Classification

CE/CZ4042 – Tutorial 3

1. Training inputs for a dichotomizer are given as:

$$\mathbf{X}_1 = \begin{bmatrix} 5 \\ 1 \end{bmatrix}; \quad \mathbf{X}_2 = \begin{bmatrix} 7 \\ 3 \end{bmatrix}; \quad \mathbf{X}_3 = \begin{bmatrix} 3 \\ 2 \end{bmatrix}; \quad \mathbf{X}_4 = \begin{bmatrix} 5 \\ 4 \end{bmatrix}; \qquad \textit{Class } 1$$

$$\mathbf{X}_5 = \begin{bmatrix} 0 \\ 0 \end{bmatrix}; \quad \mathbf{X}_6 = \begin{bmatrix} -1 \\ -3 \end{bmatrix}; \quad \mathbf{X}_7 = \begin{bmatrix} -2 \\ 3 \end{bmatrix}; \quad \mathbf{X}_8 = \begin{bmatrix} -3 \\ 0 \end{bmatrix}; \qquad \textit{Class } 2$$

(a) Determine whether the two classes of patterns are linearly separable and find the center of gravity of patterns in each class.

(b) If the center points of two clusters of class 1 and class 2 are vectors χ_1 and χ_2, show that linear decision boundary that perpendicularly passes through the middle point of the line joining the two centroids and can be expressed in the form

$$(\chi_1 - \chi_2)^T \mathbf{x} + \frac{1}{2}\left(\|\chi_2\|^2 - \|\chi_1\|^2\right) = 0$$

(c) Design a dichotomizer using a perceptron having the decision boundary as in part (b) for the given classification and determine how it recognizes the following input patterns

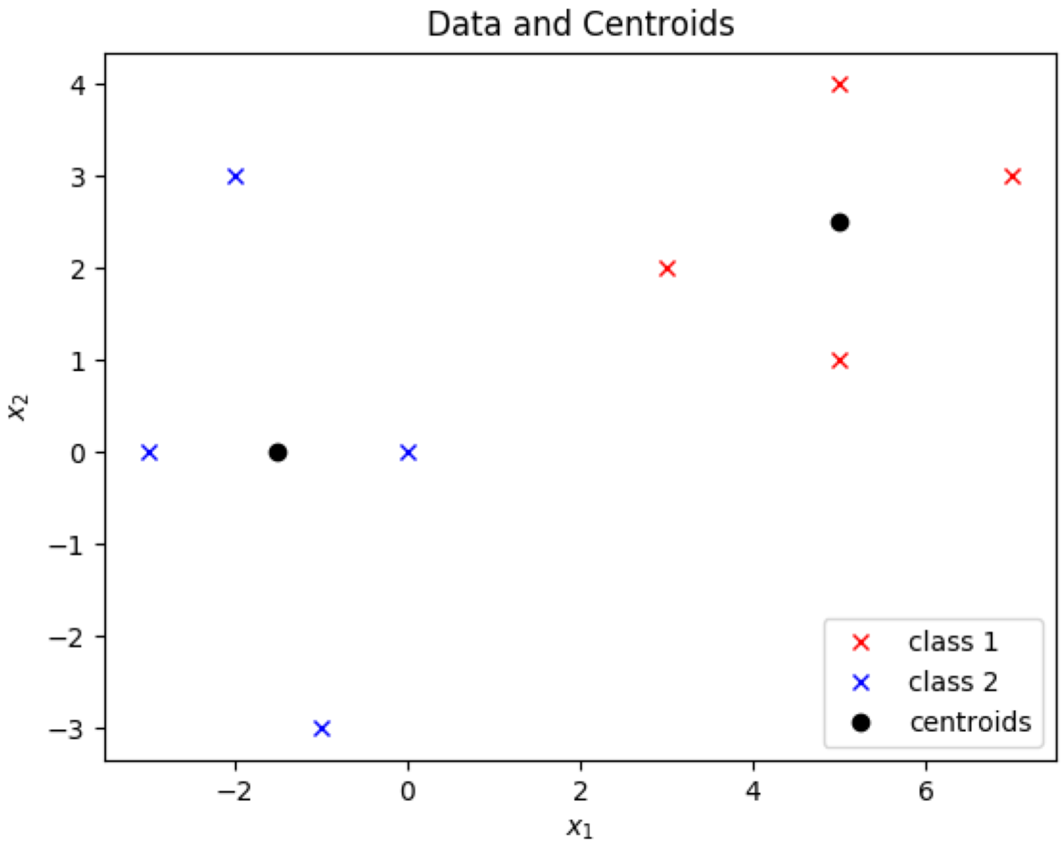
$$\mathbf{x} = \begin{bmatrix} 4 \\ 2 \end{bmatrix}; \quad \mathbf{x} = \begin{bmatrix} 0 \\ 5 \end{bmatrix}; \quad \mathbf{x} = \begin{bmatrix} \frac{36}{13} \\ 0 \end{bmatrix};$$

$\boldsymbol{x}_1 = \begin{pmatrix} 5 \\ 1 \end{pmatrix}, \boldsymbol{x}_2 = \begin{pmatrix} 7 \\ 3 \end{pmatrix}, \boldsymbol{x}_3 = \begin{pmatrix} 3 \\ 2 \end{pmatrix}, \boldsymbol{x}_4 = \begin{pmatrix} 5 \\ 4 \end{pmatrix}$ → class 1

