LOGISTIC CLASSIFICATION AND SOFTMAX

REMINDER Logistic and softmax

LOGISTIC

- **Probabilistic** interpretation of the **two-class** linear classification.

- Classifier:

$$\mathrm{h}_{\mathbf{w}}(\mathbf{x}) = p(C_1|\mathbf{x}) = \sigma\left(\mathbf{w}^{\mathrm{T}}\mathbf{x}\right)$$

  where $\sigma(a) = \frac{1}{1+\exp(-a)}$ (logistic function).

- Training:

$$\max_{\mathbf{w}} p(\mathbf{y}|\mathbf{w})$$

  where

$$p(\mathbf{y}|\mathbf{w}) = \prod_{n=1}^{N} {h^{(n)}}^{y^{(n)}} \cdot \left(1 - h^{(n)}\right)^{\left(1-y^{(n)}\right)}$$

Predicted label of nth datapoint

Real label of nth datapoint

SOFTMAX

Likelihood

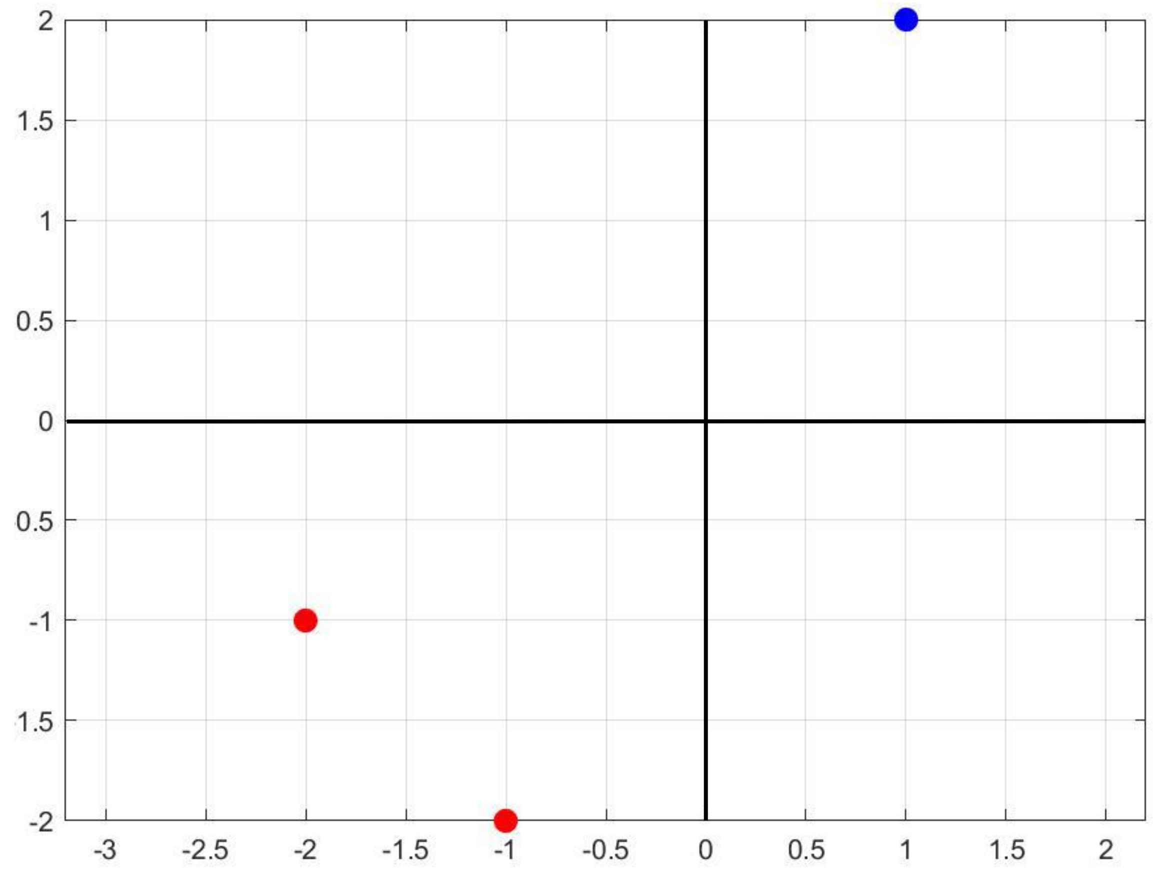
EXERCISE 2

LOGISTIC CLASSIFICATION

EXERCICE 2

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 the parametric form of a logistic classifier $h_{\mathbf{w}}(\mathbf{x})$. What is the difference with a linear classifier?

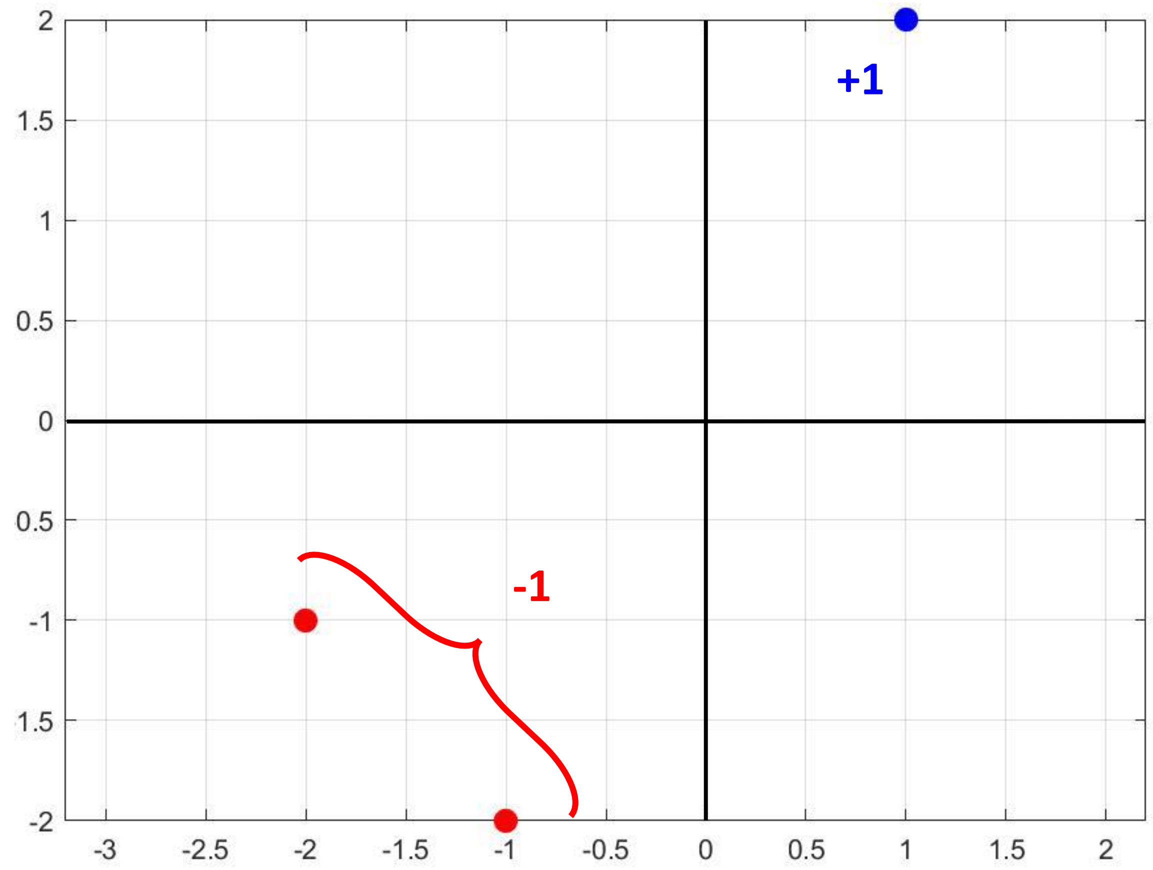


$h_{\mathbf{w}}(\mathbf{x}) = ?$

EXERCICE 2

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 the parametric form of a logistic classifier $\mathrm{h}_{\mathbf{w}}(\mathbf{x})$. What is the difference with a linear classifier?



$$\mathrm{h}_{\mathbf{w}}(\mathbf{x}) = \sigma(\mathbf{w}^{\mathrm{T}}\mathbf{x})$$

where $\mathbf{w}^{\mathrm{T}}\mathbf{x} = w_0 + w_1 x_1 + w_2 x_2$

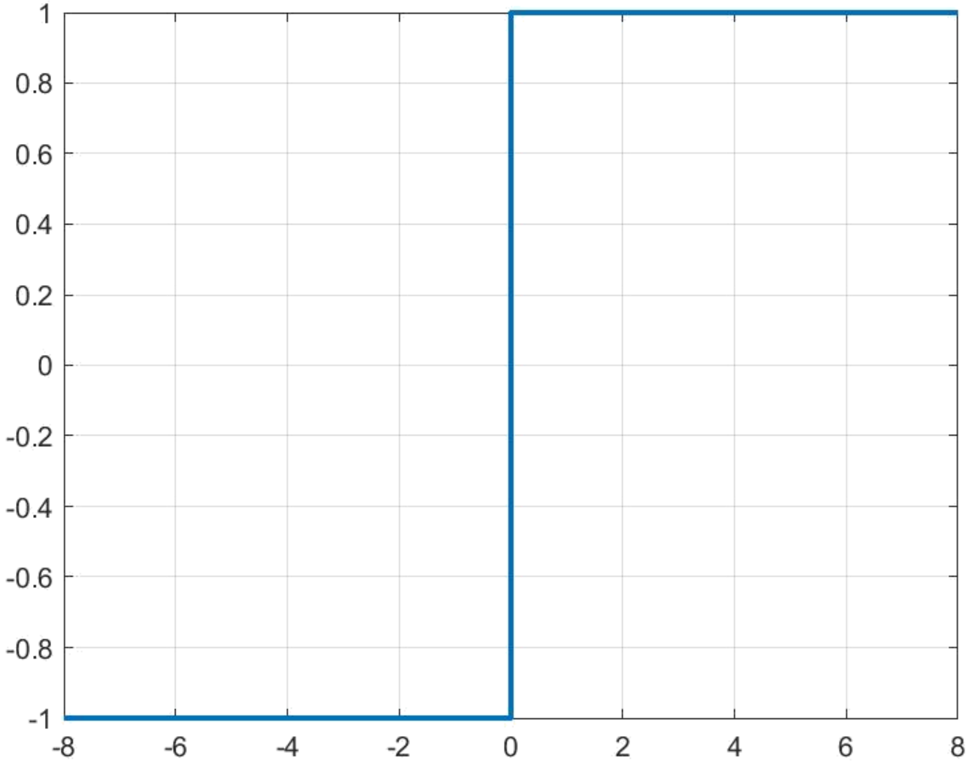
and $\boxed{\sigma(a) = \frac{1}{1+\exp(-a)}}$

