) = w_0 + w_1 x$$

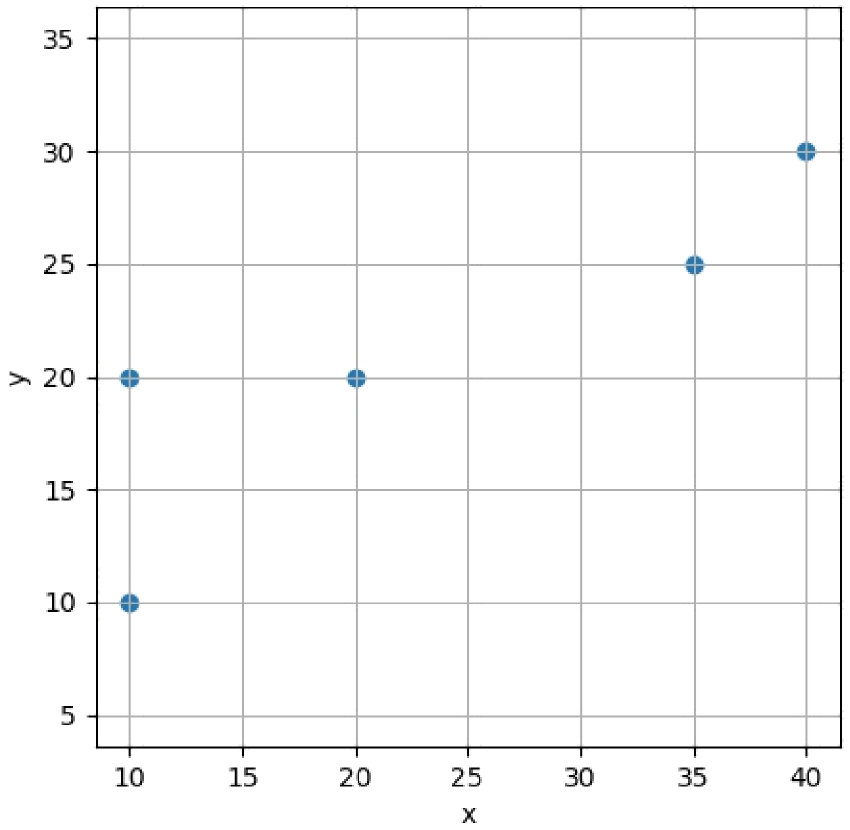
Feature variables $\{x^{(1)} = 10, x^{(2)} = 10, x^{(3)} = 20, x^{(4)} = 35, x^{(5)} = 40\}$

Target variables $\{y^{(1)} = 10, y^{(2)} = 20, y^{(3)} = 20, y^{(4)} = 25, y^{(5)} = 30\}$

(a) *Closed-form solution*

(i) Estimate $\mathbf{w}$ without regularization.
Estimate $\mathbf{w}_{\text{reg}}$ with regularization (Ridge regression), with $\lambda = 1$.

(ii) Compute the error in the training set for both models and plot the training set and both models in the same figure.



$$\mathbb{E}(\mathbf{w}) = \frac{1}{2}\sum_{n=1}^{N}\left(g\left(x^{(n)}, \mathbf{w}\right) - y^{(n)}\right)^2 = \frac{1}{2}\sum_{n=1}^{5}\left(w_0 + w_1 x^{(n)} - y^{(n)}\right)^2$$

EXERCICE 1

$$g(x) = w_0 + w_1 x$$

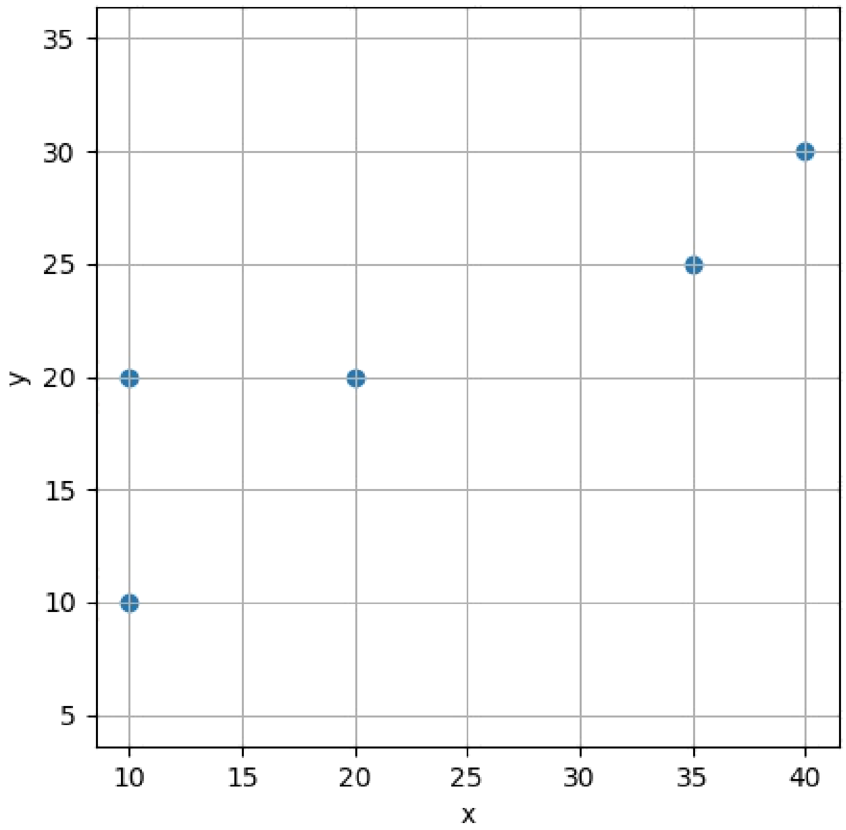
Feature variables $\{x^{(1)} = 10, x^{(2)} = 10, x^{(3)} = 20, x^{(4)} = 35, x^{(5)} = 40\}$

Target variables $\{y^{(1)} = 10, y^{(2)} = 20, y^{(3)} = 20, y^{(4)} = 25, y^{(5)} = 30\}$

(a) *Closed-form solution*

(i) Estimate $\mathbf{w}$ without regularization.
Estimate $\mathbf{w}_{\text{reg}}$ with regularization (Ridge regression), with $\lambda = 1$.

(ii) Compute the error in the training set for both models and plot the training set and both models in the same figure.



$$\mathbb{E}(\mathbf{w}) = \frac{1}{2}\sum_{n=1}^{5}\left(10.38 + 0.46\, x^{(n)} - y^{(n)}\right)^2 = \cdots = 26.92$$

$$\mathbb{E}(\mathbf{w}_{\text{reg}}) = \frac{1}{2}\sum_{n=1}^{5}\left(5.54 + 0.62\, x^{(n)} - y^{(n)}\right)^2 = \cdots = 40.29$$

$g(x) = w_0 + w_1 x$

EXERCICE 1

Feature variables $\{x^{(1)} = 10, x^{(2)} = 10, x^{(3)} = 20, x^{(4)} = 35, x^{(5)} = 40\}$

Target variables $\{y^{(1)} = 10, y^{(2)} = 20, y^{(3)} = 20, y^{(4)} = 25, y^{(5)} = 30\}$

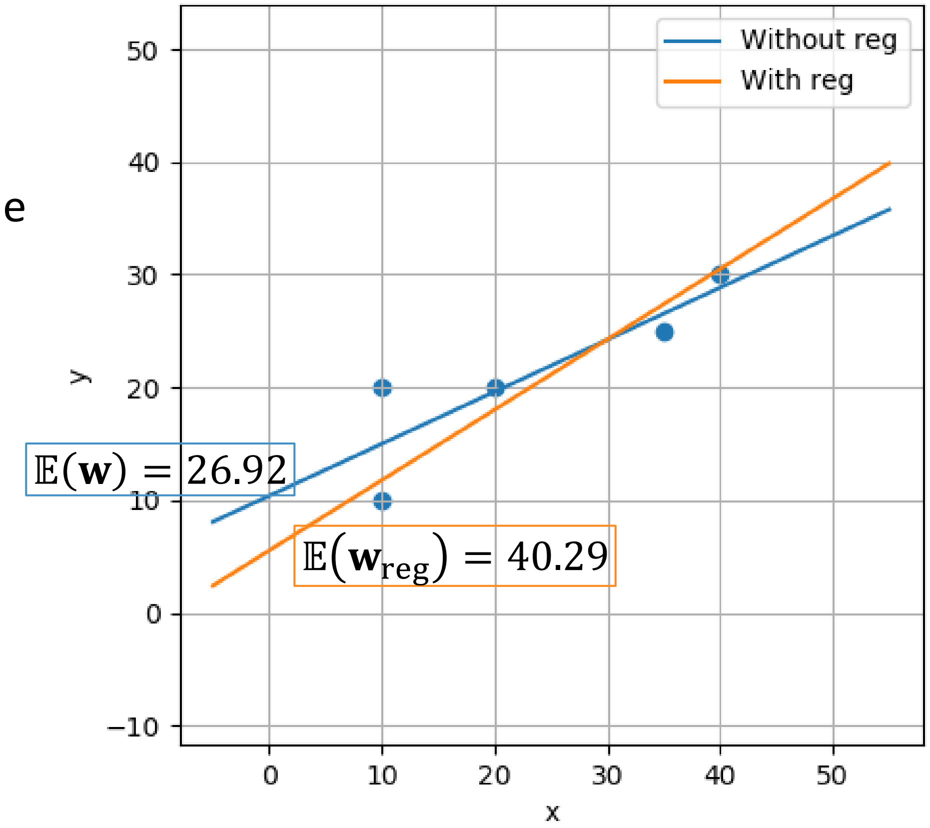
(a) *Closed-form solution*

(i) Estimate $\mathbf{w}$ without regularization.
Estimate $\mathbf{w}_{\text{reg}}$ with regularization (Ridge regression), with $\lambda = 1$.