$\boldsymbol{x}_5 = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \boldsymbol{x}_6 = \begin{pmatrix} -1 \\ -3 \end{pmatrix}, \boldsymbol{x}_7 = \begin{pmatrix} -2 \\ 3 \end{pmatrix}, \boldsymbol{x}_8 = \begin{pmatrix} -3 \\ 0 \end{pmatrix}$ → class 2

Center of class-1: $\chi_1 = \frac{1}{4}(\boldsymbol{x}_1 + \boldsymbol{x}_2 + \boldsymbol{x}_3 + \boldsymbol{x}_4) = \begin{pmatrix} 5.0 \\ 2.5 \end{pmatrix}$

Class-2: $\chi_2 = \frac{1}{4}(\boldsymbol{x}_5 + \boldsymbol{x}_6 + \boldsymbol{x}_7 + \boldsymbol{x}_8) = \begin{pmatrix} -1.5 \\ 0.0 \end{pmatrix}$


Data and Centroids
class 1
class 2
centroids
x_1
x_2
−2
0
2
4
6
4
3
2
1
0
−1
−2
−3

The middle point connecting two centroids: $\frac{1}{2}(\chi_1 + \chi_2)$

If $\boldsymbol{x}$ is any point on the boundary line, vector connecting point to the middle point:

$$\boldsymbol{x} - \frac{1}{2}(\chi_1 + \chi_2) \quad \text{--(1)}$$

The vector connecting two centroids, $\chi_1 - \chi_2$ --(2)

Given that the boundary line passes through the middle of the centers and is normal to the line connecting to centers, the inner product of (1) and (2) should be zero.

$$(\chi_1 - \chi_2)^T \left(\boldsymbol{x} - \frac{1}{2}(\chi_1 + \chi_2) \right) = 0$$

$$(\chi_1 - \chi_2)^T \boldsymbol{x} - \frac{1}{2}(\chi_1 - \chi_2)^T(\chi_1 + \chi_2) = 0$$

$$(\chi_1 - \chi_2)^T \boldsymbol{x} - \frac{1}{2}({\chi_1}^T\chi_1 + {\chi_1}^T\chi_2 - {\chi_2}^T\chi_1 - {\chi_2}^T\chi_2) = 0$$

$$(\chi_1 - \chi_2)^T \boldsymbol{x} - \frac{1}{2}({\chi_1}^T\chi_1 - {\chi_2}^T\chi_2) = 0$$

$$(\chi_1 - \chi_2)^T \boldsymbol{x} - \frac{1}{2}(\|\chi_1\|^2 - \|\chi_2\|^2) = 0$$

The above equation gives the equation of the boundary as $\boldsymbol{x}$ represents a general point on the boundary.

$$(\chi_1 - \chi_2)^T \boldsymbol{x} - \frac{1}{2}(\|\chi_1\|^2 - \|\chi_2\|^2) = 0$$

Comparing with the linear decision boundary implemented by a neuron, $u = \boldsymbol{w}^T \boldsymbol{x} + b = 0$:

$$\boldsymbol{w} = \chi_1 - \chi_2 = \begin{pmatrix} 6.5 \\ 2.5 \end{pmatrix}$$

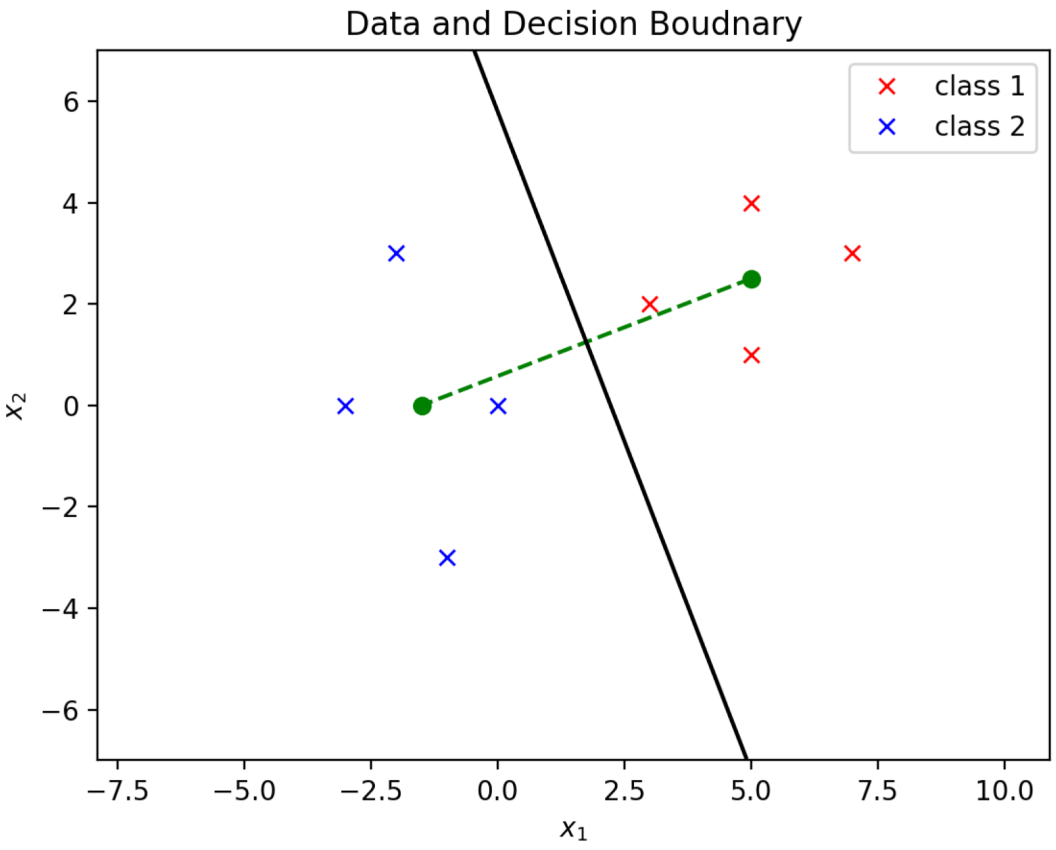
$$b = -\frac{1}{2}(\|\chi_1\|^2 - \|\chi_2\|^2) = -14.5$$