EXERCICE 2

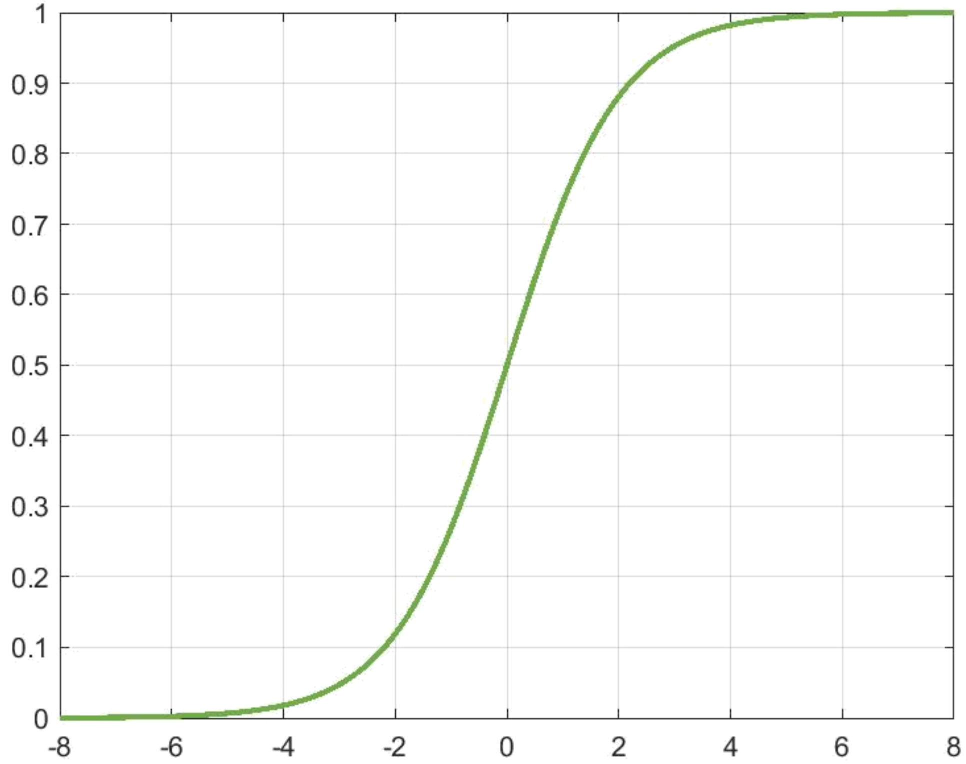
Consider two classes: the class C_1: $\left\{\mathbf{x}^{(1)} = \begin{pmatrix} 1 \\ 2 \end{pmatrix}\right\}$ with label 1 and C_2: $\left\{\mathbf{x}^{(2)} = \begin{pmatrix} -1 \\ -2 \end{pmatrix}, \mathbf{x}^{(3)} = \begin{pmatrix} -2 \\ -1 \end{pmatrix}\right\}$ with label -1

(a) Plot the points and write the parametric form of a logistic classifier $h_{\mathbf{w}}(\mathbf{x})$. What is the difference with a linear classifier?

$$\sigma(a) = \text{sign}(a)$$



$$\sigma(a) = \frac{1}{1 + \exp(-a)}$$

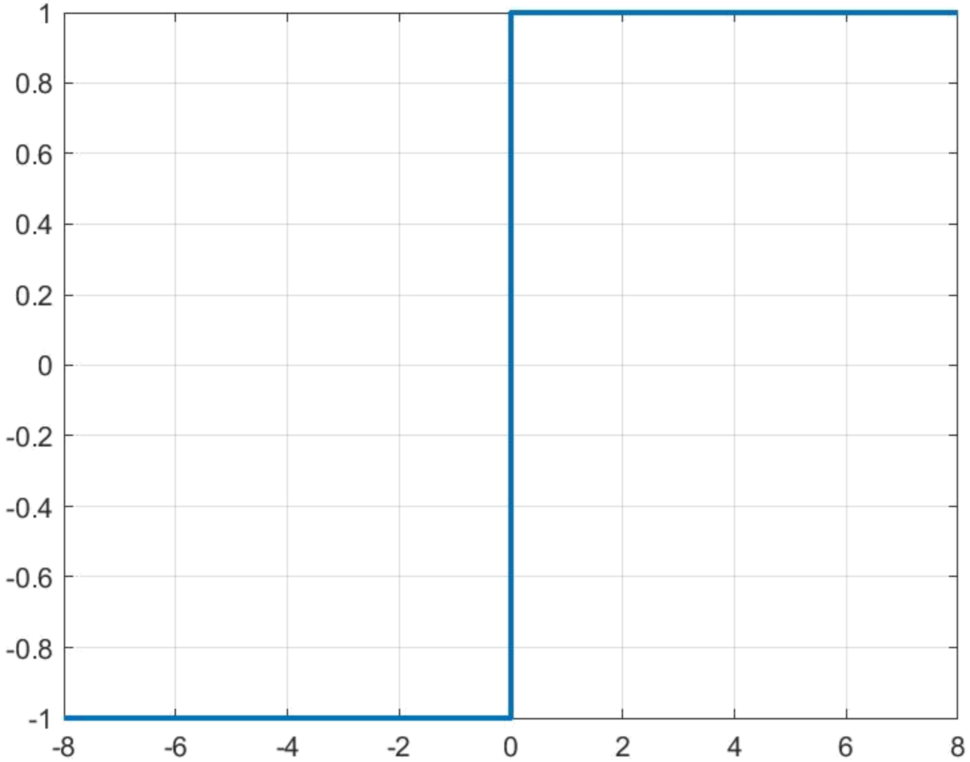


EXERCICE 2

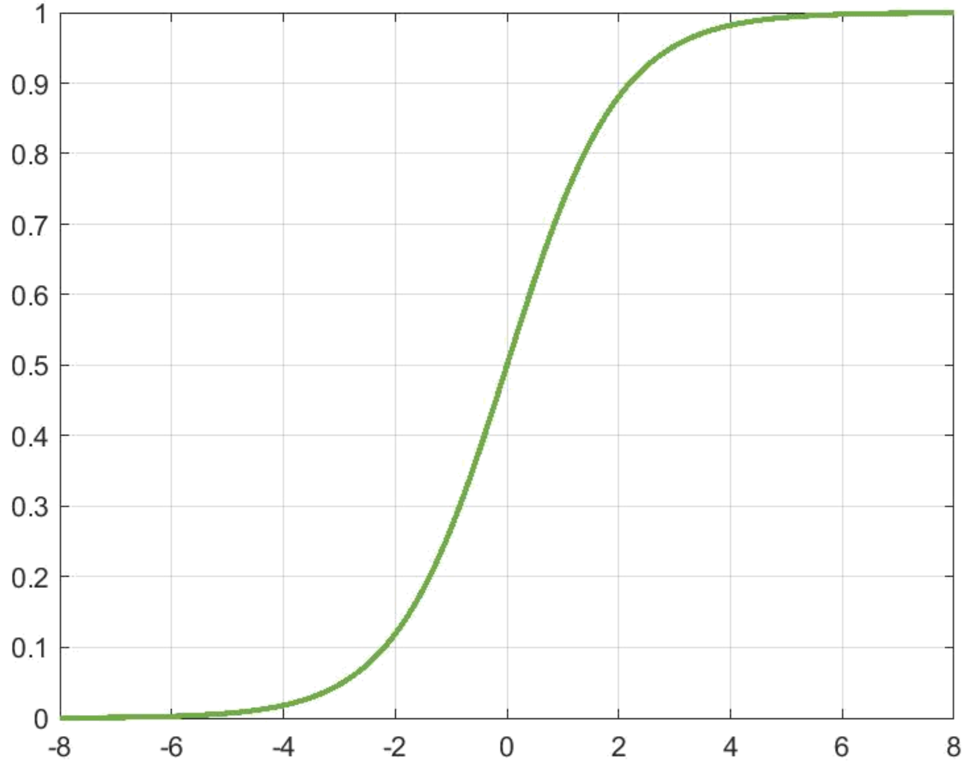
Consider two classes: the class C_1: $\left\{\mathbf{x}^{(1)} = \begin{pmatrix}1\\2\end{pmatrix}\right\}$ with label 1 and C_2: $\left\{\mathbf{x}^{(2)} = \begin{pmatrix}-1\\-2\end{pmatrix}, \mathbf{x}^{(3)} = \begin{pmatrix}-2\\-1\end{pmatrix}\right\}$ with label -1

(b) Labels 1 and -1 are not adequate when solving a two-class problem with logistic regression. Why? Which labels should we choose?

$$\sigma(a) = \mathrm{sign}(a)$$



$$\sigma(a) = \frac{1}{1 + \exp(-a)}$$

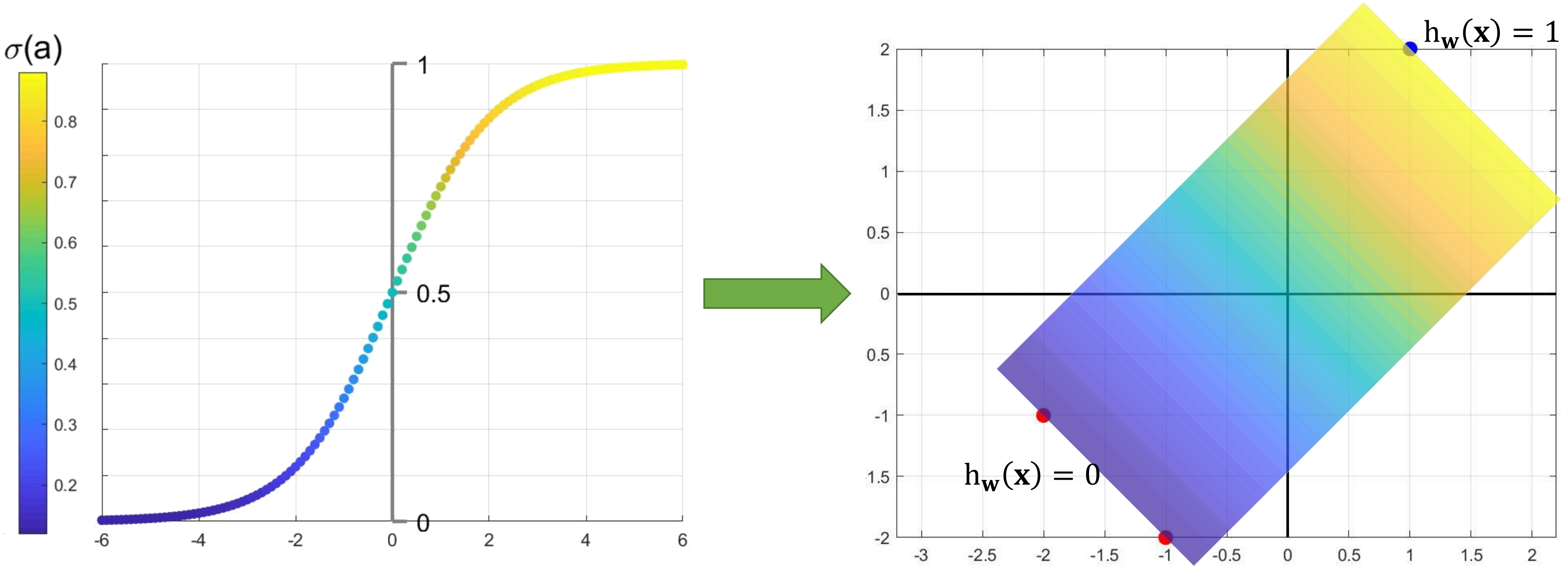


EXERCICE 2

$$h_{\mathbf{w}}(\mathbf{x}) = \sigma\left(\mathbf{w}^T\mathbf{x}\right)$$

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) Labels 1 and -1 are not adequate when solving a two-class problem with logistic regression. Why? Which labels should we choose?