(ii) Compute the error in the training set for both models and plot the training set and both models in the same figure.

$$\mathbb{E}(\mathbf{w}) = \frac{1}{2}\sum_{n=1}^{5}\left(10.38 + 0.46\, x^{(n)} - y^{(n)}\right)^2 = \cdots = 26.92$$

$$\mathbb{E}(\mathbf{w}_{\text{reg}}) = \frac{1}{2}\sum_{n=1}^{5}\left(5.54 + 0.62\, x^{(n)} - y^{(n)}\right)^2 = \cdots = 40.29$$



EXERCICE 1

$g(x) = w_0 + w_1 x$

Feature variables $\{x^{(1)} = 10, x^{(2)} = 10, x^{(3)} = 20, x^{(4)} = 35, x^{(5)} = 40\}$

Target variables $\{y^{(1)} = 10, y^{(2)} = 20, y^{(3)} = 20, y^{(4)} = 25, y^{(5)} = 30\}$

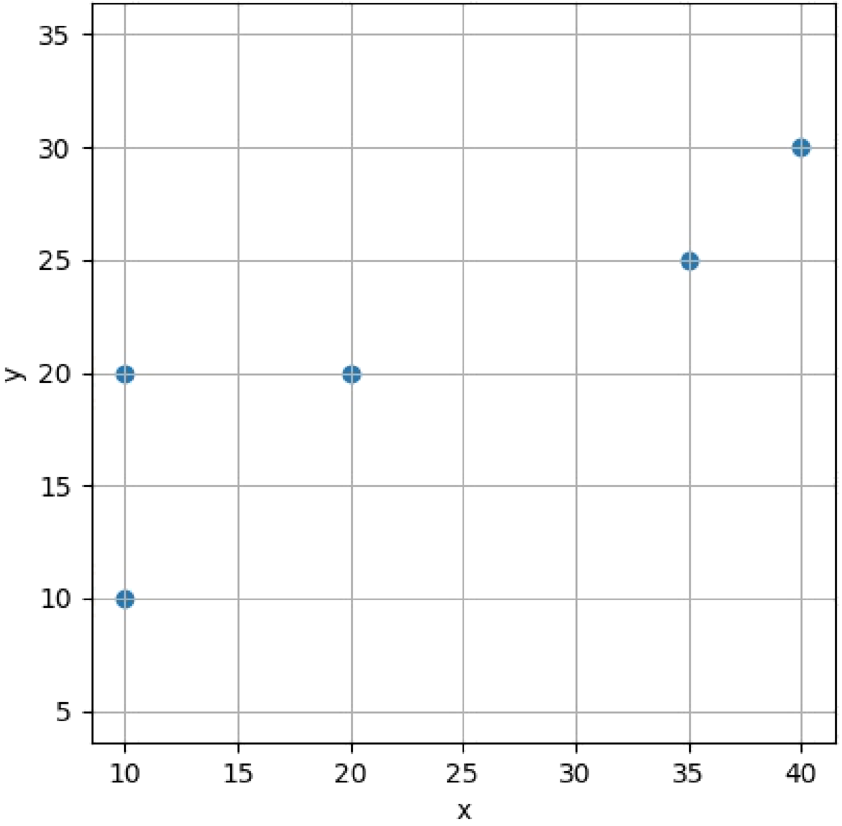
(b) *Gradient descent by hand*

(i) Parametric form of linear model and expression of error.

(ii) Expression of $\mathbf{w}$ update.

(iii) Update the parameters: initial $\mathbf{w} = (0,0)^{\mathrm{T}}$, learning rate $\alpha = 10^{-4}$. Plot both models.

(iv) Compute the error.



EXERCICE 1

$g(x) = w_0 + w_1 x$

Feature variables $\{x^{(1)} = 10, x^{(2)} = 10, x^{(3)} = 20, x^{(4)} = 35, x^{(5)} = 40\}$

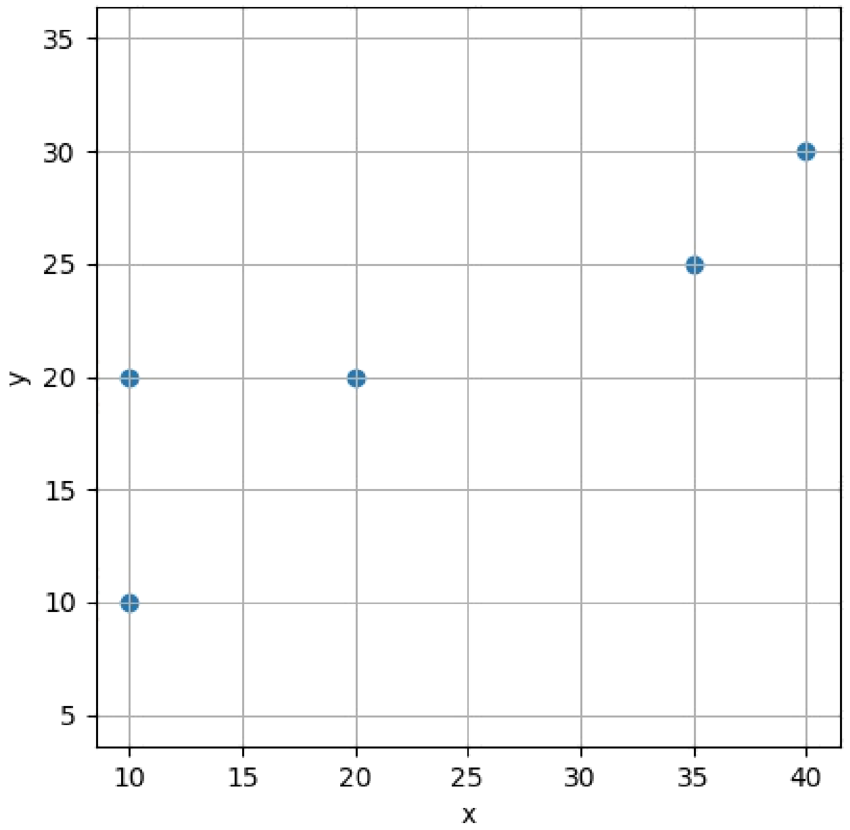
Target variables $\{y^{(1)} = 10, y^{(2)} = 20, y^{(3)} = 20, y^{(4)} = 25, y^{(5)} = 30\}$

(b) *Gradient descent by hand*

(i) Parametric form of linear model and expression of error.

$$g(x) = w_0 + w_1 x$$

$$\mathbb{E}(\mathbf{w}) = \frac{1}{2}\sum_{n=1}^{5}\left(w_0 + w_1 x^{(n)} - y^{(n)}\right)^2$$



EXERCICE 1

$g(x) = w_0 + w_1 x$

Feature variables $\{x^{(1)} = 10, x^{(2)} = 10, x^{(3)} = 20, x^{(4)} = 35, x^{(5)} = 40\}$

Target variables $\{y^{(1)} = 10, y^{(2)} = 20, y^{(3)} = 20, y^{(4)} = 25, y^{(5)} = 30\}$

(b) *Gradient descent by hand*

(ii) Expression of $\mathbf{w}$ update.

$$\mathbf{w}_{new} = \mathbf{w}_{old} - \alpha \left.\frac{\partial \mathbb{E}(\mathbf{w})}{\partial \mathbf{w}}\right|_{\mathbf{w} = \mathbf{w}_{\text{old}}}$$