If $\boldsymbol{x} = (x_1, x_2)$, the equation of the boundary line is given by

$$\boldsymbol{w}^T \boldsymbol{x} + b = (6.5 \quad 2.5)\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} - 14.5 = 0$$

$$6.5x_1 + 2.5x_2 - 14.5 = 0$$

Q1



Data and Decision Boudnary
class 1
class 2
x_1
x_2
−7.5
−5.0
−2.5
0.0
2.5
5.0
7.5
10.0
6
4
2
0
−2
−4
−6

$$u = \boldsymbol{w}^T\boldsymbol{x} + b = (6.5 \quad 2.5)\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} - 14.5$$

Training patterns

$\boldsymbol{x}$	$\begin{pmatrix} 5 \\ 1 \end{pmatrix}$	$\begin{pmatrix} 7 \\ 3 \end{pmatrix}$	$\begin{pmatrix} 3 \\ 2 \end{pmatrix}$	$\begin{pmatrix} 5 \\ 4 \end{pmatrix}$	$\begin{pmatrix} 0 \\ 0 \end{pmatrix}$	$\begin{pmatrix} -1 \\ -3 \end{pmatrix}$	$\begin{pmatrix} -2 \\ 3 \end{pmatrix}$	$\begin{pmatrix} -3 \\ 0 \end{pmatrix}$
$\boldsymbol{u}$	20.5	38.5	10.0	28.0	-14.5	-28.5	-20.0	-34.0
class	1	1	1	1	2	2	2	2

That is

$$u > 0.0 \rightarrow class\ 1$$
$$u \leq 0.0 \rightarrow class\ 2$$

$$u = \boldsymbol{w}^T\boldsymbol{x} + b = (6.5 \quad 2.5)\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} - 14.5$$

$$u > 0.0 \rightarrow class\ 1$$
$$u \leq 0.0 \rightarrow class\ 2$$

<u>Testing patterns:</u>

$$\boldsymbol{x} = \begin{pmatrix} 4 \\ 2 \end{pmatrix}, u = 16.5 > 0 \rightarrow class\ 1$$

$$\boldsymbol{x} = \begin{pmatrix} 0 \\ 5 \end{pmatrix}, u = -2 \leq 0 \rightarrow class\ 2$$

$$\boldsymbol{x} = \begin{pmatrix} 36/13 \\ 0 \end{pmatrix}, u = 3.5 > 0 \rightarrow class\ 1$$

2. Design a dichotomizer that performs the following classification:

$$\boldsymbol{x}_1 = \begin{pmatrix} 0.8 \\ 0.5 \\ 0.0 \end{pmatrix}, x_2 = \begin{pmatrix} 0.9 \\ 0.7 \\ 0.3 \end{pmatrix}, \boldsymbol{x}_3 = \begin{pmatrix} 1.0 \\ 0.8 \\ 0.5 \end{pmatrix} \rightarrow Class\ A$$

$$\boldsymbol{x}_4 = \begin{pmatrix} 0.0 \\ 0.2 \\ 0.3 \end{pmatrix}, \boldsymbol{x}_5 = \begin{pmatrix} 0.2 \\ 0.3 \\ 0.5 \end{pmatrix}, \boldsymbol{x}_6 = \begin{pmatrix} 0.4 \\ 0.7 \\ 0.8 \end{pmatrix} \rightarrow Class\ B$$

Demonstrate one iterations of learning of above two-class classification with

(a) a discrete perceptron using discrete perceptron learning algorithm.

(b) a logistic regression neuron using stochastic gradient descent learning.

Draw the decision boundary learned by the neurons.

Initialize weights randomly and biases to 0.0. Set the learning parameter $\alpha = 0.1$.