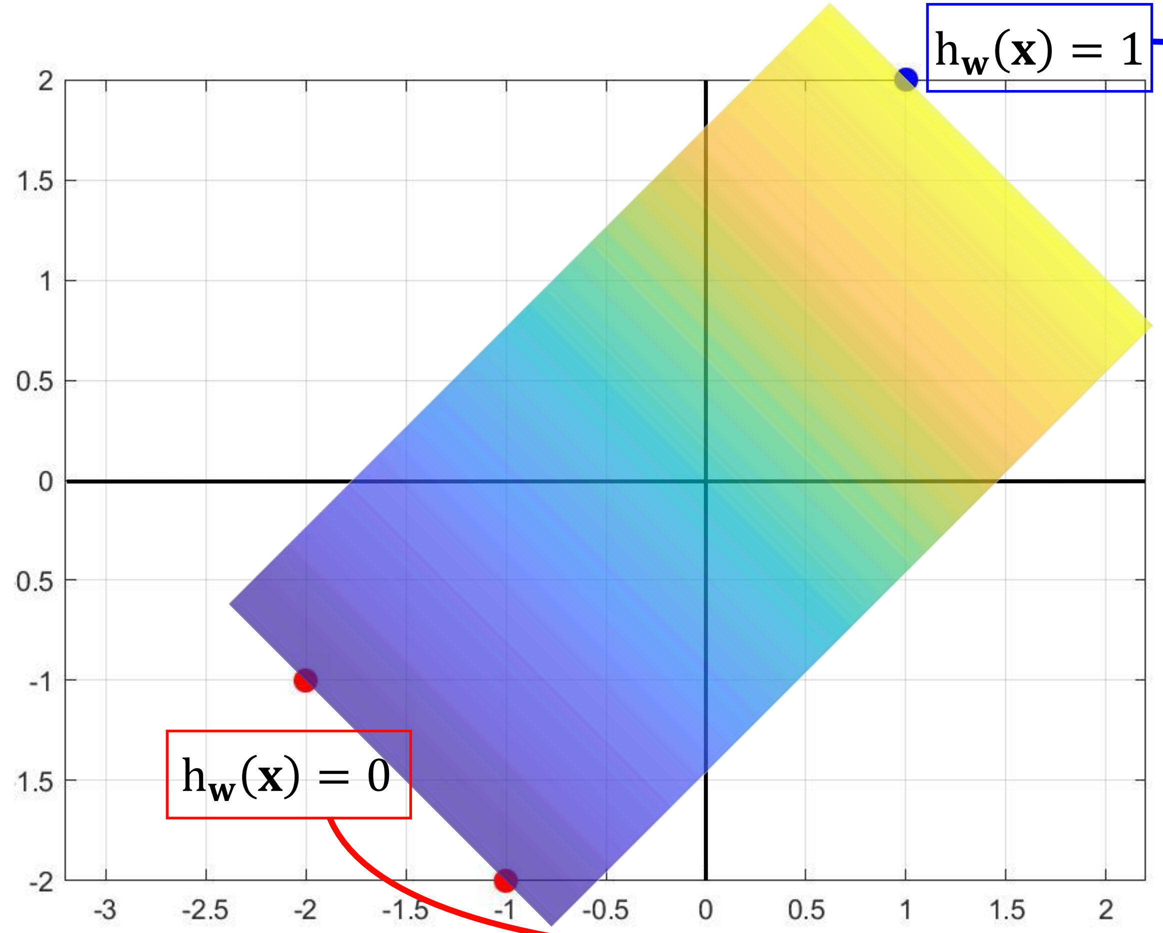


$$h_{\mathbf{w}}(\mathbf{x}) = \sigma(\mathbf{w}^T\mathbf{x})$$

EXERCICE 2

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) Labels 1 and -1 are not adequate when solving a two-class problem with logistic regression. Why? Which labels should we choose?



$C_1: \left\{\mathbf{x}^{(1)} = \begin{pmatrix}1\\2\end{pmatrix}\right\}$ with label 1 $\longrightarrow$ 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 $-1 \longrightarrow$ label 0

EXERCICE 2

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

(c) Write the error function that is minimized when estimating the logistic classifier that separates the given data and develop it.

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

EXERCICE 2

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

(c) Write the error function that is minimized when estimating the logistic classifier that separates the given data and develop it.

$$\max_{\mathbf{w}} p(\mathbf{y}|\mathbf{w}) \quad \text{where} \quad p(\mathbf{y}|\mathbf{w}) = \prod_{n=1}^{N} {h^{(n)}}^{y^{(n)}} \cdot \left(1 - h^{(n)}\right)^{\left(1-y^{(n)}\right)}$$

$$h^{(n)} = h_{\mathbf{w}}\left(\mathbf{x}^{(n)}\right) = \sigma\left(\mathbf{w}^{\mathrm{T}}\mathbf{x}^{(n)}\right) = \frac{1}{1 + \exp\left(\mathbf{w}^{\mathrm{T}}\mathbf{x}^{(n)}\right)}$$

the prediction of the nth point

is equivalent to

$$\min_{\mathbf{w}} \mathbb{E}(\mathbf{w}) \quad \text{where} \quad \mathbb{E}(\mathbf{w}) = -\ln p(\mathbf{y}|\mathbf{w}) = -\sum_{n=1}^{N} y^{(n)} \cdot \ln h^{(n)} + \left(1 - y^{(n)}\right) \ln\left(1 - h^{(n)}\right)$$

EXERCICE 2

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

(d) *(Optional)* Derive the expression of the error from the previous exercise with respect to the weight vector.

$$\mathbb{E}(\mathbf{w}) = -\ln p(\mathbf{y}|\mathbf{w}) = -\sum_{n=1}^{N} y^{(n)} \cdot \ln h^{(n)} + \left(1 - y^{(n)}\right) \ln\left(1 - h^{(n)}\right)$$

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

... OPTIONAL HOMEWORK ...

EXERCICE 2

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

(e) *(Jupyter)* Generate a gradient descend algorithm with ridge regression. To do so, you only have to add to the gradient an extra term $\lambda \mathbf{w}$, where λ is the regularization parameter. Plot the data and the resulting linear classifier in the same figure.

REMINDER – Gradient descent

Until convergence:

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

Linear classifier: $\frac{\partial \mathbb{E}(\mathbf{w})}{\partial \mathbf{w}} = \sum_{n=1}^{5} \left(\mathbf{w}^{\mathrm{T}}\mathbf{x}^{(n)} - y^{(n)}\right) \cdot \mathbf{x}^{(n)}$, where $\mathbf{w}^{\mathrm{T}}\mathbf{x}^{(n)} = h_{\mathbf{w}}\left(\mathbf{x}^{(n)}\right)$ linear

Logistic classifier: $\frac{\partial \mathbb{E}(\mathbf{w})}{\partial \mathbf{w}} = \sum_{n=1}^{5} \left(\sigma\left(\mathbf{w}^{\mathrm{T}}\mathbf{x}^{(n)}\right) - y^{(n)}\right) \cdot \mathbf{x}^{(n)}$, where $\sigma\left(\mathbf{w}^{\mathrm{T}}\mathbf{x}^{(n)}\right) = h_{\mathbf{w}}\left(\mathbf{x}^{(n)}\right)$ logistic

EXERCICE 2

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

(e) *(Jupyter)* Generate a gradient descend algorithm with ridge regression. To do so, you only have to add to the gradient an extra term $\lambda\mathbf{w}$, where λ is the regularization parameter. Plot the data and the resulting linear classifier in the same figure.

REMINDER – Gradient descent

Until convergence:

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

Linear classifier: $$\frac{\partial \mathbb{E}(\mathbf{w})}{\partial \mathbf{w}} = \sum_{n=1}^{5}\left(\overbrace{\mathbf{w}^{\mathrm{T}}\mathbf{x}^{(n)}}^{h_{\mathbf{w}}(\mathbf{x}^{(n)})\text{ linear}} - y^{(n)}\right)\cdot\mathbf{x}^{(n)}$$

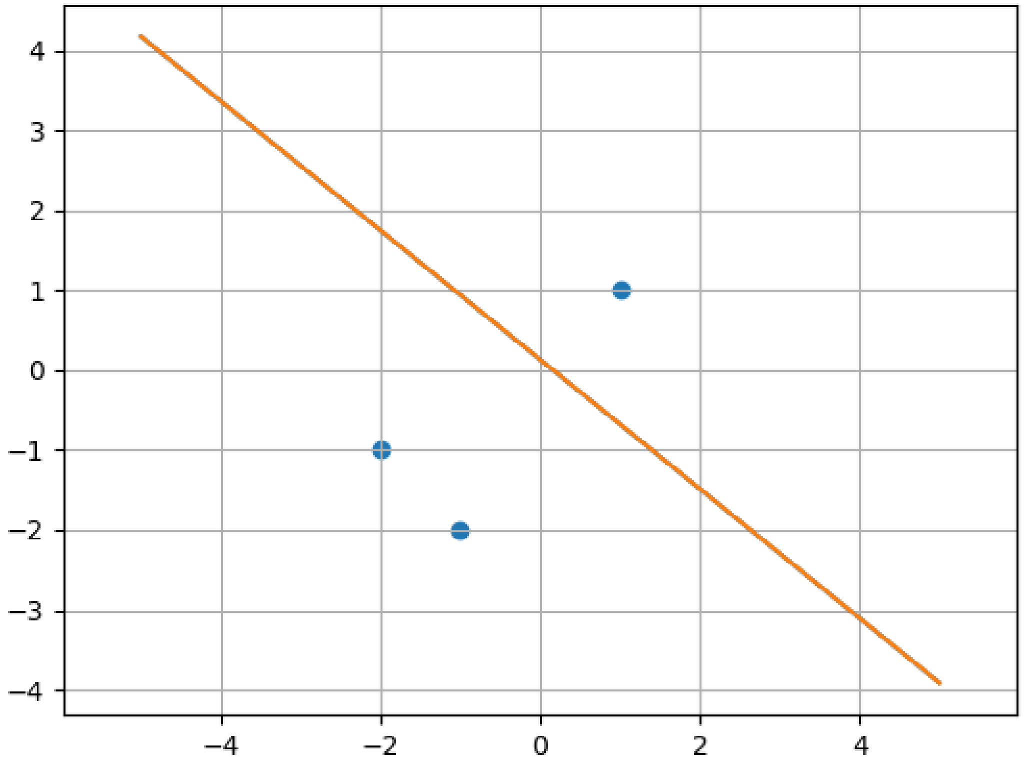
Logistic classifier: $$\frac{\partial \mathbb{E}(\mathbf{w})}{\partial \mathbf{w}} = \sum_{n=1}^{5}\left(\underbrace{\sigma\left(\mathbf{w}^{\mathrm{T}}\mathbf{x}^{(n)}\right)}_{h_{\mathbf{w}}(\mathbf{x}^{(n)})\text{ logistic}} - y^{(n)}\right)\cdot\mathbf{x}^{(n)}$$

Regularisation term
Rigde Regression

EXERCICE 2

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

(e) *(Jupyter)* Generate a gradient descend algorithm with ridge regression. To do so, you only have to add to the gradient an extra term $\lambda \mathbf{w}$, where λ is the regularization parameter. Plot the data and the resulting linear classifier in the same figure.



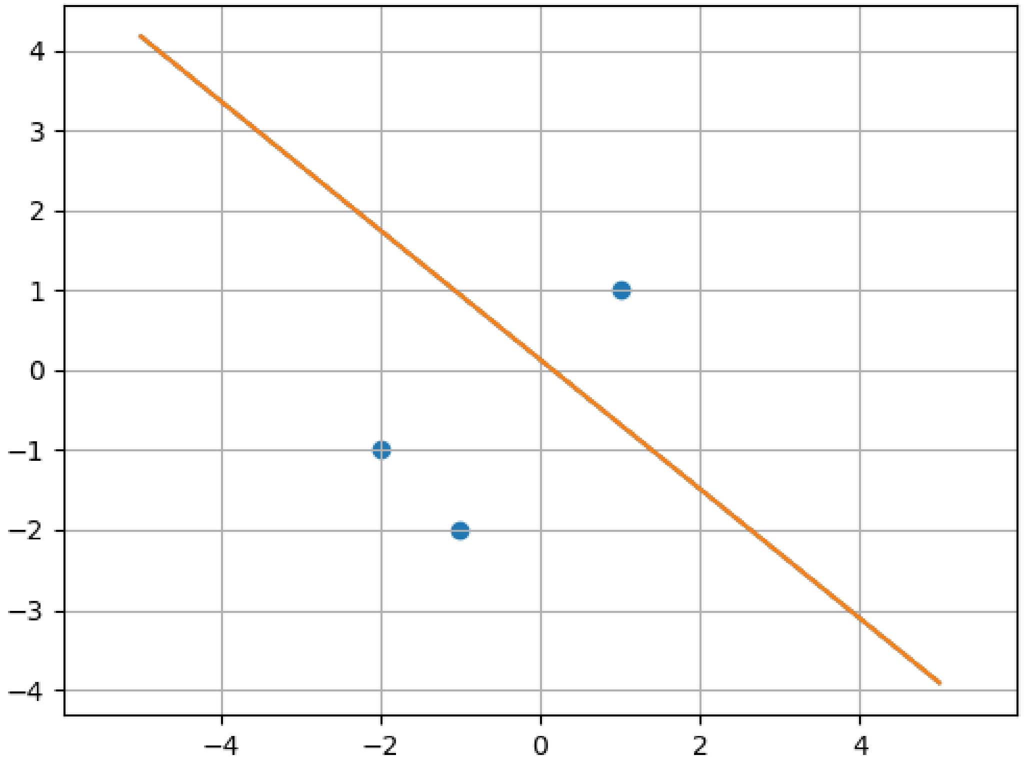
$\lambda = 10^{-3}, \alpha = 0.1$

Remember the effect?