This means that the derivative should be evaluated at $\mathbf{w}_{\text{old}}$

where $\mathbb{E}(\mathbf{w}) = \frac{1}{2}\sum_{n=1}^{5}\left(w_0 + w_1 x^{(n)} - y^{(n)}\right)^2$

EXERCICE 1

$g(x) = w_0 + w_1 x$

Feature variables $\{x^{(1)} = 10, x^{(2)} = 10, x^{(3)} = 20, x^{(4)} = 35, x^{(5)} = 40\}$

Target variables $\{y^{(1)} = 10, y^{(2)} = 20, y^{(3)} = 20, y^{(4)} = 25, y^{(5)} = 30\}$

(b) *Gradient descent by hand*

(ii) Expression of $\mathbf{w}$ update.

$$\mathbf{w}_{new} = \mathbf{w}_{old} - \alpha \left.\frac{\partial \mathbb{E}(\mathbf{w})}{\partial \mathbf{w}}\right|_{\mathbf{w} = \mathbf{w}_{\text{old}}} \qquad \text{where} \qquad \mathbb{E}(\mathbf{w}) = \frac{1}{2}\sum_{n=1}^{5}\left(w_0 + w_1 x^{(n)} - y^{(n)}\right)^2$$

$$\frac{\partial \mathbb{E}(\mathbf{w})}{\partial \mathbf{w}} = ?$$

EXERCICE 1

$$g(x) = w_0 + w_1 x$$

Feature variables $\{x^{(1)} = 10, x^{(2)} = 10, x^{(3)} = 20, x^{(4)} = 35, x^{(5)} = 40\}$

Target variables $\{y^{(1)} = 10, y^{(2)} = 20, y^{(3)} = 20, y^{(4)} = 25, y^{(5)} = 30\}$

(b) *Gradient descent by hand*

(ii) Expression of $\mathbf{w}$ update.

$$\mathbf{w}_{new} = \mathbf{w}_{old} - \alpha \frac{\partial \mathbb{E}(\mathbf{w})}{\partial \mathbf{w}}\bigg|_{\mathbf{w} = \mathbf{w}_{\text{old}}} \qquad \text{where} \qquad \mathbb{E}(\mathbf{w}) = \frac{1}{2}\sum_{n=1}^{5}\left(w_0 + w_1 x^{(n)} - y^{(n)}\right)^2$$

$$\frac{\partial \mathbb{E}(\mathbf{w})}{\partial \mathbf{w}} = \frac{\partial}{\partial \mathbf{w}} \frac{1}{2}\sum_{n=1}^{5}\left(\mathbf{w}^{\mathrm{T}}\mathbf{x}^{(n)} - y^{(n)}\right)^2 = \sum_{n=1}^{5}\left(\mathbf{w}^{\mathrm{T}}\mathbf{x}^{(n)} - y^{(n)}\right)\cdot \mathbf{x}^{(n)}$$

$g(x) = \mathbf{w}^{\mathrm{T}}\mathbf{x}$, where $\mathbf{w}^{\mathrm{T}} = (w_0, w_1)$ and $\mathbf{x} = \begin{pmatrix} 1 \\ x \end{pmatrix}$.

EXERCICE 1

$$g(x) = w_0 + w_1 x$$

Feature variables $\{x^{(1)} = 10, x^{(2)} = 10, x^{(3)} = 20, x^{(4)} = 35, x^{(5)} = 40\}$

Target variables $\{y^{(1)} = 10, y^{(2)} = 20, y^{(3)} = 20, y^{(4)} = 25, y^{(5)} = 30\}$

(b) *Gradient descent by hand*

(ii) Expression of $\mathbf{w}$ update.

$$\mathbf{w}_{new} = \mathbf{w}_{old} - \alpha \left. \frac{\partial \mathbb{E}(\mathbf{w})}{\partial \mathbf{w}} \right|_{\mathbf{w} = \mathbf{w}_{\text{old}}} \qquad \text{where} \qquad \frac{\partial \mathbb{E}(\mathbf{w})}{\partial \mathbf{w}} = \sum_{n=1}^{5} \left(\mathbf{w}^{\mathrm{T}} \mathbf{x}^{(n)} - y^{(n)} \right) \cdot \mathbf{x}^{(n)}$$

$$\mathbf{x} = \begin{pmatrix} 1 \\ x \end{pmatrix}$$

$$w_0 = w_0 - \alpha \sum_{n=1}^{5} \left(\mathbf{w}^{\mathrm{T}} \mathbf{x}^{(n)} - y^{(n)} \right) \cdot \mathbf{x}_0^{(n)} = w_0 - \alpha \sum_{n=1}^{5} \left(w_0 + w_1 x^{(n)} - y^{(n)} \right)$$

$$w_1 = w_1 - \alpha \sum_{n=1}^{5} \left(\mathbf{w}^{\mathrm{T}} \mathbf{x}^{(n)} - y^{(n)} \right) \cdot \mathbf{x}_1^{(n)} = w_1 - \alpha \sum_{n=1}^{5} \left(w_0 + w_1 x^{(n)} - y^{(n)} \right) x^{(n)}$$

EXERCICE 1

$g(x) = w_0 + w_1 x$

Feature variables $\{x^{(1)} = 10, x^{(2)} = 10, x^{(3)} = 20, x^{(4)} = 35, x^{(5)} = 40\}$

Target variables $\{y^{(1)} = 10, y^{(2)} = 20, y^{(3)} = 20, y^{(4)} = 25, y^{(5)} = 30\}$

(b) *Gradient descent by hand*

(iii) Update the parameters: initial $\boxed{\mathbf{w} = (0,0)^\mathrm{T}}$, learning rate $\alpha = 10^{-4}$.
Plot both models.

$$w_0 = w_0 - \alpha \sum_{n=1}^{5} \left(w_0 + w_1 x^{(n)} - y^{(n)} \right) =$$

$$w_1 = w_1 - \alpha \sum_{n=1}^{5} \left(w_0 + w_1 x^{(n)} - y^{(n)} \right) x^{(n)} =$$

EXERCICE 1

$g(x) = w_0 + w_1 x$

Feature variables $\{x^{(1)} = 10, x^{(2)} = 10, x^{(3)} = 20, x^{(4)} = 35, x^{(5)} = 40\}$

Target variables $\{y^{(1)} = 10, y^{(2)} = 20, y^{(3)} = 20, y^{(4)} = 25, y^{(5)} = 30\}$

(b) *Gradient descent by hand*

(iii) Update the parameters: initial $\boxed{\mathbf{w} = (0,0)^{\mathrm{T}}}$, learning rate $\alpha = 10^{-4}$. Plot both models.

$$w_0 = w_0 - \alpha \sum_{n=1}^{5} \left(w_0 + w_1 x^{(n)} - y^{(n)}\right) = w_0 - \alpha \sum_{n=1}^{5} -y^{(n)} = ?$$

$$w_1 = w_1 - \alpha \sum_{n=1}^{5} \left(w_0 + w_1 x^{(n)} - y^{(n)}\right) x^{(n)} = w_1 - \alpha \sum_{n=1}^{5} -y^{(n)} x^{(n)} = ?$$

EXERCICE 1

$g(x) = w_0 + w_1 x$

Feature variables $\{x^{(1)} = 10, x^{(2)} = 10, x^{(3)} = 20, x^{(4)} = 35, x^{(5)} = 40\}$

Target variables $\{y^{(1)} = 10, y^{(2)} = 20, y^{(3)} = 20, y^{(4)} = 25, y^{(5)} = 30\}$

(b) *Gradient descent by hand*

(iii) Update the parameters: initial $\mathbf{w} = (0,0)^{\mathrm{T}}$, learning rate $\alpha = 10^{-4}$.
Plot both models.

$$w_0 = w_0 - \alpha \sum_{n=1}^{5} \left(w_0 + w_1 x^{(n)} - y^{(n)}\right) = w_0 - \alpha \sum_{n=1}^{5} -y^{(n)} = 1.05 \cdot 10^{-2}$$

$$w_1 = w_1 - \alpha \sum_{n=1}^{5} \left(w_0 + w_1 x^{(n)} - y^{(n)}\right) x^{(n)} = w_1 - \alpha \sum_{n=1}^{5} -y^{(n)} x^{(n)} = 0.2775$$

EXERCICE 1

$g(x) = w_0 + w_1 x$

Feature variables $\{x^{(1)} = 10, x^{(2)} = 10, x^{(3)} = 20, x^{(4)} = 35, x^{(5)} = 40\}$