Discrete perceptron:

$$y = f(u) = \begin{cases} 1, & u > 0 \\ 0, & u \leq 0 \end{cases}$$

Learning is by discrete perceptron learning algorithm

Discrete perceptron learning algorithm

Given a training dataset $\{(\boldsymbol{x}_p, d_p)\}_{p=1}^{P}$

Set the learning parameter α

Initialize $\boldsymbol{w}$ and b

Repeat until convergence:

For every training pattern $(\boldsymbol{x}_p, d_p)$:

$$u_p = \boldsymbol{x}_p{}^T \boldsymbol{w} + b$$

$$y_p = 1(u_p > 0)$$

$$\boldsymbol{w} \leftarrow \boldsymbol{w} + \alpha(d_p - y_p)\boldsymbol{x}_p$$

$$b \leftarrow b + \alpha(d_p - y_p)$$

$$\boldsymbol{x}_1 = \begin{pmatrix} 0.8 \\ 0.5 \\ 0.0 \end{pmatrix}, \boldsymbol{x}_2 = \begin{pmatrix} 0.9 \\ 0.7 \\ 0.3 \end{pmatrix}, \boldsymbol{x}_3 = \begin{pmatrix} 1.0 \\ 0.8 \\ 0.5 \end{pmatrix} \rightarrow Class\ A$$

$$\boldsymbol{x}_4 = \begin{pmatrix} 0.0 \\ 0.2 \\ 0.3 \end{pmatrix}, \boldsymbol{x}_5 = \begin{pmatrix} 0.2 \\ 0.3 \\ 0.5 \end{pmatrix}, \boldsymbol{x}_6 = \begin{pmatrix} 0.4 \\ 0.7 \\ 0.8 \end{pmatrix} \rightarrow Class\ B$$

$$y = f(u) = \begin{cases} 1, & u > 0 \\ 0, & u \leq 0 \end{cases}$$

Let targets: For class A, $d = 0$; and class B, $d = 1$

Learning factor $\alpha = 0.1$.

Weights are initialized randomly and biases to zero:

$$\boldsymbol{w} = \begin{pmatrix} 0.54 \\ 0.28 \\ 0.42 \end{pmatrix}, b = 0.0$$

Iteration 1

Shuffle the input patterns and apply to the network in random order.

Apply pattern $\boldsymbol{x} = \begin{pmatrix} 0.9 \\ 0.7 \\ 0.3 \end{pmatrix}$, $d = 0$

$$u = \boldsymbol{x}^T \boldsymbol{w} + b = \begin{pmatrix} 0.9 & 0.7 & 0.3 \end{pmatrix} \begin{pmatrix} 0.54 \\ 0.28 \\ 0.42 \end{pmatrix} + 0.0 = 0.81$$

$$y = f(u) = 1.0$$
$$\delta = d - y = -1.0$$

$$\boldsymbol{w} = \boldsymbol{w} + \alpha\delta\boldsymbol{x} = \begin{pmatrix} 0.54 \\ 0.28 \\ 0.42 \end{pmatrix} + 0.1 \times (-1.0) \begin{pmatrix} 0.9 \\ 0.7 \\ 0.3 \end{pmatrix} = \begin{pmatrix} 0.45 \\ 0.21 \\ 0.39 \end{pmatrix}$$
$$b = b + \alpha\delta = 0.0 + 0.1 \times (-1.0) = -0.1$$

Apply pattern $\boldsymbol{x} = \begin{pmatrix} 0.0 \\ 0.2 \\ 0.3 \end{pmatrix}, d = 1$

$$u = \boldsymbol{x}^T \boldsymbol{w} + b = \begin{pmatrix} 0.0 & 0.2 & 0.3 \end{pmatrix} \begin{pmatrix} 0.45 \\ 0.21 \\ 0.39 \end{pmatrix} - 0.1 = 0.06$$

$$y = f(u) = 1.0$$
$$\delta = d - y = 0$$

$$\boldsymbol{w} = \boldsymbol{w} + \alpha\delta\boldsymbol{x} = \begin{pmatrix} 0.45 \\ 0.21 \\ 0.39 \end{pmatrix} + 0\times(-1.0) \begin{pmatrix} 0.0 \\ 0.2 \\ 0.3 \end{pmatrix} = \begin{pmatrix} 0.45 \\ 0.21 \\ 0.39 \end{pmatrix}$$
$$b = b + \alpha\delta = -0.1 + 0.1\times(0.0) = -0.1$$