EXERCICE 2

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

(f) *(Jupyter, Optional)* Estimate the logistic classifier using the function *sklearn.linear model.LogisticRegression*. Draw the data points and the resulting linear classier in the same figure.



Should be similar to ours.

Even tough it will depend on the regularisation parameter λ and the learning rate α.

EXERCISE 3

SOFTMAX CLASSIFICATION

REMINDER Logistic and softmax

LOGISTIC

- **Probabilistic** interpretation of the **two-class** linear classification.
- Classifier:

$$\mathrm{h}_{\mathbf{w}}(\mathbf{x}) = p(C_1|\mathbf{x}) = \sigma(\mathbf{w}^{\mathrm{T}}\mathbf{x})$$

where $\sigma(a) = \frac{1}{1+\exp(-a)}$ (logistic function).

- Training:

$$\min_{\mathbf{w}} \mathbb{E}(\mathbf{w})$$

where

$$\mathbb{E}(\mathbf{w}) = -\sum_{n=1}^{N} y^{(n)} \cdot \ln h^{(n)} + \left(1 - y^{(n)}\right) \ln\left(1 - h^{(n)}\right)$$

Log-likelihood

SOFTMAX

- **Generalisation** of the logistic to **multi-class** classification.
- Classifier:

$$\textit{assign } \mathbf{x} \textit{ to class } C_k \textit{ if } h_k(x) > h_l(x)\ \forall l \neq k$$

where

$$\mathrm{h}_k(\mathbf{x}) = p(C_k|\mathbf{x}) = \frac{\exp(\mathbf{w}_k^{\mathrm{T}}\mathbf{x})}{\sum_{j=1}^{K} \exp(\mathbf{w}_j^{\mathrm{T}}\mathbf{x})}.$$

Softmax function

- Training:

$$\min_{\mathbf{w}} \mathbb{E}(\mathbf{w})$$

where

$$\mathbb{E}(\mathbf{w}) = -\sum_{n=1}^{N}\sum_{k=1}^{K} y_k^{(n)} \cdot \ln h_k^{(n)}$$

Cross-entropy

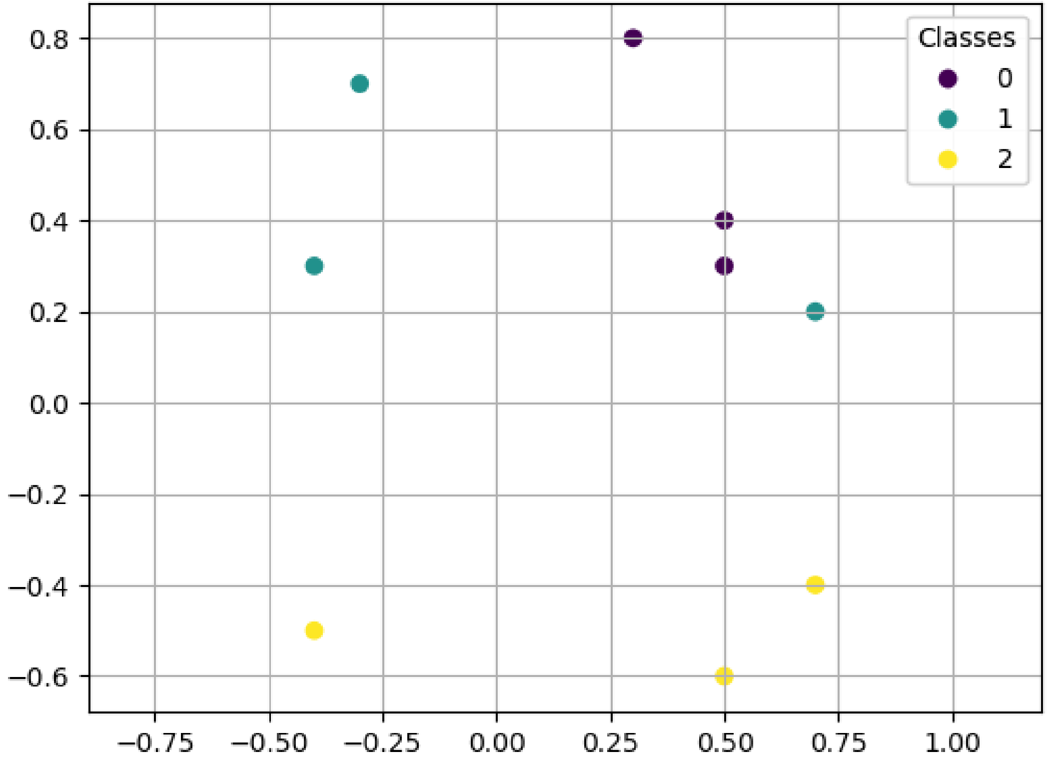
EXERCICE 3

Consider **three** classes:

Class 0 C_0: $\left\{\mathbf{x}^{(1)} = \begin{pmatrix}0.5\\0.4\end{pmatrix}, \mathbf{x}^{(2)} = \begin{pmatrix}0.8\\0.3\end{pmatrix}, \mathbf{x}^{(3)} = \begin{pmatrix}0.3\\0.8\end{pmatrix}\right\}$,

Class 1 C_1: $\left\{\mathbf{x}^{(4)} = \begin{pmatrix}-0.4\\0.3\end{pmatrix}, \mathbf{x}^{(5)} = \begin{pmatrix}-0.3\\0.7\end{pmatrix}, \mathbf{x}^{(6)} = \begin{pmatrix}0.7\\0.2\end{pmatrix}\right\}$,

Class 2 C_2: $\left\{\mathbf{x}^{(7)} = \begin{pmatrix}0.7\\-0.4\end{pmatrix}, \mathbf{x}^{(8)} = \begin{pmatrix}0.5\\-0.6\end{pmatrix}, \mathbf{x}^{(9)} = \begin{pmatrix}-0.4\\-0.5\end{pmatrix}\right\}$



Assume that the values obtained for the three discriminants are $\mathbf{w}_0 = \begin{pmatrix}-0.3\\0.87\\1.47\end{pmatrix}$, $\mathbf{w}_1 = \begin{pmatrix}-0.01\\0.58\\1.02\end{pmatrix}$ and $\mathbf{w}_2 = \begin{pmatrix}0.43\\-1.90\\0.33\end{pmatrix}$

EXERCICE 3

(a) Draw the points and codify the class of the vectors according to $1 - \text{of} - K$ encoding.

$\boxed{1 - \text{of} - K \text{ encoding}}$ → for class j, target output $y = (0, \dots, 0, 1, 0, \dots, 0)$

a.k.a hot encoding or dummy encoding

jth coefficient

Class 0 C_0: $\left\{\mathbf{x}^{(1)} = \begin{pmatrix}0.5\\0.4\end{pmatrix}, \mathbf{x}^{(2)} = \begin{pmatrix}0.8\\0.3\end{pmatrix}, \mathbf{x}^{(3)} = \begin{pmatrix}0.3\\0.8\end{pmatrix}\right\}$, → $\mathbf{y}^{(1)} = \mathbf{y}^{(2)} = \mathbf{y}^{(3)} = ?$

Class 1 C_1: $\left\{\mathbf{x}^{(4)} = \begin{pmatrix}-0.4\\0.3\end{pmatrix}, \mathbf{x}^{(5)} = \begin{pmatrix}-0.3\\0.7\end{pmatrix}, \mathbf{x}^{(6)} = \begin{pmatrix}0.7\\0.2\end{pmatrix}\right\}$, → $\mathbf{y}^{(4)} = \mathbf{y}^{(5)} = \mathbf{y}^{(6)} = ?$

Class 2 C_2: $\left\{\mathbf{x}^{(7)} = \begin{pmatrix}0.7\\-0.4\end{pmatrix}, \mathbf{x}^{(8)} = \begin{pmatrix}0.5\\-0.6\end{pmatrix}, \mathbf{x}^{(9)} = \begin{pmatrix}-0.4\\-0.5\end{pmatrix}\right\}$ → $\mathbf{y}^{(7)} = \mathbf{y}^{(8)} = \mathbf{y}^{(9)} = ?$

EXERCICE 3

(a) Draw the points and codify the class of the vectors according to $1-\text{of}-K$ encoding.

$\boxed{1-\text{of}-K \text{ encoding}}$ → for class j, target output $y = (0, \dots, 0, 1, 0, \dots, 0)$

jth coefficient

Class 0 C_0: $\left\{\mathbf{x}^{(1)} = \begin{pmatrix} 0.5 \\ 0.4 \end{pmatrix}, \mathbf{x}^{(2)} = \begin{pmatrix} 0.8 \\ 0.3 \end{pmatrix}, \mathbf{x}^{(3)} = \begin{pmatrix} 0.3 \\ 0.8 \end{pmatrix}\right\}$, → $\mathbf{y}^{(1)} = \mathbf{y}^{(2)} = \mathbf{y}^{(3)} = (1,0,0)^{\mathrm{T}}$

Class 1 C_1: $\left\{\mathbf{x}^{(4)} = \begin{pmatrix} -0.4 \\ 0.3 \end{pmatrix}, \mathbf{x}^{(5)} = \begin{pmatrix} -0.3 \\ 0.7 \end{pmatrix}, \mathbf{x}^{(6)} = \begin{pmatrix} 0.7 \\ 0.2 \end{pmatrix}\right\}$, → $\mathbf{y}^{(4)} = \mathbf{y}^{(5)} = \mathbf{y}^{(6)} = (0,1,0)^{\mathrm{T}}$

Class 2 C_2: $\left\{\mathbf{x}^{(7)} = \begin{pmatrix} 0.7 \\ -0.4 \end{pmatrix}, \mathbf{x}^{(8)} = \begin{pmatrix} 0.5 \\ -0.6 \end{pmatrix}, \mathbf{x}^{(9)} = \begin{pmatrix} -0.4 \\ -0.5 \end{pmatrix}\right\}$ → $\mathbf{y}^{(7)} = \mathbf{y}^{(8)} = \mathbf{y}^{(9)} = (0,0,1)^{\mathrm{T}}$

EXERCICE 3

Class 0 C_0: $\left\{\mathbf{x}^{(1)} = \begin{pmatrix}0.5\\0.4\end{pmatrix}, \mathbf{x}^{(2)} = \begin{pmatrix}0.8\\0.3\end{pmatrix}, \mathbf{x}^{(3)} = \begin{pmatrix}0.3\\0.8\end{pmatrix}\right\}$,

Class 1 C_1: $\left\{\mathbf{x}^{(4)} = \begin{pmatrix}-0.4\\0.3\end{pmatrix}, \mathbf{x}^{(5)} = \begin{pmatrix}-0.3\\0.7\end{pmatrix}, \mathbf{x}^{(6)} = \begin{pmatrix}0.7\\0.2\end{pmatrix}\right\}$,

Class 2 C_2: $\left\{\mathbf{x}^{(7)} = \begin{pmatrix}0.7\\-0.4\end{pmatrix}, \mathbf{x}^{(8)} = \begin{pmatrix}0.5\\-0.6\end{pmatrix}, \mathbf{x}^{(9)} = \begin{pmatrix}-0.4\\-0.5\end{pmatrix}\right\}$

$$\mathbf{w}_0 = \begin{pmatrix}-0.3\\0.87\\1.47\end{pmatrix}, \mathbf{w}_1 = \begin{pmatrix}-0.01\\0.58\\1.02\end{pmatrix}$$

$$\mathbf{w}_2 = \begin{pmatrix}0.43\\-1.90\\0.33\end{pmatrix}$$

(b) Determine at which class the calculated softmax will classify the points raw the points $\mathbf{x}^{(1)}$, $\mathbf{x}^{(4)}$ and $\mathbf{x}^{(7)}$.