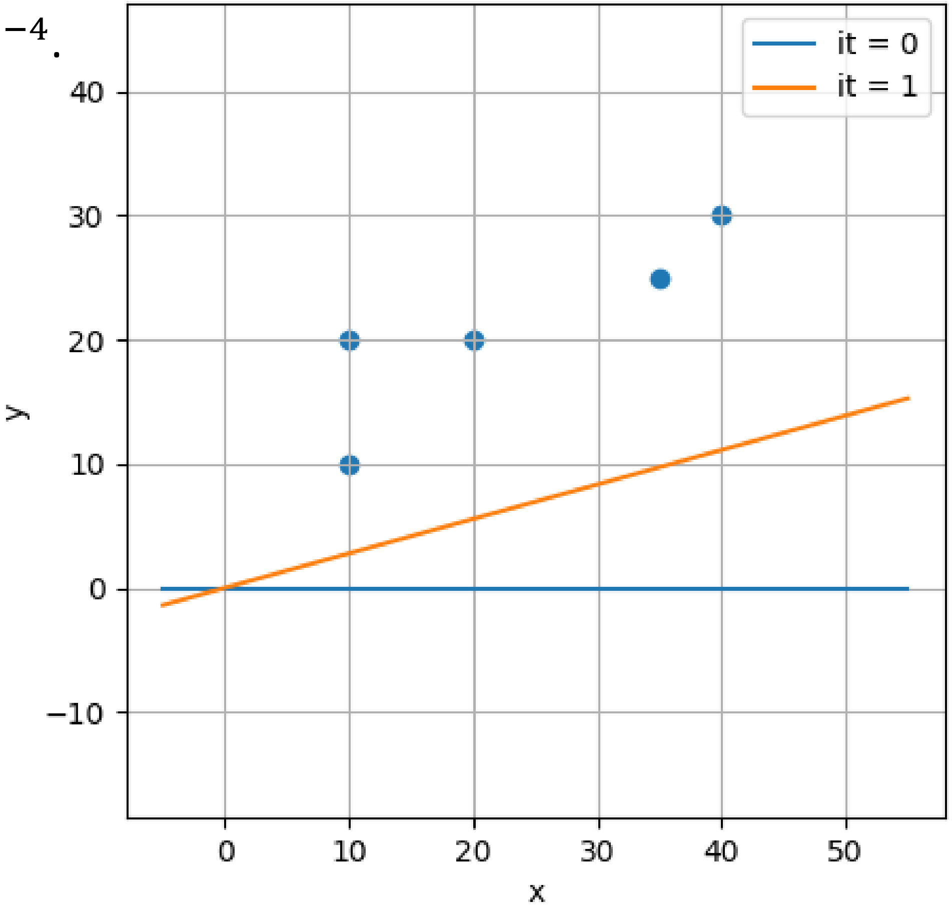
Target variables $\{y^{(1)} = 10, y^{(2)} = 20, y^{(3)} = 20, y^{(4)} = 25, y^{(5)} = 30\}$

(b) *Gradient descent by hand*

(iii) Update the parameters: initial $\mathbf{w} = (0,0)^{\mathrm{T}}$, learning rate $\alpha = 10^{-4}$. Plot both models.

$\boxed{\text{it} = 0}$ $w_0 = 0$, $w_1 = 0$

$\boxed{\text{it} = 1}$ $w_0 = 1.05 \cdot 10^{-2}$, $w_1 = 0.2775$



EXERCICE 1

$g(x) = w_0 + w_1 x$

Feature variables $\{x^{(1)} = 10, x^{(2)} = 10, x^{(3)} = 20, x^{(4)} = 35, x^{(5)} = 40\}$

Target variables $\{y^{(1)} = 10, y^{(2)} = 20, y^{(3)} = 20, y^{(4)} = 25, y^{(5)} = 30\}$

(b) *Gradient descent by hand*

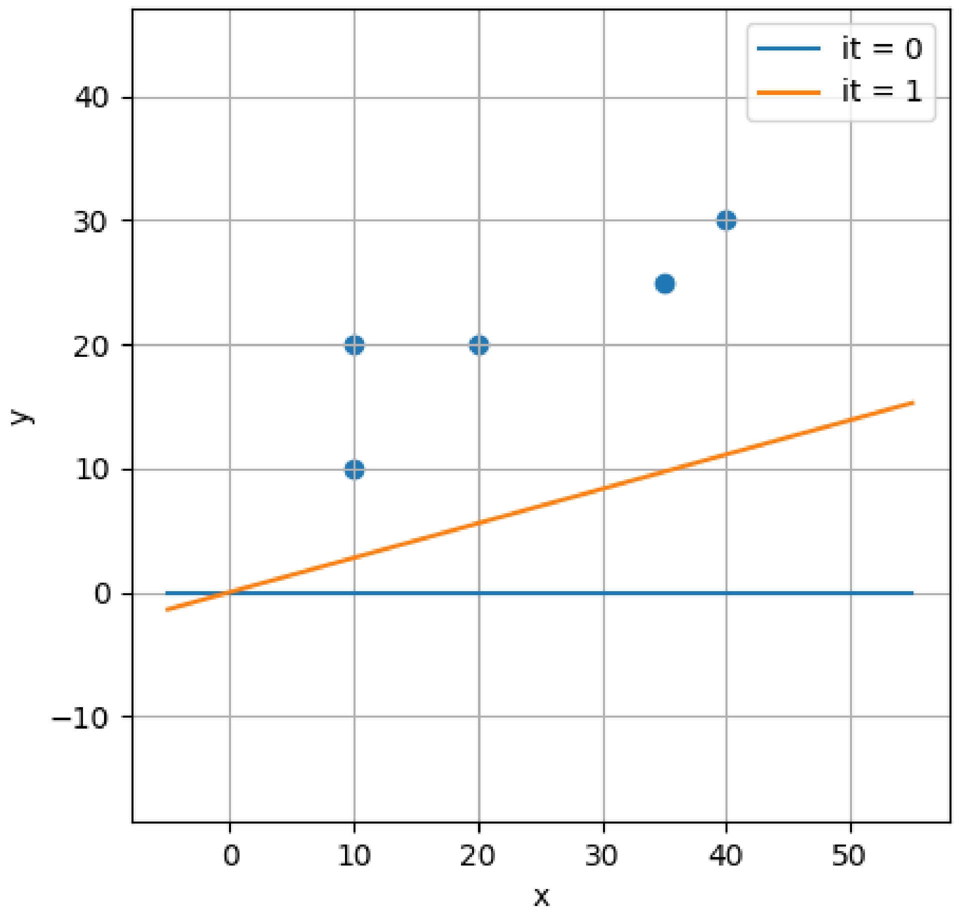
(iv) Compute error. $\mathbb{E}(\mathbf{w}) = \frac{1}{2}\sum_{n=1}^{5}\left(w_0 + w_1 x^{(n)} - y^{(n)}\right)^2$

$\boxed{\text{it} = 0}$ $w_0 = 0$ $w_1 = 0$

$$\mathbb{E}\left(\mathbf{w}^{\text{it}=0}\right) = \frac{1}{2}\sum_{n=1}^{5}\left(-y^{(n)}\right)^2 = ?$$

$\boxed{\text{it} = 1}$ $w_0 = 1.05 \cdot 10^{-2}$ $w_1 = 0.2775$

$$\mathbb{E}\left(\mathbf{w}^{\text{it}=1}\right) = \frac{1}{2}\sum_{n=1}^{5}\left(1.05 \cdot 10^{-2} + 0.2775\, x^{(n)} - y^{(n)}\right)^2 = ?$$



EXERCICE 1

$g(x) = w_0 + w_1 x$

Feature variables $\{x^{(1)} = 10, x^{(2)} = 10, x^{(3)} = 20, x^{(4)} = 35, x^{(5)} = 40\}$

Target variables $\{y^{(1)} = 10, y^{(2)} = 20, y^{(3)} = 20, y^{(4)} = 25, y^{(5)} = 30\}$