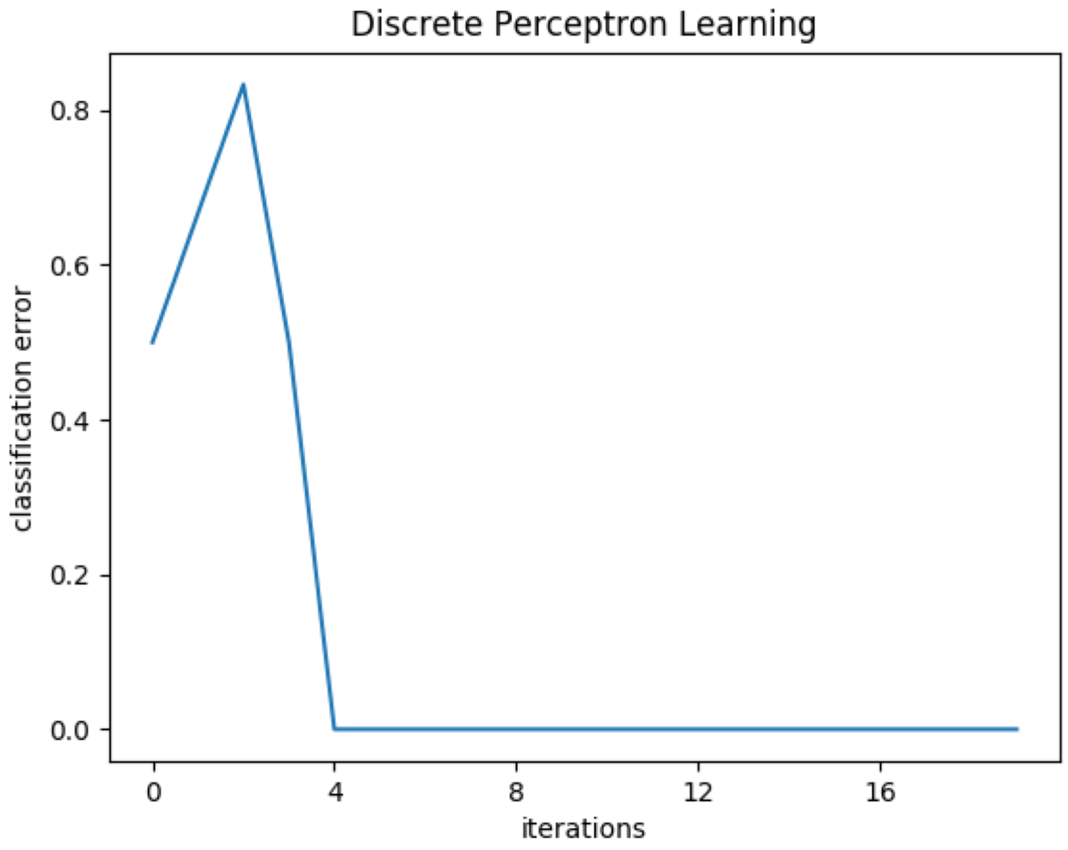
Continue applying the rest of the patterns

Then. continue epochs 2, 3, …… until convergence.

Iteration 1

x	d	u	y	$\boldsymbol{\delta}$	$\boldsymbol{w}$	b
$\begin{pmatrix}0.9\\0.7\\0.3\end{pmatrix}$	0	0.81	1	-1	$\begin{pmatrix}0.45\\0.21\\0.39\end{pmatrix}$	-0.1
$\begin{pmatrix}0.0\\0.2\\0.3\end{pmatrix}$	1	0.06	1	0	$\begin{pmatrix}0.45\\0.21\\0.39\end{pmatrix}$	-0.1
$\begin{pmatrix}0.4\\0.7\\0.8\end{pmatrix}$	1	0.54	1	0	$\begin{pmatrix}0.45\\0.21\\0.39\end{pmatrix}$	-0.1
$\begin{pmatrix}0.2\\0.3\\0.5\end{pmatrix}$	1	0.25	1	0	$\begin{pmatrix}0.45\\0.21\\0.39\end{pmatrix}$	-0.1
$\begin{pmatrix}1.0\\0.8\\0.5\end{pmatrix}$	0	0.72	1	-1	$\begin{pmatrix}0.35\\0.13\\0.34\end{pmatrix}$	-0.2
$\begin{pmatrix}0.8\\0.5\\0.0\end{pmatrix}$	0	0.15	1	-1	$\begin{pmatrix}0.27\\0.08\\0.34\end{pmatrix}$	-0.3

Classification errors = 3


Discrete Perceptron Learning
classification error
0.0
0.2
0.4
0.6
0.8
0
4
8
12
16
iterations

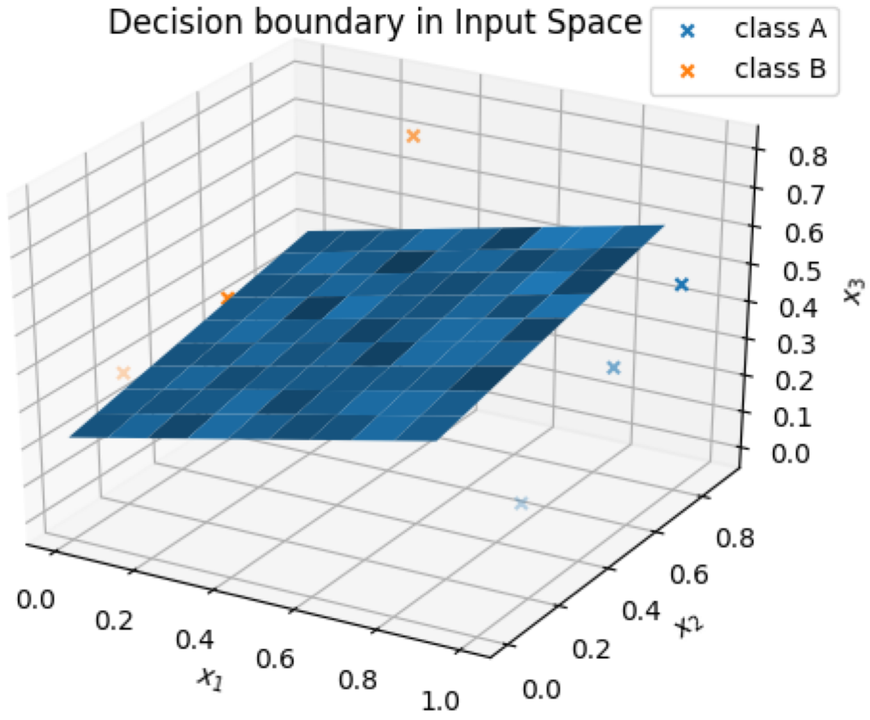
At convergence, $\boldsymbol{w} = \begin{pmatrix} -0.11 \\ -0.08 \\ 0.45 \end{pmatrix}$, $b = -0.1$