REMINDER – Softmax

$$h_k(\mathbf{x}) = p(C_k|\mathbf{x}) = \frac{\exp\left(\mathbf{w}_k^{\mathrm{T}}\mathbf{x}\right)}{\sum_{j=1}^{K}\exp\left(\mathbf{w}_j^{\mathrm{T}}\mathbf{x}\right)} \quad \forall k = 0,1,2 \longrightarrow$$

We classify the point $\mathbf{x}$ to the class with **higher probability**.

Probability of $\mathbf{x}$ of being class k.

EXERCICE 3

Class 0 C_0: $\left\{\mathbf{x}^{(1)} = \begin{pmatrix}0.5\\0.4\end{pmatrix}, \mathbf{x}^{(2)} = \begin{pmatrix}0.8\\0.3\end{pmatrix}, \mathbf{x}^{(3)} = \begin{pmatrix}0.3\\0.8\end{pmatrix}\right\}$,

Class 1 C_1: $\left\{\mathbf{x}^{(4)} = \begin{pmatrix}-0.4\\0.3\end{pmatrix}, \mathbf{x}^{(5)} = \begin{pmatrix}-0.3\\0.7\end{pmatrix}, \mathbf{x}^{(6)} = \begin{pmatrix}0.7\\0.2\end{pmatrix}\right\}$,

Class 2 C_2: $\left\{\mathbf{x}^{(7)} = \begin{pmatrix}0.7\\-0.4\end{pmatrix}, \mathbf{x}^{(8)} = \begin{pmatrix}0.5\\-0.6\end{pmatrix}, \mathbf{x}^{(9)} = \begin{pmatrix}-0.4\\-0.5\end{pmatrix}\right\}$

$$\mathbf{w}_0 = \begin{pmatrix}-0.3\\0.87\\1.47\end{pmatrix}, \mathbf{w}_1 = \begin{pmatrix}-0.01\\0.58\\1.02\end{pmatrix}, \mathbf{w}_2 = \begin{pmatrix}0.43\\-1.90\\0.33\end{pmatrix}$$

(b) Determine at which class the calculated softmax will classify the points raw the points $\mathbf{x}^{(1)}$, $\mathbf{x}^{(4)}$ and $\mathbf{x}^{(7)}$.

REMINDER – Softmax

$$p(C_k|\mathbf{x}) = \frac{\exp\left(\mathbf{w}_k^{\mathrm{T}}\mathbf{x}\right)}{\sum_{j=1}^{K}\exp\left(\mathbf{w}_j^{\mathrm{T}}\mathbf{x}\right)} \quad \forall k = 0,1,2$$

Efficient way of computing softmax output

$$\left.\begin{matrix}\mathbf{w}_0^{\mathrm{T}}\mathbf{x}\\[1em]\mathbf{w}_1^{\mathrm{T}}\mathbf{x}\\[1em]\mathbf{w}_2^{\mathrm{T}}\mathbf{x}\end{matrix}\right\} \quad \begin{pmatrix}\mathbf{w}_0^{\mathrm{T}}\mathbf{x}\\\mathbf{w}_1^{\mathrm{T}}\mathbf{x}\\\mathbf{w}_2^{\mathrm{T}}\mathbf{x}\end{pmatrix} = \overbrace{\begin{pmatrix}\cdots & \mathbf{w}_0^{\mathrm{T}} & \cdots\\\cdots & \mathbf{w}_1^{\mathrm{T}} & \cdots\\\cdots & \mathbf{w}_2^{\mathrm{T}} & \cdots\end{pmatrix}}^{1} \cdot \mathbf{x} \longrightarrow \overset{4}{\frac{\overset{2}{\exp}\begin{pmatrix}\mathbf{w}_0^{\mathrm{T}}\mathbf{x}\\\mathbf{w}_1^{\mathrm{T}}\mathbf{x}\\\mathbf{w}_2^{\mathrm{T}}\mathbf{x}\end{pmatrix}}{\underset{3}{\text{sum}}\left[\exp\begin{pmatrix}\mathbf{w}_0^{\mathrm{T}}\mathbf{x}\\\mathbf{w}_1^{\mathrm{T}}\mathbf{x}\\\mathbf{w}_2^{\mathrm{T}}\mathbf{x}\end{pmatrix}\right]}} = \begin{pmatrix}p(C_0|\mathbf{x})\\p(C_1|\mathbf{x})\\p(C_2|\mathbf{x})\end{pmatrix}$$

EXERCICE 3

Class 0 C_0: $\left\{\mathbf{x}^{(1)} = \begin{pmatrix}0.5\\0.4\end{pmatrix}, \mathbf{x}^{(2)} = \begin{pmatrix}0.8\\0.3\end{pmatrix}, \mathbf{x}^{(3)} = \begin{pmatrix}0.3\\0.8\end{pmatrix}\right\}$,

Class 1 C_1: $\left\{\mathbf{x}^{(4)} = \begin{pmatrix}-0.4\\0.3\end{pmatrix}, \mathbf{x}^{(5)} = \begin{pmatrix}-0.3\\0.7\end{pmatrix}, \mathbf{x}^{(6)} = \begin{pmatrix}0.7\\0.2\end{pmatrix}\right\}$,

Class 2 C_2: $\left\{\mathbf{x}^{(7)} = \begin{pmatrix}0.7\\-0.4\end{pmatrix}, \mathbf{x}^{(8)} = \begin{pmatrix}0.5\\-0.6\end{pmatrix}, \mathbf{x}^{(9)} = \begin{pmatrix}-0.4\\-0.5\end{pmatrix}\right\}$

$$\mathbf{w}_0 = \begin{pmatrix}-0.3\\0.87\\1.47\end{pmatrix}, \mathbf{w}_1 = \begin{pmatrix}-0.01\\0.58\\1.02\end{pmatrix}$$

$$\mathbf{w}_2 = \begin{pmatrix}0.43\\-1.90\\0.33\end{pmatrix}$$

(b) Determine at which class the calculated softmax will classify the points raw the points $\mathbf{x}^{(1)}$, $\mathbf{x}^{(4)}$ and $\mathbf{x}^{(7)}$.

1

$$\begin{pmatrix}\mathbf{w}_0^T\mathbf{x}^{(1)}\\ \mathbf{w}_1^T\mathbf{x}^{(1)}\\ \mathbf{w}_2^T\mathbf{x}^{(1)}\end{pmatrix} = \begin{pmatrix}\cdots & \mathbf{w}_0^T & \cdots\\ \cdots & \mathbf{w}_1^T & \cdots\\ \cdots & \mathbf{w}_2^T & \cdots\end{pmatrix} \cdot \mathbf{x}^{(1)} = ?$$

EXERCICE 3

Class 0 C_0: $\left\{\mathbf{x}^{(1)} = \begin{pmatrix}0.5\\0.4\end{pmatrix}, \mathbf{x}^{(2)} = \begin{pmatrix}0.8\\0.3\end{pmatrix}, \mathbf{x}^{(3)} = \begin{pmatrix}0.3\\0.8\end{pmatrix}\right\}$,

Class 1 C_1: $\left\{\mathbf{x}^{(4)} = \begin{pmatrix}-0.4\\0.3\end{pmatrix}, \mathbf{x}^{(5)} = \begin{pmatrix}-0.3\\0.7\end{pmatrix}, \mathbf{x}^{(6)} = \begin{pmatrix}0.7\\0.2\end{pmatrix}\right\}$,

Class 2 C_2: $\left\{\mathbf{x}^{(7)} = \begin{pmatrix}0.7\\-0.4\end{pmatrix}, \mathbf{x}^{(8)} = \begin{pmatrix}0.5\\-0.6\end{pmatrix}, \mathbf{x}^{(9)} = \begin{pmatrix}-0.4\\-0.5\end{pmatrix}\right\}$

$$\mathbf{w}_0 = \begin{pmatrix}-0.3\\0.87\\1.47\end{pmatrix}, \mathbf{w}_1 = \begin{pmatrix}-0.01\\0.58\\1.02\end{pmatrix}$$
$$\mathbf{w}_2 = \begin{pmatrix}0.43\\-1.90\\0.33\end{pmatrix}$$

(b) Determine at which class the calculated softmax will classify the points raw the points $\mathbf{x}^{(1)}$, $\mathbf{x}^{(4)}$ and $\mathbf{x}^{(7)}$.

1

Bias term: $\mathbf{w}^{\mathrm{T}}\mathbf{x} = w_0 + w_1 x_1 + w_2 x_2$

$$\begin{pmatrix}\mathbf{w}_0^{\mathrm{T}}\mathbf{x}^{(1)}\\ \mathbf{w}_1^{\mathrm{T}}\mathbf{x}^{(1)}\\ \mathbf{w}_2^{\mathrm{T}}\mathbf{x}^{(1)}\end{pmatrix} = \begin{pmatrix}\cdots & \mathbf{w}_0^{\mathrm{T}} & \cdots\\ \cdots & \mathbf{w}_1^{\mathrm{T}} & \cdots\\ \cdots & \mathbf{w}_2^{\mathrm{T}} & \cdots\end{pmatrix} \cdot \mathbf{x}^{(1)} = \begin{pmatrix}-0.3 & 0.87 & 1.47\\ -0.01 & 0.58 & 1.02\\ 0.43 & -1.90 & 0.33\end{pmatrix} \cdot \begin{pmatrix}1\\0.5\\0.4\end{pmatrix} = \begin{pmatrix}0.72\\0.69\\-0.39\end{pmatrix}$$

EXERCICE 3

Class 0 C_0: $\left\{\mathbf{x}^{(1)} = \begin{pmatrix}0.5\\0.4\end{pmatrix}, \mathbf{x}^{(2)} = \begin{pmatrix}0.8\\0.3\end{pmatrix}, \mathbf{x}^{(3)} = \begin{pmatrix}0.3\\0.8\end{pmatrix}\right\}$,

Class 1 C_1: $\left\{\mathbf{x}^{(4)} = \begin{pmatrix}-0.4\\0.3\end{pmatrix}, \mathbf{x}^{(5)} = \begin{pmatrix}-0.3\\0.7\end{pmatrix}, \mathbf{x}^{(6)} = \begin{pmatrix}0.7\\0.2\end{pmatrix}\right\}$,

Class 2 C_2: $\left\{\mathbf{x}^{(7)} = \begin{pmatrix}0.7\\-0.4\end{pmatrix}, \mathbf{x}^{(8)} = \begin{pmatrix}0.5\\-0.6\end{pmatrix}, \mathbf{x}^{(9)} = \begin{pmatrix}-0.4\\-0.5\end{pmatrix}\right\}$

$$\mathbf{w}_0 = \begin{pmatrix}-0.3\\0.87\\1.47\end{pmatrix}, \mathbf{w}_1 = \begin{pmatrix}-0.01\\0.58\\1.02\end{pmatrix}$$

$$\mathbf{w}_2 = \begin{pmatrix}0.43\\-1.90\\0.33\end{pmatrix}$$

(b) Determine at which class the calculated softmax will classify the points raw the points $\mathbf{x}^{(1)}$, $\mathbf{x}^{(4)}$ and $\mathbf{x}^{(7)}$.