(b) *Gradient descent by hand*

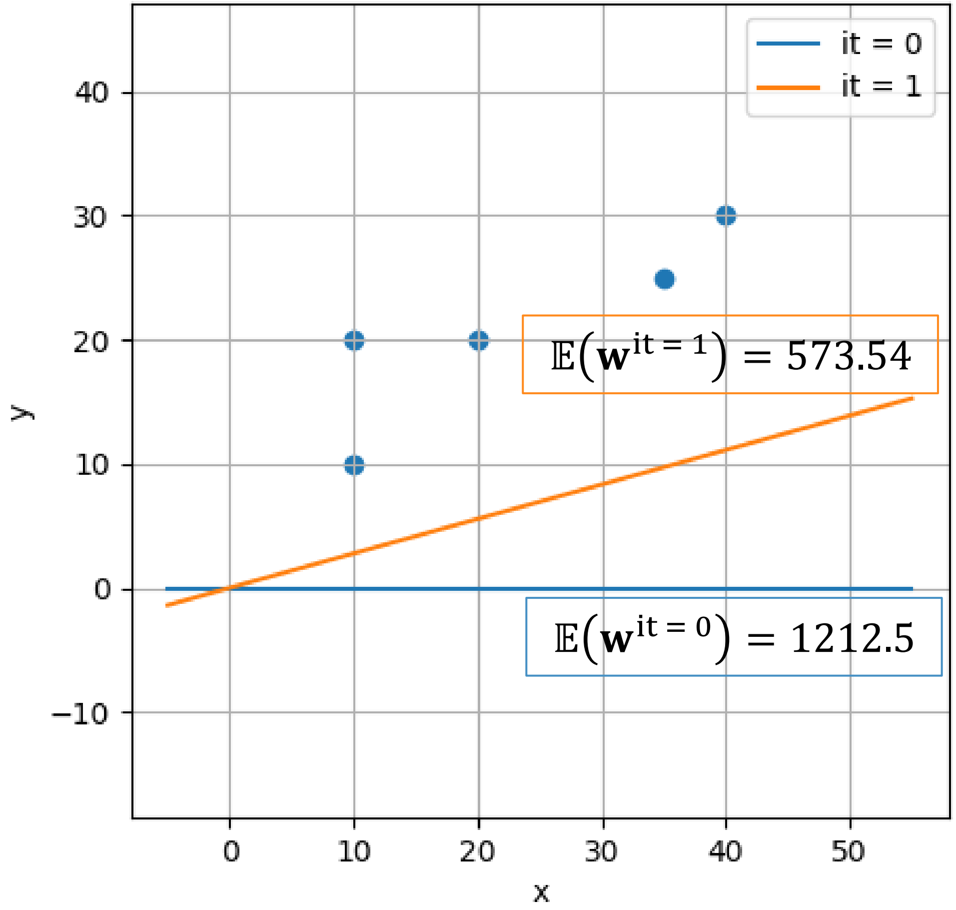
(iv) Compute error. $\mathbb{E}(\mathbf{w}) = \frac{1}{2}\sum_{n=1}^{5}\left(w_0 + w_1 x^{(n)} - y^{(n)}\right)^2$

$\boxed{\text{it} = 0}$ $w_0 = 0$ $w_1 = 0$

$$\mathbb{E}\left(\mathbf{w}^{\text{it}=0}\right) = 1212.5$$

$\boxed{\text{it} = 1}$ $w_0 = 1.05 \cdot 10^{-2}$ $w_1 = 0.2775$

$$\mathbb{E}\left(\mathbf{w}^{\text{it}=1}\right) = 573.54$$



EXERCICE 1

$$g(x) = w_0 + w_1 x$$

Feature variables $\{x^{(1)} = 10, x^{(2)} = 10, x^{(3)} = 20, x^{(4)} = 35, x^{(5)} = 40\}$

Target variables $\{y^{(1)} = 10, y^{(2)} = 20, y^{(3)} = 20, y^{(4)} = 25, y^{(5)} = 30\}$

Interactive 3D in Jupyter Notebook

(b) *Gradient descent by hand*

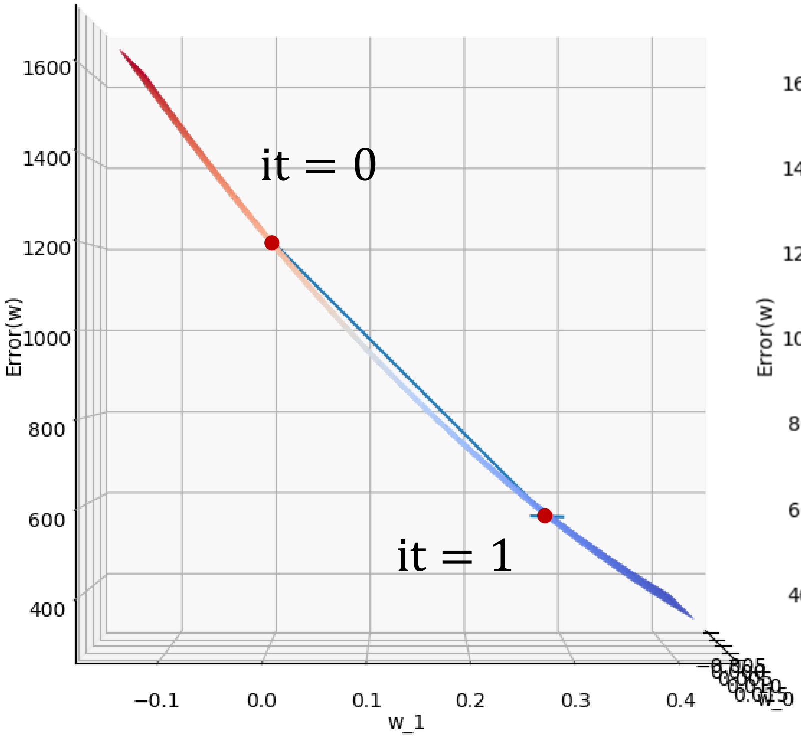
(iv) Compute error.

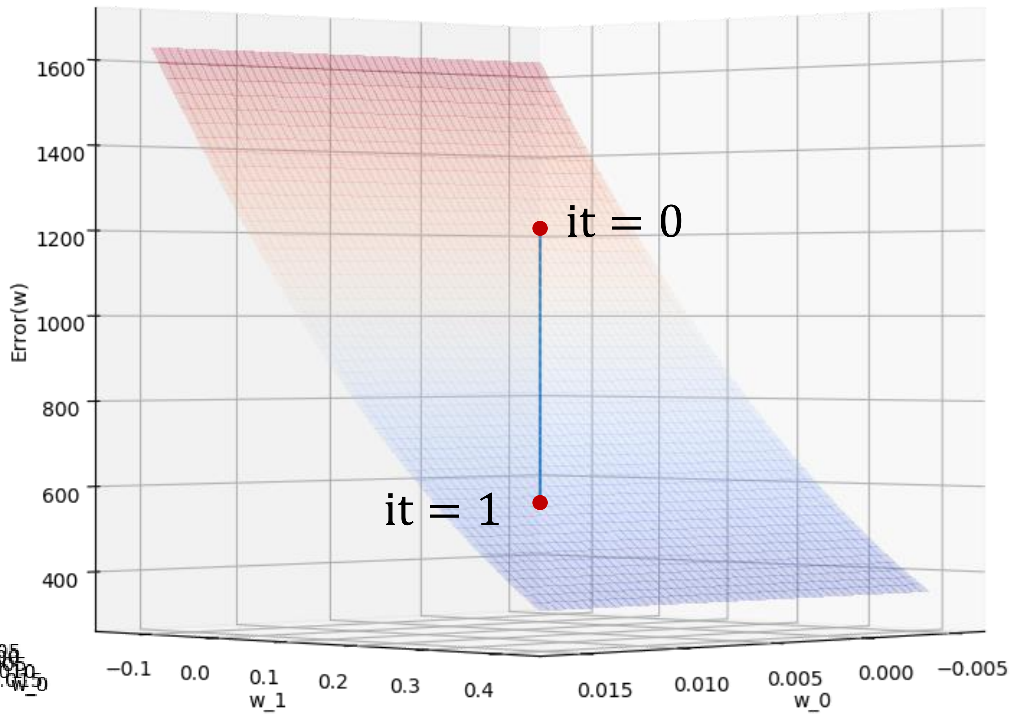
$\boxed{\text{it} = 0}$ $w_0 = 0 \quad w_1 = 0$

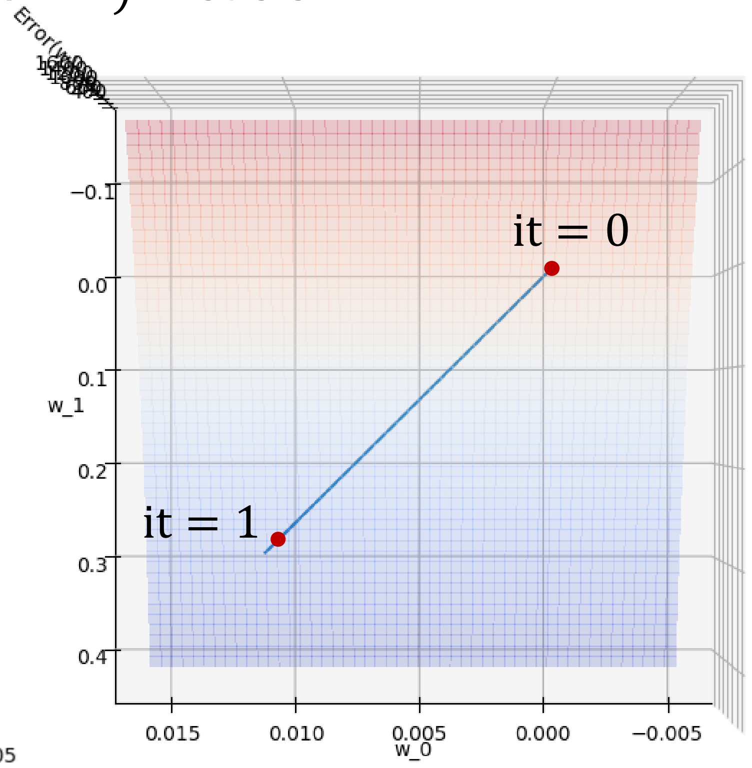
$\mathbb{E}\left(\mathbf{w}^{\text{it}=0}\right) = 1212.5$