Decision boundary:

$$\boldsymbol{w}^T \boldsymbol{x} + b = 0$$

$$\begin{pmatrix} -0.11 & -0.08 & 0.45 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} - 0.1 = 0$$

$$-0.11x_1 - 0.08x_2 + 0.45x_3 - 0.1 = 0$$


Decision boundary in Input Space
class A
class B
0.0
0.2
0.4
0.6
0.8
1.0
x_1
0.0
0.2
0.4
0.6
0.8
x_2
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
x_3

Stochastic Gradient descent for logistic regression neuron

Given a training dataset $\{(\boldsymbol{x}_p, d_p)\}_{p=1}^{P}$

Set learning rate α

Initialize $\boldsymbol{w}$ and b

Iterate until convergence:

For every pattern $(\boldsymbol{x}_p, d_p)$:

Synaptic input $u_p = \boldsymbol{x}_p{}^T \boldsymbol{w} + b$

$$f(u_p) = \frac{1}{1+e^{-u_p}}$$

$$\boldsymbol{w} \leftarrow \boldsymbol{w} + \alpha\left(d_p - f(u_p)\right)\boldsymbol{x}_p$$

$$b \leftarrow b + \alpha\left(d_p - f(u_p)\right)$$

$$\boldsymbol{x}_1 = \begin{pmatrix} 0.8 \\ 0.5 \\ 0.0 \end{pmatrix}, \boldsymbol{x}_2 = \begin{pmatrix} 0.9 \\ 0.7 \\ 0.3 \end{pmatrix}, \boldsymbol{x}_3 = \begin{pmatrix} 1.0 \\ 0.8 \\ 0.5 \end{pmatrix} \rightarrow Class\ A$$

$$\boldsymbol{x}_4 = \begin{pmatrix} 0.0 \\ 0.2 \\ 0.3 \end{pmatrix}, \boldsymbol{x}_5 = \begin{pmatrix} 0.2 \\ 0.3 \\ 0.5 \end{pmatrix}, \boldsymbol{x}_6 = \begin{pmatrix} 0.4 \\ 0.7 \\ 0.8 \end{pmatrix} \rightarrow Class\ B$$

$$f(u) = \frac{1}{1 + e^{-u}}$$

$$y = 1(f(u) > 0.5)$$

Let targets: For class A, $d = 0$; and class B, $d = 1$

Learning factor $\alpha = 0.1$.

Weights are initialized randomly and biases to zero:

$$\boldsymbol{w} = \begin{pmatrix} 0.54 \\ 0.28 \\ 0.42 \end{pmatrix}, b = 0.0$$

Iteration 1

Shuffle the input patterns and apply to the network in random order.

Apply pattern $\boldsymbol{x} = \begin{pmatrix} 0.9 \\ 0.7 \\ 0.3 \end{pmatrix}$, $d = 0$

$$u = \boldsymbol{x}^T \boldsymbol{w} + b = (0.9 \quad 0.7 \quad 0.3) \begin{pmatrix} 0.54 \\ 0.28 \\ 0.42 \end{pmatrix} + 0.0 = 0.81$$

$$f(u) = \frac{1}{1 + e^{-u}} = 0.69$$

Entropy $= -d\log\big(f(u)\big) - (1-d)\log\big(1 - f(u)\big) = -\log(1 - f(u))$ = 1.18

$$\boldsymbol{w} = \boldsymbol{w} + \alpha\big(d - f(u)\big)\boldsymbol{x} = \begin{pmatrix} 0.54 \\ 0.28 \\ 0.42 \end{pmatrix} + 0.1\times(0 - 0.69) \begin{pmatrix} 0.9 \\ 0.7 \\ 0.3 \end{pmatrix} = \begin{pmatrix} 0.48 \\ 0.23 \\ 0.40 \end{pmatrix}$$

$$b = b + \alpha\big(d - f(u)\big) = 0.0 + 0.1\times(0 - 0.69) = -0.07$$

Apply pattern $\boldsymbol{x} = \begin{pmatrix} 0.0 \\ 0.2 \\ 0.3 \end{pmatrix}$, $d = 1$

$$u = \boldsymbol{x}^T \boldsymbol{w} + b = (0.0 \quad 0.2 \quad 0.3) \begin{pmatrix} 0.48 \\ 0.23 \\ 0.40 \end{pmatrix} - 0.07 = 0.10$$

$$f(u) = 0.52$$

Entropy $= -d\log\big(f(u)\big) - (1-d)\log\big(1 - f(u)\big) = -\log(f(u))$ = 0.645

$$\boldsymbol{w} = \boldsymbol{w} + \alpha\big(d - f(u)\big)\boldsymbol{x} = \begin{pmatrix} 0.48 \\ 0.23 \\ 0.40 \end{pmatrix} + 0.1\times(1 - 0.52) \begin{pmatrix} 0.0 \\ 0.2 \\ 0.3 \end{pmatrix} = \begin{pmatrix} 0.48 \\ 0.24 \\ 0.42 \end{pmatrix}$$