2

$$\begin{pmatrix}\mathbf{w}_0^{\mathrm{T}}\mathbf{x}^{(1)}\\\mathbf{w}_1^{\mathrm{T}}\mathbf{x}^{(1)}\\\mathbf{w}_2^{\mathrm{T}}\mathbf{x}^{(1)}\end{pmatrix} = \begin{pmatrix}0.72\\0.69\\-0.39\end{pmatrix}$$

$$\exp\begin{pmatrix}\mathbf{w}_0^{\mathrm{T}}\mathbf{x}^{(1)}\\\mathbf{w}_1^{\mathrm{T}}\mathbf{x}^{(1)}\\\mathbf{w}_2^{\mathrm{T}}\mathbf{x}^{(1)}\end{pmatrix} =?$$

EXERCICE 3

Class 0 C_0: $\left\{\mathbf{x}^{(1)} = \begin{pmatrix}0.5\\0.4\end{pmatrix}, \mathbf{x}^{(2)} = \begin{pmatrix}0.8\\0.3\end{pmatrix}, \mathbf{x}^{(3)} = \begin{pmatrix}0.3\\0.8\end{pmatrix}\right\}$,

Class 1 C_1: $\left\{\mathbf{x}^{(4)} = \begin{pmatrix}-0.4\\0.3\end{pmatrix}, \mathbf{x}^{(5)} = \begin{pmatrix}-0.3\\0.7\end{pmatrix}, \mathbf{x}^{(6)} = \begin{pmatrix}0.7\\0.2\end{pmatrix}\right\}$,

Class 2 C_2: $\left\{\mathbf{x}^{(7)} = \begin{pmatrix}0.7\\-0.4\end{pmatrix}, \mathbf{x}^{(8)} = \begin{pmatrix}0.5\\-0.6\end{pmatrix}, \mathbf{x}^{(9)} = \begin{pmatrix}-0.4\\-0.5\end{pmatrix}\right\}$

$$\mathbf{w}_0 = \begin{pmatrix}-0.3\\0.87\\1.47\end{pmatrix}, \mathbf{w}_1 = \begin{pmatrix}-0.01\\0.58\\1.02\end{pmatrix}$$

$$\mathbf{w}_2 = \begin{pmatrix}0.43\\-1.90\\0.33\end{pmatrix}$$

(b) Determine at which class the calculated softmax will classify the points raw the points $\mathbf{x}^{(1)}$, $\mathbf{x}^{(4)}$ and $\mathbf{x}^{(7)}$.

2

$$\begin{pmatrix}\mathbf{w}_0^{\mathrm{T}}\mathbf{x}^{(1)}\\\mathbf{w}_1^{\mathrm{T}}\mathbf{x}^{(1)}\\\mathbf{w}_2^{\mathrm{T}}\mathbf{x}^{(1)}\end{pmatrix} = \begin{pmatrix}0.72\\0.69\\-0.39\end{pmatrix}$$

$$\exp\begin{pmatrix}\mathbf{w}_0^{\mathrm{T}}\mathbf{x}^{(1)}\\\mathbf{w}_1^{\mathrm{T}}\mathbf{x}^{(1)}\\\mathbf{w}_2^{\mathrm{T}}\mathbf{x}^{(1)}\end{pmatrix} = \begin{pmatrix}2.06\\1.99\\0.68\end{pmatrix}$$

EXERCICE 3

Class 0 C_0: $\left\{\mathbf{x}^{(1)} = \begin{pmatrix}0.5\\0.4\end{pmatrix}, \mathbf{x}^{(2)} = \begin{pmatrix}0.8\\0.3\end{pmatrix}, \mathbf{x}^{(3)} = \begin{pmatrix}0.3\\0.8\end{pmatrix}\right\}$,

Class 1 C_1: $\left\{\mathbf{x}^{(4)} = \begin{pmatrix}-0.4\\0.3\end{pmatrix}, \mathbf{x}^{(5)} = \begin{pmatrix}-0.3\\0.7\end{pmatrix}, \mathbf{x}^{(6)} = \begin{pmatrix}0.7\\0.2\end{pmatrix}\right\}$,

Class 2 C_2: $\left\{\mathbf{x}^{(7)} = \begin{pmatrix}0.7\\-0.4\end{pmatrix}, \mathbf{x}^{(8)} = \begin{pmatrix}0.5\\-0.6\end{pmatrix}, \mathbf{x}^{(9)} = \begin{pmatrix}-0.4\\-0.5\end{pmatrix}\right\}$

$$\mathbf{w}_0 = \begin{pmatrix}-0.3\\0.87\\1.47\end{pmatrix}, \mathbf{w}_1 = \begin{pmatrix}-0.01\\0.58\\1.02\end{pmatrix}$$

$$\mathbf{w}_2 = \begin{pmatrix}0.43\\-1.90\\0.33\end{pmatrix}$$

(b) Determine at which class the calculated softmax will classify the points raw the points $\mathbf{x}^{(1)}$, $\mathbf{x}^{(4)}$ and $\mathbf{x}^{(7)}$.

3

$$\exp\begin{pmatrix}\mathbf{w}_0^{\mathrm{T}}\mathbf{x}^{(1)}\\\mathbf{w}_1^{\mathrm{T}}\mathbf{x}^{(1)}\\\mathbf{w}_2^{\mathrm{T}}\mathbf{x}^{(1)}\end{pmatrix} = \begin{pmatrix}2.06\\1.99\\0.68\end{pmatrix}$$

$$\mathrm{sum}\left[\exp\begin{pmatrix}\mathbf{w}_0^{\mathrm{T}}\mathbf{x}^{(1)}\\\mathbf{w}_1^{\mathrm{T}}\mathbf{x}^{(1)}\\\mathbf{w}_2^{\mathrm{T}}\mathbf{x}^{(1)}\end{pmatrix}\right] = ?$$

EXERCICE 3

Class 0 C_0: $\left\{\mathbf{x}^{(1)} = \begin{pmatrix}0.5\\0.4\end{pmatrix}, \mathbf{x}^{(2)} = \begin{pmatrix}0.8\\0.3\end{pmatrix}, \mathbf{x}^{(3)} = \begin{pmatrix}0.3\\0.8\end{pmatrix}\right\}$,

Class 1 C_1: $\left\{\mathbf{x}^{(4)} = \begin{pmatrix}-0.4\\0.3\end{pmatrix}, \mathbf{x}^{(5)} = \begin{pmatrix}-0.3\\0.7\end{pmatrix}, \mathbf{x}^{(6)} = \begin{pmatrix}0.7\\0.2\end{pmatrix}\right\}$,

Class 2 C_2: $\left\{\mathbf{x}^{(7)} = \begin{pmatrix}0.7\\-0.4\end{pmatrix}, \mathbf{x}^{(8)} = \begin{pmatrix}0.5\\-0.6\end{pmatrix}, \mathbf{x}^{(9)} = \begin{pmatrix}-0.4\\-0.5\end{pmatrix}\right\}$

$$\mathbf{w}_0 = \begin{pmatrix}-0.3\\0.87\\1.47\end{pmatrix}, \mathbf{w}_1 = \begin{pmatrix}-0.01\\0.58\\1.02\end{pmatrix}$$

$$\mathbf{w}_2 = \begin{pmatrix}0.43\\-1.90\\0.33\end{pmatrix}$$

(b) Determine at which class the calculated softmax will classify the points raw the points $\mathbf{x}^{(1)}$, $\mathbf{x}^{(4)}$ and $\mathbf{x}^{(7)}$.

3

$$\exp\begin{pmatrix}\mathbf{w}_0^{\mathrm{T}}\mathbf{x}^{(1)}\\\mathbf{w}_1^{\mathrm{T}}\mathbf{x}^{(1)}\\\mathbf{w}_2^{\mathrm{T}}\mathbf{x}^{(1)}\end{pmatrix} = \begin{pmatrix}2.06\\1.99\\0.68\end{pmatrix}$$

$$\text{sum}\left[\exp\begin{pmatrix}\mathbf{w}_0^{\mathrm{T}}\mathbf{x}^{(1)}\\\mathbf{w}_1^{\mathrm{T}}\mathbf{x}^{(1)}\\\mathbf{w}_2^{\mathrm{T}}\mathbf{x}^{(1)}\end{pmatrix}\right] = 2.06 + 1.99 + 0.68 = 4.73$$

EXERCICE 3

Class 0 C_0: $\left\{\mathbf{x}^{(1)} = \begin{pmatrix}0.5\\0.4\end{pmatrix}, \mathbf{x}^{(2)} = \begin{pmatrix}0.8\\0.3\end{pmatrix}, \mathbf{x}^{(3)} = \begin{pmatrix}0.3\\0.8\end{pmatrix}\right\}$,

Class 1 C_1: $\left\{\mathbf{x}^{(4)} = \begin{pmatrix}-0.4\\0.3\end{pmatrix}, \mathbf{x}^{(5)} = \begin{pmatrix}-0.3\\0.7\end{pmatrix}, \mathbf{x}^{(6)} = \begin{pmatrix}0.7\\0.2\end{pmatrix}\right\}$,

Class 2 C_2: $\left\{\mathbf{x}^{(7)} = \begin{pmatrix}0.7\\-0.4\end{pmatrix}, \mathbf{x}^{(8)} = \begin{pmatrix}0.5\\-0.6\end{pmatrix}, \mathbf{x}^{(9)} = \begin{pmatrix}-0.4\\-0.5\end{pmatrix}\right\}$

$$\mathbf{w}_0 = \begin{pmatrix}-0.3\\0.87\\1.47\end{pmatrix}, \mathbf{w}_1 = \begin{pmatrix}-0.01\\0.58\\1.02\end{pmatrix}$$

$$\mathbf{w}_2 = \begin{pmatrix}0.43\\-1.90\\0.33\end{pmatrix}$$

(b) Determine at which class the calculated softmax will classify the points raw the points $\mathbf{x}^{(1)}$, $\mathbf{x}^{(4)}$ and $\mathbf{x}^{(7)}$.