$\boxed{\text{it} = 1}$ $w_0 = 1.05 \cdot 10^{-2} \quad w_1 = 0.2775$

$\mathbb{E}\left(\mathbf{w}^{\text{it}=1}\right) = 573.54$







EXERCICE 1

$$g(x) = w_0 + w_1 x$$

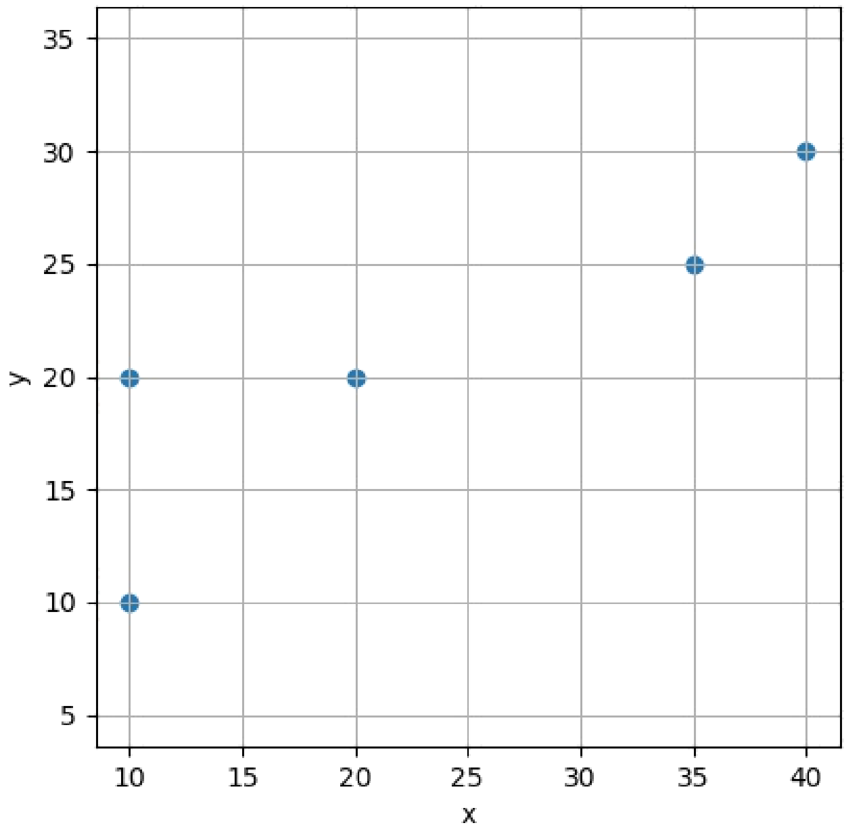
Feature variables $\{x^{(1)} = 10, x^{(2)} = 10, x^{(3)} = 20, x^{(4)} = 35, x^{(5)} = 40\}$

Target variables $\{y^{(1)} = 10, y^{(2)} = 20, y^{(3)} = 20, y^{(4)} = 25, y^{(5)} = 30\}$

(c) *Gradient descent*

(i) Without regularization and Ridge regression ($\lambda = 500$).
Compute error.
Plot both models.
Plot the evolution of $\mathbf{w}$ w.r.t the error.
Which is the effect the regularization?

(ii) Different values for the learning rate: $\alpha = 10^{-6}, 10^{-4}, 3 \cdot 10^{-3}$.
Plot the evolution of $\mathbf{w}$ w.r.t the error.
What are the difference between all the learning parameters?



EXERCICE 1

$g(x) = w_0 + w_1 x$

Feature variables $\{x^{(1)} = 10, x^{(2)} = 10, x^{(3)} = 20, x^{(4)} = 35, x^{(5)} = 40\}$

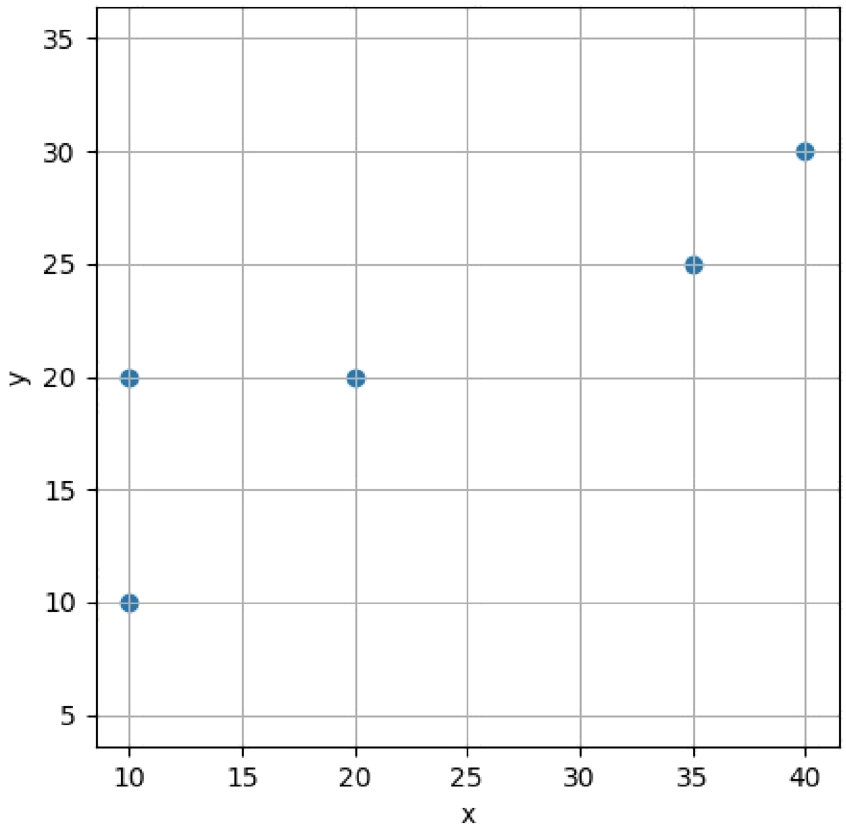
Target variables $\{y^{(1)} = 10, y^{(2)} = 20, y^{(3)} = 20, y^{(4)} = 25, y^{(5)} = 30\}$

(c) *Gradient descent*

(i) Without regularization and Ridge regression ($\lambda = 500$).

DONE IN JUPYTER NOTEBOOK

$$\mathbf{w} = \begin{pmatrix} 0.04 \\ 0.78 \end{pmatrix} \qquad \mathbf{w}_{\text{reg}} = \begin{pmatrix} 0.02 \\ 0.47 \end{pmatrix}$$



EXERCICE 1

$g(x) = w_0 + w_1 x$

Feature variables $\{x^{(1)} = 10, x^{(2)} = 10, x^{(3)} = 20, x^{(4)} = 35, x^{(5)} = 40\}$

Target variables $\{y^{(1)} = 10, y^{(2)} = 20, y^{(3)} = 20, y^{(4)} = 25, y^{(5)} = 30\}$

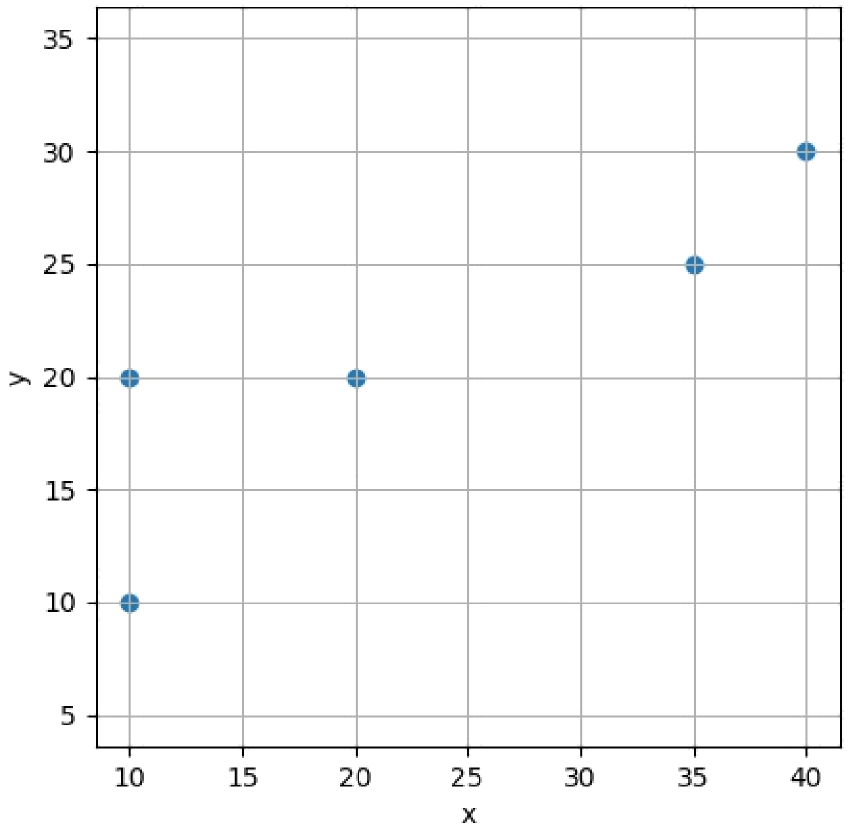
(c) *Gradient descent*

(i) Without regularization and Ridge regression ($\lambda = 500$).
Compute error.