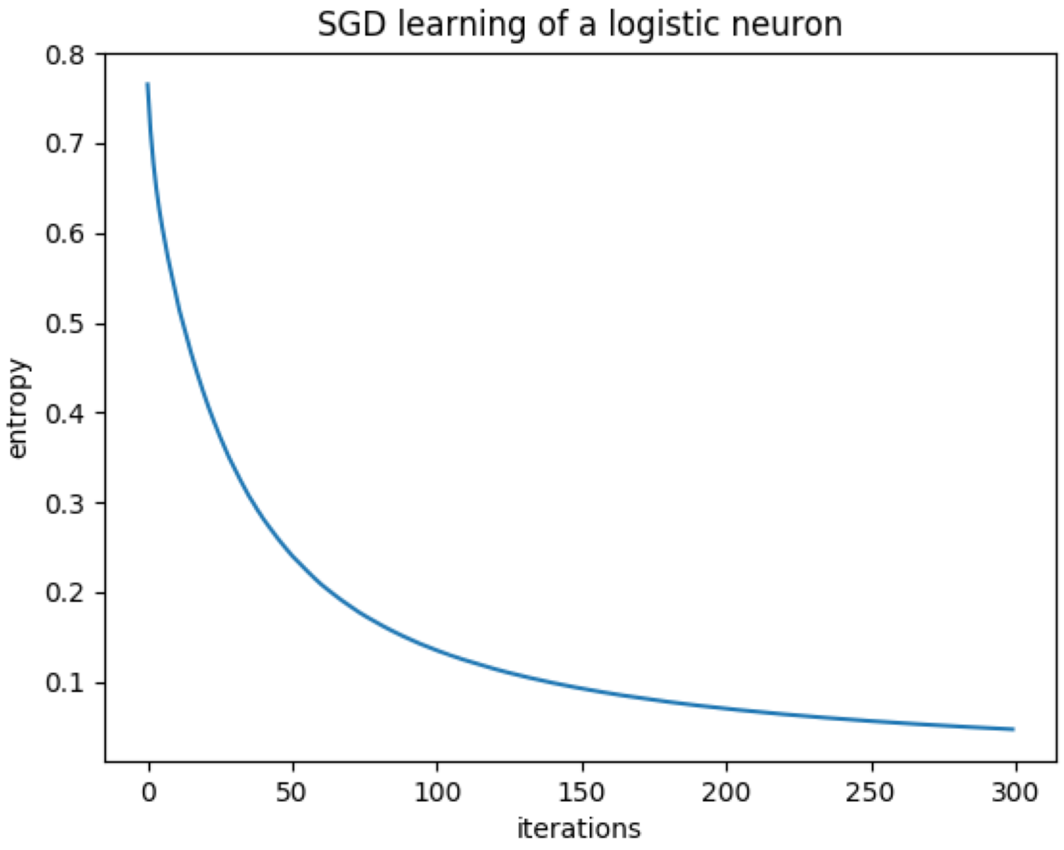
$$b = b + \alpha\big(d - f(u)\big) = 0.07 + 0.1\times(1 - 0.52) = -0.02$$

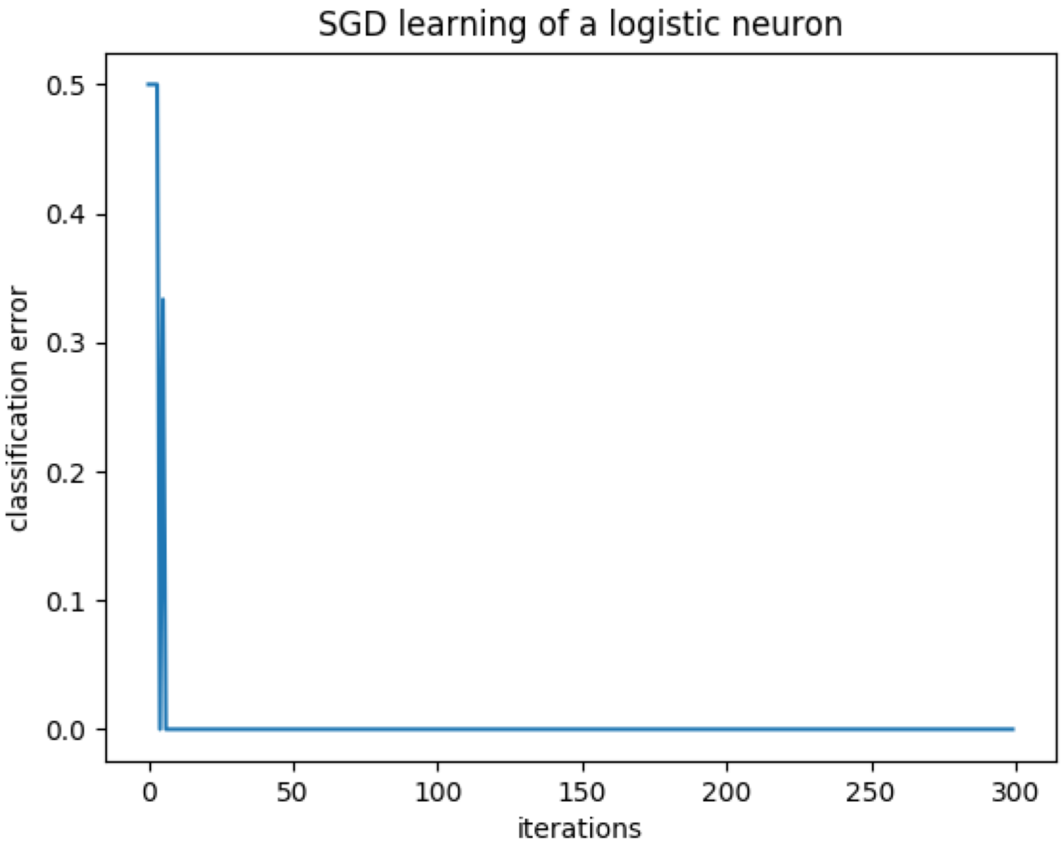
Continue applying the rest of the patterns