4

$$\exp\begin{pmatrix}\mathbf{w}_0^{\mathrm{T}}\mathbf{x}^{(1)}\\ \mathbf{w}_1^{\mathrm{T}}\mathbf{x}^{(1)}\\ \mathbf{w}_2^{\mathrm{T}}\mathbf{x}^{(1)}\end{pmatrix} = \begin{pmatrix}2.06\\1.99\\0.68\end{pmatrix} \qquad \mathrm{sum}\left[\exp\begin{pmatrix}\mathbf{w}_0^{\mathrm{T}}\mathbf{x}^{(1)}\\ \mathbf{w}_1^{\mathrm{T}}\mathbf{x}^{(1)}\\ \mathbf{w}_2^{\mathrm{T}}\mathbf{x}^{(1)}\end{pmatrix}\right] = 4.73$$

$$\begin{pmatrix}p(C_0|\mathbf{x}^{(1)})\\ p(C_1|\mathbf{x}^{(1)})\\ p(C_2|\mathbf{x}^{(1)})\end{pmatrix} = \frac{\exp\begin{pmatrix}\mathbf{w}_0^{\mathrm{T}}\mathbf{x}^{(1)}\\ \mathbf{w}_1^{\mathrm{T}}\mathbf{x}^{(1)}\\ \mathbf{w}_2^{\mathrm{T}}\mathbf{x}^{(1)}\end{pmatrix}}{\mathrm{sum}\left[\exp\begin{pmatrix}\mathbf{w}_0^{\mathrm{T}}\mathbf{x}^{(1)}\\ \mathbf{w}_1^{\mathrm{T}}\mathbf{x}^{(1)}\\ \mathbf{w}_2^{\mathrm{T}}\mathbf{x}^{(1)}\end{pmatrix}\right]} = ?$$

EXERCICE 3

Class 0 C_0: $\left\{\mathbf{x}^{(1)} = \begin{pmatrix}0.5\\0.4\end{pmatrix}, \mathbf{x}^{(2)} = \begin{pmatrix}0.8\\0.3\end{pmatrix}, \mathbf{x}^{(3)} = \begin{pmatrix}0.3\\0.8\end{pmatrix}\right\}$,

Class 1 C_1: $\left\{\mathbf{x}^{(4)} = \begin{pmatrix}-0.4\\0.3\end{pmatrix}, \mathbf{x}^{(5)} = \begin{pmatrix}-0.3\\0.7\end{pmatrix}, \mathbf{x}^{(6)} = \begin{pmatrix}0.7\\0.2\end{pmatrix}\right\}$,

Class 2 C_2: $\left\{\mathbf{x}^{(7)} = \begin{pmatrix}0.7\\-0.4\end{pmatrix}, \mathbf{x}^{(8)} = \begin{pmatrix}0.5\\-0.6\end{pmatrix}, \mathbf{x}^{(9)} = \begin{pmatrix}-0.4\\-0.5\end{pmatrix}\right\}$

$$\mathbf{w}_0 = \begin{pmatrix}-0.3\\0.87\\1.47\end{pmatrix}, \mathbf{w}_1 = \begin{pmatrix}-0.01\\0.58\\1.02\end{pmatrix}$$

$$\mathbf{w}_2 = \begin{pmatrix}0.43\\-1.90\\0.33\end{pmatrix}$$

(b) Determine at which class the calculated softmax will classify the points raw the points $\mathbf{x}^{(1)}$, $\mathbf{x}^{(4)}$ and $\mathbf{x}^{(7)}$.

4

$$\exp\begin{pmatrix}\mathbf{w}_0^{\mathrm{T}}\mathbf{x}^{(1)}\\\mathbf{w}_1^{\mathrm{T}}\mathbf{x}^{(1)}\\\mathbf{w}_2^{\mathrm{T}}\mathbf{x}^{(1)}\end{pmatrix} = \begin{pmatrix}2.06\\1.99\\0.68\end{pmatrix} \qquad \text{sum}\left[\exp\begin{pmatrix}\mathbf{w}_0^{\mathrm{T}}\mathbf{x}^{(1)}\\\mathbf{w}_1^{\mathrm{T}}\mathbf{x}^{(1)}\\\mathbf{w}_2^{\mathrm{T}}\mathbf{x}^{(1)}\end{pmatrix}\right] = 4.73$$

$$\begin{pmatrix}p(C_0|\mathbf{x}^{(1)})\\p(C_1|\mathbf{x}^{(1)})\\p(C_2|\mathbf{x}^{(1)})\end{pmatrix} = \frac{\exp\begin{pmatrix}\mathbf{w}_0^{\mathrm{T}}\mathbf{x}^{(1)}\\\mathbf{w}_1^{\mathrm{T}}\mathbf{x}^{(1)}\\\mathbf{w}_2^{\mathrm{T}}\mathbf{x}^{(1)}\end{pmatrix}}{\text{sum}\left[\exp\begin{pmatrix}\mathbf{w}_0^{\mathrm{T}}\mathbf{x}^{(1)}\\\mathbf{w}_1^{\mathrm{T}}\mathbf{x}^{(1)}\\\mathbf{w}_2^{\mathrm{T}}\mathbf{x}^{(1)}\end{pmatrix}\right]} = \begin{pmatrix}\frac{2.06}{4.73}\\\frac{1.99}{4.73}\\\frac{0.68}{4.73}\end{pmatrix} = \begin{pmatrix}0.44\\0.42\\0.14\end{pmatrix}$$

The largest probability is for class 0, then, this classifier would assign $\mathbf{x}^{(1)}$ to class 0.

EXERCICE 3

Class 0 C_0: $\left\{\mathbf{x}^{(1)} = \begin{pmatrix}0.5\\0.4\end{pmatrix}, \mathbf{x}^{(2)} = \begin{pmatrix}0.8\\0.3\end{pmatrix}, \mathbf{x}^{(3)} = \begin{pmatrix}0.3\\0.8\end{pmatrix}\right\}$,

Class 1 C_1: $\left\{\mathbf{x}^{(4)} = \begin{pmatrix}-0.4\\0.3\end{pmatrix}, \mathbf{x}^{(5)} = \begin{pmatrix}-0.3\\0.7\end{pmatrix}, \mathbf{x}^{(6)} = \begin{pmatrix}0.7\\0.2\end{pmatrix}\right\}$,

Class 2 C_2: $\left\{\mathbf{x}^{(7)} = \begin{pmatrix}0.7\\-0.4\end{pmatrix}, \mathbf{x}^{(8)} = \begin{pmatrix}0.5\\-0.6\end{pmatrix}, \mathbf{x}^{(9)} = \begin{pmatrix}-0.4\\-0.5\end{pmatrix}\right\}$

$$\mathbf{w}_0 = \begin{pmatrix}-0.3\\0.87\\1.47\end{pmatrix}, \mathbf{w}_1 = \begin{pmatrix}-0.01\\0.58\\1.02\end{pmatrix}$$

$$\mathbf{w}_2 = \begin{pmatrix}0.43\\-1.90\\0.33\end{pmatrix}$$

(b) Determine at which class the calculated softmax will classify the points raw the points $\mathbf{x}^{(1)}$, $\mathbf{x}^{(4)}$ and $\mathbf{x}^{(7)}$.

Repeat for $\mathbf{x}^{(4)}$ and $\mathbf{x}^{(7)}$...

$$\begin{pmatrix}p(C_0|\mathbf{x}^{(4)})\\p(C_1|\mathbf{x}^{(4)})\\p(C_2|\mathbf{x}^{(4)})\end{pmatrix} = \begin{pmatrix}0.15\\0.19\\0.66\end{pmatrix}$$

$\mathbf{x}^{(4)}$ to class 2

$$\begin{pmatrix}p(C_0|\mathbf{x}^{(7)})\\p(C_1|\mathbf{x}^{(7)})\\p(C_2|\mathbf{x}^{(7)})\end{pmatrix} = \begin{pmatrix}0.36\\0.47\\0.17\end{pmatrix}$$

$\mathbf{x}^{(7)}$ to class 1

EXERCICE 3

Class 0 C_0: $\left\{\mathbf{x}^{(1)} = \begin{pmatrix}0.5\\0.4\end{pmatrix}, \mathbf{x}^{(2)} = \begin{pmatrix}0.8\\0.3\end{pmatrix}, \mathbf{x}^{(3)} = \begin{pmatrix}0.3\\0.8\end{pmatrix}\right\}$,

Class 1 C_1: $\left\{\mathbf{x}^{(4)} = \begin{pmatrix}-0.4\\0.3\end{pmatrix}, \mathbf{x}^{(5)} = \begin{pmatrix}-0.3\\0.7\end{pmatrix}, \mathbf{x}^{(6)} = \begin{pmatrix}0.7\\0.2\end{pmatrix}\right\}$,

Class 2 C_2: $\left\{\mathbf{x}^{(7)} = \begin{pmatrix}0.7\\-0.4\end{pmatrix}, \mathbf{x}^{(8)} = \begin{pmatrix}0.5\\-0.6\end{pmatrix}, \mathbf{x}^{(9)} = \begin{pmatrix}-0.4\\-0.5\end{pmatrix}\right\}$

$$\mathbf{w}_0 = \begin{pmatrix}-0.3\\0.87\\1.47\end{pmatrix}, \mathbf{w}_1 = \begin{pmatrix}-0.01\\0.58\\1.02\end{pmatrix}$$

$$\mathbf{w}_2 = \begin{pmatrix}0.43\\-1.90\\0.33\end{pmatrix}$$

(c) Calculate and plot the discriminant surfaces.

REMINDER – Discriminant surfaces (DS)

DS between class j and class k $\left(\mathrm{DS}_{jk}\right)$:

$$\mathrm{DS}_{jk} = \{\mathbf{x} \mid g_j(\mathbf{x}) = g_k(\mathbf{x})\} \qquad \text{where} \qquad g_k(\mathbf{x}) = \mathbf{w}_k^{\mathrm{T}}\mathbf{x} \qquad \text{(analogously for } j\text{)}$$

The set of points $\mathbf{x}$ that satisfy the condition $g_j(\mathbf{x}) = g_k(\mathbf{x})$.

EXERCICE 3

Class 0 C_0: $\left\{\mathbf{x}^{(1)} = \begin{pmatrix}0.5\\0.4\end{pmatrix}, \mathbf{x}^{(2)} = \begin{pmatrix}0.8\\0.3\end{pmatrix}, \mathbf{x}^{(3)} = \begin{pmatrix}0.3\\0.8\end{pmatrix}\right\}$,

Class 1 C_1: $\left\{\mathbf{x}^{(4)} = \begin{pmatrix}-0.4\\0.3\end{pmatrix}, \mathbf{x}^{(5)} = \begin{pmatrix}-0.3\\0.7\end{pmatrix}, \mathbf{x}^{(6)} = \begin{pmatrix}0.7\\0.2\end{pmatrix}\right\}$,

Class 2 C_2: $\left\{\mathbf{x}^{(7)} = \begin{pmatrix}0.7\\-0.4\end{pmatrix}, \mathbf{x}^{(8)} = \begin{pmatrix}0.5\\-0.6\end{pmatrix}, \mathbf{x}^{(9)} = \begin{pmatrix}-0.4\\-0.5\end{pmatrix}\right\}$

$$\mathbf{w}_0 = \begin{pmatrix}-0.3\\0.87\\1.47\end{pmatrix}, \mathbf{w}_1 = \begin{pmatrix}-0.01\\0.58\\1.02\end{pmatrix}$$

$$\mathbf{w}_2 = \begin{pmatrix}0.43\\-1.90\\0.33\end{pmatrix}$$

(c) Calculate and plot the discriminant surfaces.

$$g_k(\mathbf{x}) = \mathbf{w}_k^{\mathrm{T}}\mathbf{x}$$

$$\mathrm{DS}_{01} = \{\mathbf{x} \mid g_0(\mathbf{x}) = g_1(\mathbf{x})\} = \{\mathbf{x} \mid ?\}$$

$$\mathrm{DS}_{02} = \{\mathbf{x} \mid g_0(\mathbf{x}) = g_2(\mathbf{x})\} = \{\mathbf{x} \mid ?\}$$

$$\mathrm{DS}_{12} = \{\mathbf{x} \mid g_1(\mathbf{x}) = g_2(\mathbf{x})\} = \{\mathbf{x} \mid ?\}$$

EXERCICE 3

Class 0 C_0: $\left\{\mathbf{x}^{(1)} = \begin{pmatrix}0.5\\0.4\end{pmatrix}, \mathbf{x}^{(2)} = \begin{pmatrix}0.8\\0.3\end{pmatrix}, \mathbf{x}^{(3)} = \begin{pmatrix}0.3\\0.8\end{pmatrix}\right\}$,

Class 1 C_1: $\left\{\mathbf{x}^{(4)} = \begin{pmatrix}-0.4\\0.3\end{pmatrix}, \mathbf{x}^{(5)} = \begin{pmatrix}-0.3\\0.7\end{pmatrix}, \mathbf{x}^{(6)} = \begin{pmatrix}0.7\\0.2\end{pmatrix}\right\}$,