DONE IN JUPYTER NOTEBOOK

$$\mathbf{w} = \begin{pmatrix} 0.04 \\ 0.78 \end{pmatrix} \qquad \mathbf{w}_{\mathbf{reg}} = \begin{pmatrix} 0.02 \\ 0.47 \end{pmatrix}$$

$$\mathbb{E}(\mathbf{w}) = 89.04 < \mathbb{E}(\mathbf{w}_{\mathbf{reg}}) = 289.34$$



EXERCICE 1

$$g(x) = w_0 + w_1 x$$

Feature variables $\{x^{(1)} = 10, x^{(2)} = 10, x^{(3)} = 20, x^{(4)} = 35, x^{(5)} = 40\}$

Target variables $\{y^{(1)} = 10, y^{(2)} = 20, y^{(3)} = 20, y^{(4)} = 25, y^{(5)} = 30\}$

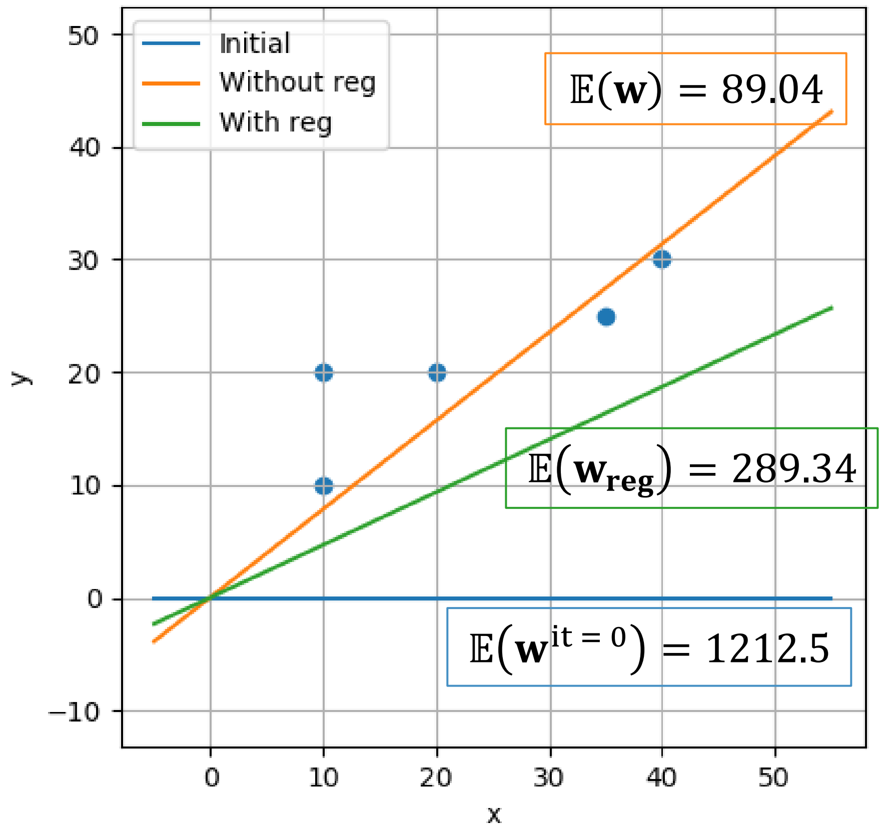
(c) *Gradient descent*

(i) Without regularization and Ridge regression ($\lambda = 500$).
Compute error.
Plot both models.

DONE IN JUPYTER NOTEBOOK

$$\mathbf{w} = \begin{pmatrix} 0.04 \\ 0.78 \end{pmatrix} \qquad \mathbf{w}_{\mathbf{reg}} = \begin{pmatrix} 0.02 \\ 0.47 \end{pmatrix}$$

$$\mathbb{E}(\mathbf{w}) = 89.04 \quad < \quad \mathbb{E}(\mathbf{w}_{\mathbf{reg}}) = 289.34$$



EXERCICE 1

$g(x) = w_0 + w_1 x$

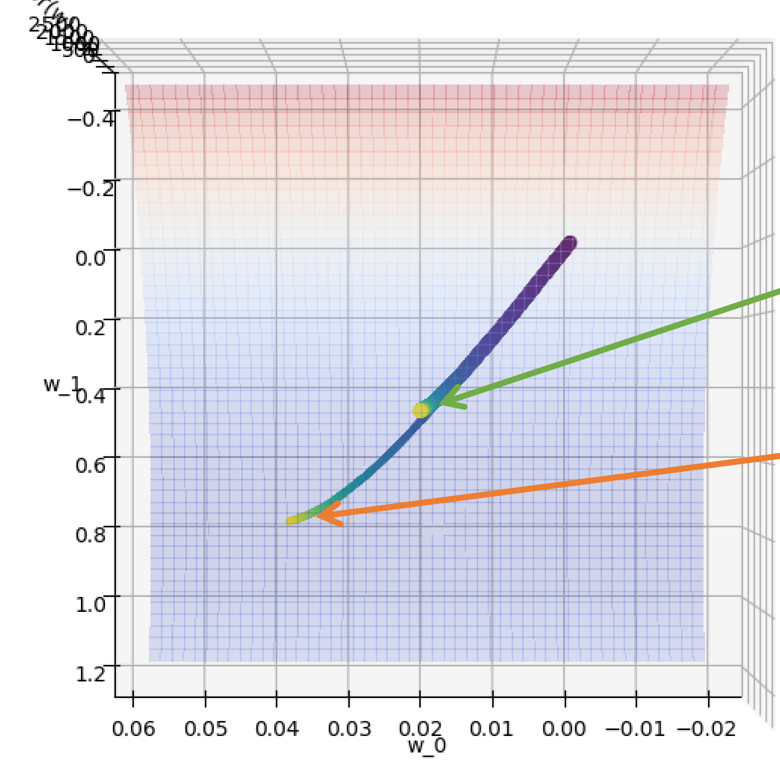
Interactive 3D in Jupyter Notebook

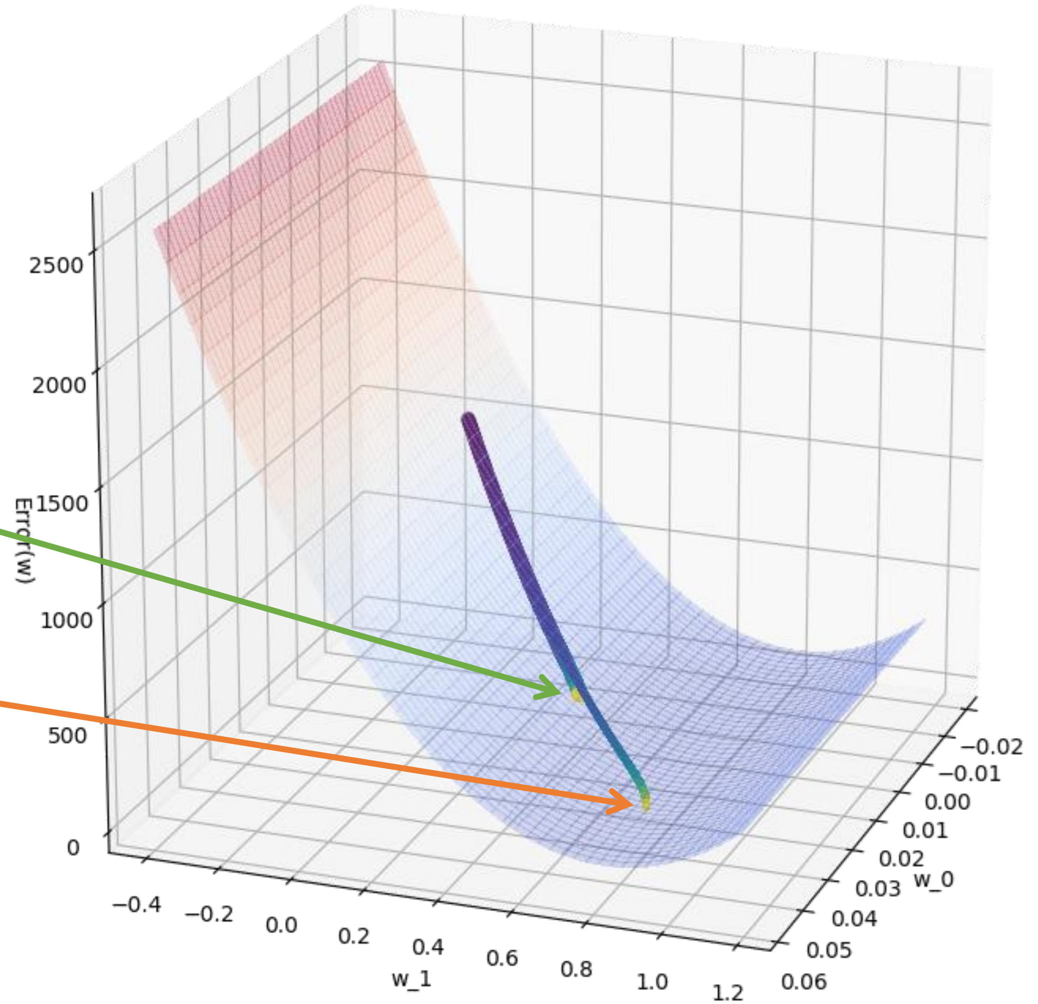
Feature variables $\{x^{(1)} = 10, x^{(2)} = 10, x^{(3)} = 20, x^{(4)} = 35, x^{(5)} = 40\}$

Target variables $\{y^{(1)} = 10, y^{(2)} = 20, y^{(3)} = 20, y^{(4)} = 25, y^{(5)} = 30\}$

(c) *Gradient descent*

(i) Without regularization and Ridge regression ($\lambda = 500$).
Plot the evolution of $\mathbf{w}$ w.r.t the error.





EXERCICE 1

$g(x) = w_0 + w_1 x$

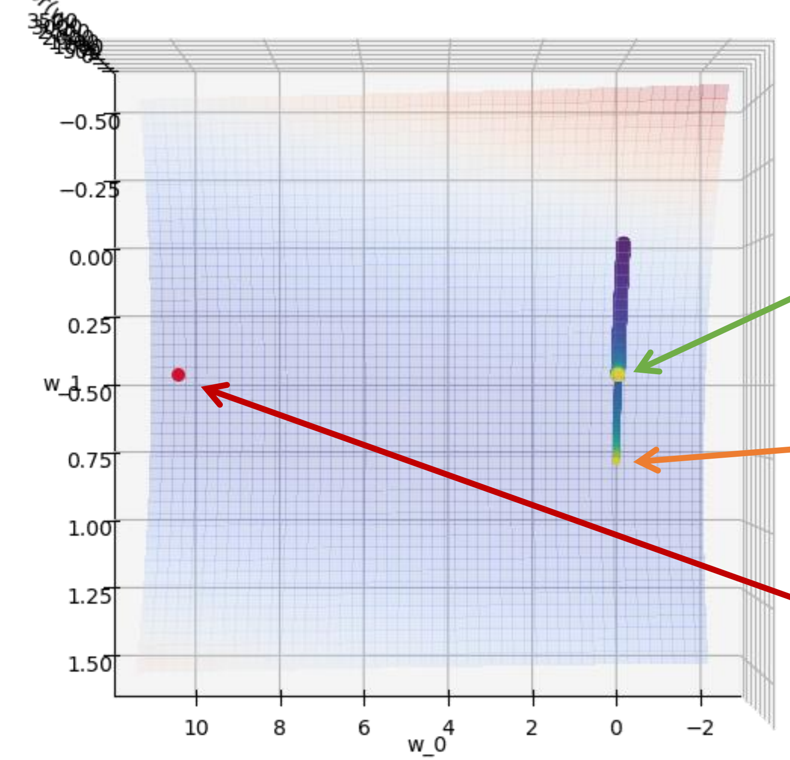
Feature variables $\{x^{(1)} = 10, x^{(2)} = 10, x^{(3)} = 20, x^{(4)} = 35, x^{(5)} = 40\}$

Target variables $\{y^{(1)} = 10, y^{(2)} = 20, y^{(3)} = 20, y^{(4)} = 25, y^{(5)} = 30\}$

Interactive 3D in Jupyter Notebook

(c) *Gradient descent*

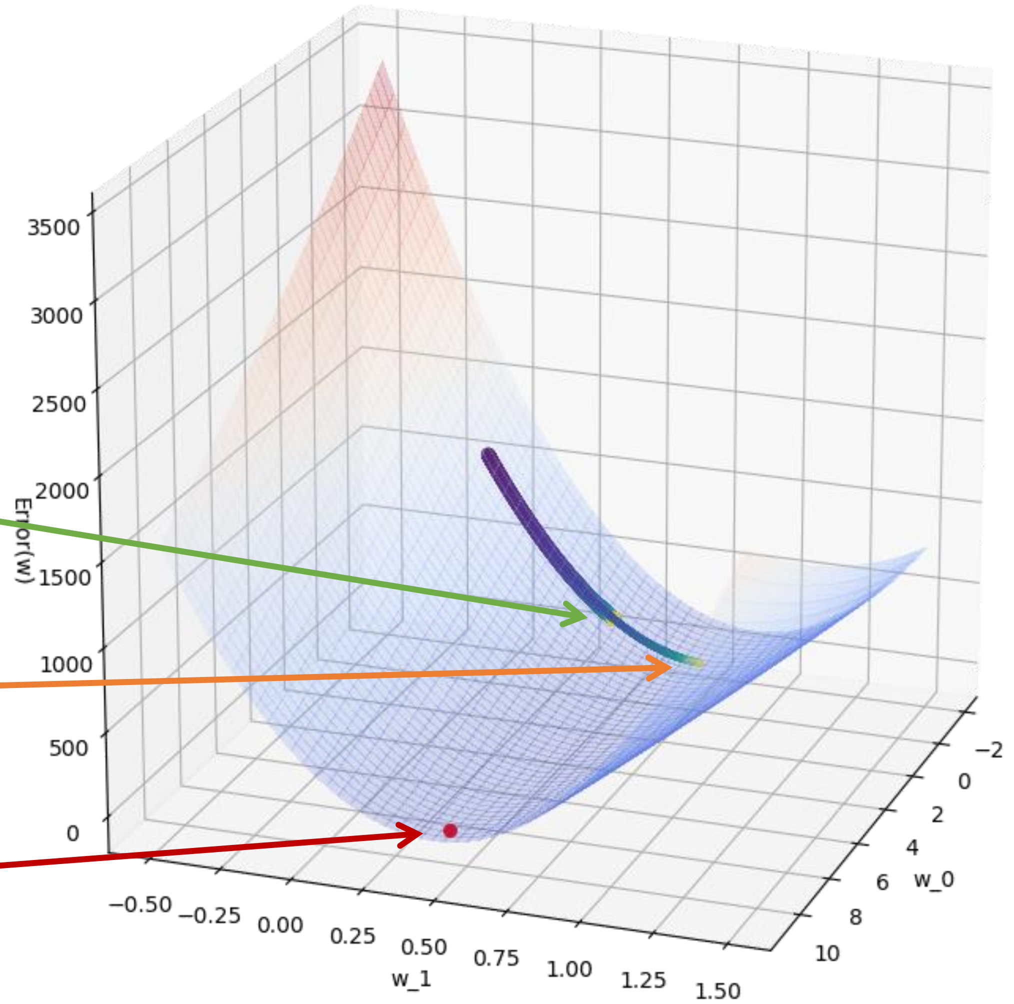
(i) Without regularization and Ridge regression ($\lambda = 500$).
Plot the evolution of $\mathbf{w}$ w.r.t the error.



With regularisation

Without regularisation

Closed form solution



EXERCICE 1

$g(x) = w_0 + w_1 x$

Interactive 3D in Jupyter Notebook

(c) *Gradient descent*

(ii) Different values for the learning rate: $\alpha = 10^{-6}, 10^{-4}, 3 \cdot 10^{-3}$.
Plot the evolution of $\mathbf{w}$ w.r.t the error.