Class 2 C_2: $\left\{\mathbf{x}^{(7)} = \begin{pmatrix}0.7\\-0.4\end{pmatrix}, \mathbf{x}^{(8)} = \begin{pmatrix}0.5\\-0.6\end{pmatrix}, \mathbf{x}^{(9)} = \begin{pmatrix}-0.4\\-0.5\end{pmatrix}\right\}$

$$\mathbf{w}_0 = \begin{pmatrix}-0.3\\0.87\\1.47\end{pmatrix}, \mathbf{w}_1 = \begin{pmatrix}-0.01\\0.58\\1.02\end{pmatrix}$$

$$\mathbf{w}_2 = \begin{pmatrix}0.43\\-1.90\\0.33\end{pmatrix}$$

(c) Calculate and plot the discriminant surfaces.

$$g_k(\mathbf{x}) = \mathbf{w}_k^{\mathrm{T}}\mathbf{x}$$

$$\mathrm{DS}_{01} = \{\mathbf{x} \mid g_0(\mathbf{x}) = g_1(\mathbf{x})\} = \{\mathbf{x} \mid -0.3 + 0.87x_1 + 1.47x_2 = -0.01 + 0.58x_1 + 1.02x_2\}$$
$$= \{\mathbf{x} \mid -0.29 + 0.29x_1 + 0.45x_2 = 0\}$$

$\mathrm{DS}_{01} = \{g_0(x) - g_1(x) = 0\}$
Then
if $\mathrm{DS}_{01} > 0 \Longrightarrow g_0 > g_1$
if $\mathrm{DS}_{01} < 0 \Longrightarrow g_0 < g_1$

$$\mathrm{DS}_{02} = \{\mathbf{x} \mid g_0(\mathbf{x}) = g_2(\mathbf{x})\} = \{\mathbf{x} \mid -0.73 + 2.77x_1 + 1.14x_2 = 0\}$$

$$\mathrm{DS}_{02} = \{g_0(x) - g_2(x) = 0\}$$

$$\mathrm{DS}_{12} = \{\mathbf{x} \mid g_1(\mathbf{x}) = g_2(\mathbf{x})\} = \{\mathbf{x} \mid -0.44 + 2.48x_1 + 0.69x_2 = 0\}$$

$$\mathrm{DS}_{12} = \{g_1(x) - g_2(x) = 0\}$$

EXERCICE 3

Class 0 C_0: $\left\{\mathbf{x}^{(1)} = \begin{pmatrix}0.5\\0.4\end{pmatrix}, \mathbf{x}^{(2)} = \begin{pmatrix}0.8\\0.3\end{pmatrix}, \mathbf{x}^{(3)} = \begin{pmatrix}0.3\\0.8\end{pmatrix}\right\}$,

Class 1 C_1: $\left\{\mathbf{x}^{(4)} = \begin{pmatrix}-0.4\\0.3\end{pmatrix}, \mathbf{x}^{(5)} = \begin{pmatrix}-0.3\\0.7\end{pmatrix}, \mathbf{x}^{(6)} = \begin{pmatrix}0.7\\0.2\end{pmatrix}\right\}$,

Class 2 C_2: $\left\{\mathbf{x}^{(7)} = \begin{pmatrix}0.7\\-0.4\end{pmatrix}, \mathbf{x}^{(8)} = \begin{pmatrix}0.5\\-0.6\end{pmatrix}, \mathbf{x}^{(9)} = \begin{pmatrix}-0.4\\-0.5\end{pmatrix}\right\}$

$$\mathbf{w}_0 = \begin{pmatrix}-0.3\\0.87\\1.47\end{pmatrix}, \mathbf{w}_1 = \begin{pmatrix}-0.01\\0.58\\1.02\end{pmatrix}$$

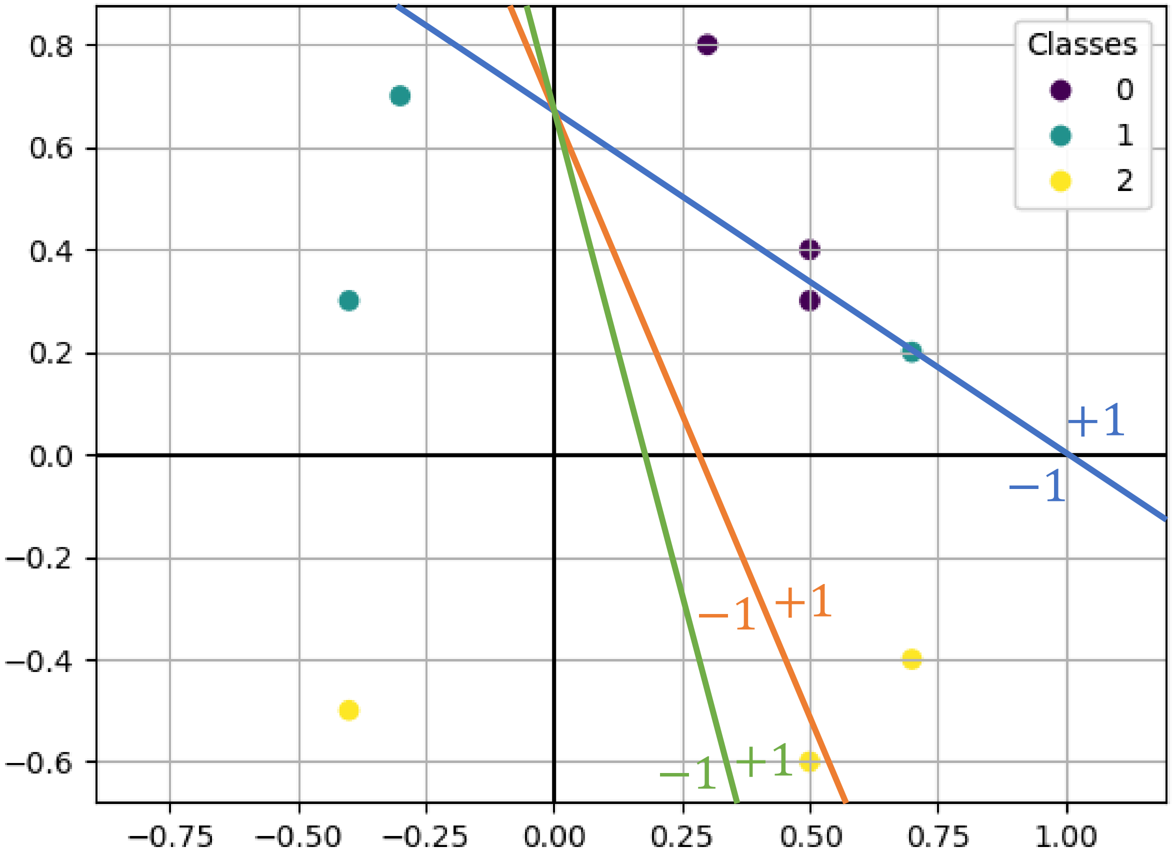
$$\mathbf{w}_2 = \begin{pmatrix}0.43\\-1.90\\0.33\end{pmatrix}$$

(c) Calculate and plot the discriminant surfaces.

$$\mathrm{DS}_{01} = \{\mathbf{x} \mid -0.29 + 0.29x_1 + 0.45x_2 = 0\}$$

$$\mathrm{DS}_{02} = \{\mathbf{x} \mid -0.73 + 2.77x_1 + 1.14x_2 = 0\}$$

$$\mathrm{DS}_{12} = \{\mathbf{x} \mid -0.44 + 2.48x_1 + 0.69x_2 = 0\}$$



EXERCICE 3

Class 0 C_0: $\left\{\mathbf{x}^{(1)} = \begin{pmatrix}0.5\\0.4\end{pmatrix}, \mathbf{x}^{(2)} = \begin{pmatrix}0.8\\0.3\end{pmatrix}, \mathbf{x}^{(3)} = \begin{pmatrix}0.3\\0.8\end{pmatrix}\right\}$,

Class 1 C_1: $\left\{\mathbf{x}^{(4)} = \begin{pmatrix}-0.4\\0.3\end{pmatrix}, \mathbf{x}^{(5)} = \begin{pmatrix}-0.3\\0.7\end{pmatrix}, \mathbf{x}^{(6)} = \begin{pmatrix}0.7\\0.2\end{pmatrix}\right\}$,

Class 2 C_2: $\left\{\mathbf{x}^{(7)} = \begin{pmatrix}0.7\\-0.4\end{pmatrix}, \mathbf{x}^{(8)} = \begin{pmatrix}0.5\\-0.6\end{pmatrix}, \mathbf{x}^{(9)} = \begin{pmatrix}-0.4\\-0.5\end{pmatrix}\right\}$

$$\mathbf{w}_0 = \begin{pmatrix}-0.3\\0.87\\1.47\end{pmatrix}, \mathbf{w}_1 = \begin{pmatrix}-0.01\\0.58\\1.02\end{pmatrix}$$

$$\mathbf{w}_2 = \begin{pmatrix}0.43\\-1.90\\0.33\end{pmatrix}$$

(d) Calculate the error for the given solutions.

REMINDER – CROSS-ENTROPY

$$\mathbb{E}(\mathbf{w}_0, \mathbf{w}_1, \mathbf{w}_2) = -\sum_{n=1}^{N}\sum_{k=1}^{K} y_k^{(n)} \cdot \ln h_k^{(n)}$$

where

$$y_k^{(n)} \begin{cases} = 1 & if\ \mathbf{x}^{(n)} \text{ belongs to class } k \\ = 0 & \text{otherwise} \end{cases}$$

$$h_k^{(n)} = p\left(C_k \middle| \mathbf{x}^{(n)}\right)$$

Already computed in exercise b

EXERCICE 3

Class 0 C_0: $\left\{\mathbf{x}^{(1)} = \begin{pmatrix}0.5\\0.4\end{pmatrix}, \mathbf{x}^{(2)} = \begin{pmatrix}0.8\\0.3\end{pmatrix}, \mathbf{x}^{(3)} = \begin{pmatrix}0.3\\0.8\end{pmatrix}\right\}$,

Class 1 C_1: $\left\{\mathbf{x}^{(4)} = \begin{pmatrix}-0.4\\0.3\end{pmatrix}, \mathbf{x}^{(5)} = \begin{pmatrix}-0.3\\0.7\end{pmatrix}, \mathbf{x}^{(6)} = \begin{pmatrix}0.7\\0.2\end{pmatrix}\right\}$,

Class 2 C_2: $\left\{\mathbf{x}^{(7)} = \begin{pmatrix}0.7\\-0.4\end{pmatrix}, \mathbf{x}^{(8)} = \begin{pmatrix}0.5\\-0.6\end{pmatrix}, \mathbf{x}^{(9)} = \begin{pmatrix}-0.4\\-0.5\end{pmatrix}\right\}$

$$\mathbf{w}_0 = \begin{pmatrix}-0.3\\0.87\\1.47\end{pmatrix}, \mathbf{w}_1 = \begin{pmatrix}-0.01\\0.58\\1.02\end{pmatrix}$$

$$\mathbf{w}_2 = \begin{pmatrix}0.43\\-1.90\\0.33\end{pmatrix}$$

(d) Calculate the error for the given solutions.

REMINDER – CROSS-ENTROPY

$$\mathbb{E}(\mathbf{w}_0, \mathbf{w}_1, \mathbf{w}_2) = -\sum_{n=1}^{N}\sum_{k=1}^{K} y_k^{(n)} \cdot \ln h_k^{(n)} = -\sum_{n=1}^{N} \mathbf{y}^{(n)^{\mathrm{T}}} \cdot \ln \mathbf{h}^{(n)} = -\begin{pmatrix}1 & 0 & 0\end{pmatrix}\begin{pmatrix}\ln p\left(C_0 \middle| \mathbf{x}^{(1)}\right)\\ \ln p\left(C_1 \middle| \mathbf{x}^{(1)}\right)\\ \ln p\left(C_2 \middle| \mathbf{x}^{(1)}\right)\end{pmatrix} - \cdots = 9.55$$