Then. continue epochs 2, 3, …… until convergence.

Iteration 1

x	d	u	$f(u)$	loss	$\boldsymbol{w}$	b
$\begin{pmatrix}0.9\\0.7\\0.3\end{pmatrix}$	0	0.81	0.69	1.18	$\begin{pmatrix}0.48\\0.23\\0.40\end{pmatrix}$	-0.07
$\begin{pmatrix}0.0\\0.2\\0.3\end{pmatrix}$	1	0.10	0.52	0.64	$\begin{pmatrix}0.48\\0.24\\0.42\end{pmatrix}$	-0.02
$\begin{pmatrix}0.4\\0.7\\0.8\end{pmatrix}$	1	0.67	0.66	0.41	$\begin{pmatrix}0.49\\0.26\\0.45\end{pmatrix}$	0.01
$\begin{pmatrix}0.2\\0.3\\0.5\end{pmatrix}$	1	0.41	0.60	0.51	$\begin{pmatrix}0.50\\0.28\\0.46\end{pmatrix}$	0.05
$\begin{pmatrix}1.0\\0.8\\0.5\end{pmatrix}$	0	1.01	0.73	1.32	$\begin{pmatrix}0.43\\0.22\\0.43\end{pmatrix}$	-0.02
$\begin{pmatrix}0.8\\0.5\\0.0\end{pmatrix}$	0	0.43	0.61	0.93	$\begin{pmatrix}0.38\\0.19\\0.43\end{pmatrix}$	-0.08


SGD learning of a logistic neuron
entropy
iterations
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
50
100
150
200
250
300


SGD learning of a logistic neuron
classification error
0.5
0.4
0.3
0.2
0.1
0.0
0
50
100
150
200
250
300
iterations

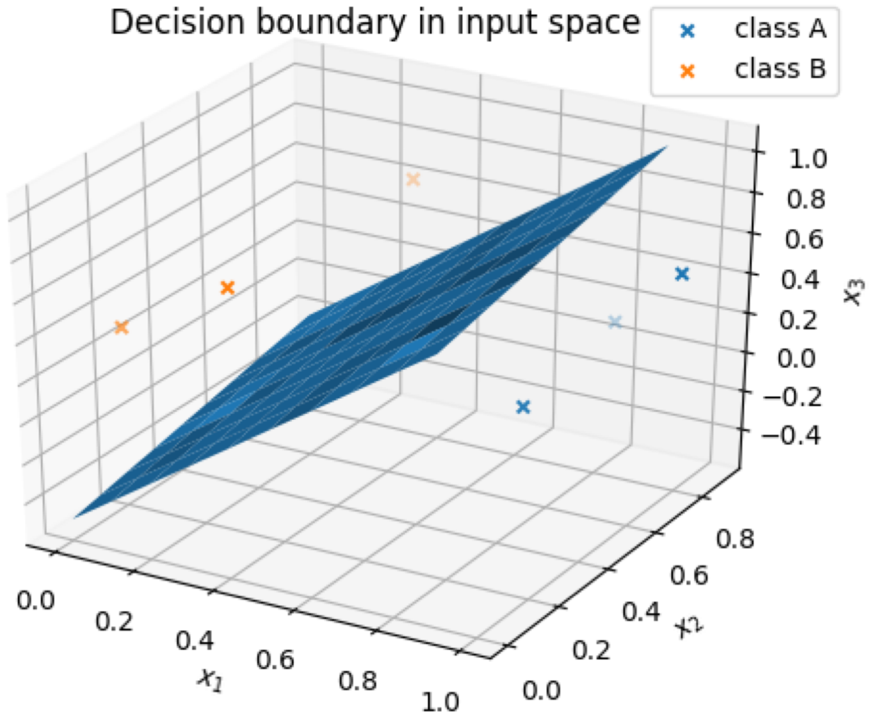
At convergence, $\boldsymbol{w} = \begin{pmatrix} -6.46 \\ -1.47 \\ 4.73 \end{pmatrix}$, $b = 2.35$

Decision boundary:

$$\boldsymbol{w}^T \boldsymbol{x} + b = 0$$

$$\begin{pmatrix} -6.46 & -1.47 & 4.73 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} + 2.35 = 0$$

$$-6.46x_1 - 1.47x_2 + 4.73x_3 + 2.35 = 0$$



Decision boundary in input space
class A
class B
1.0
0.8
0.6
0.4
0.2
0.0
−0.2
−0.4
x_3
0.0
0.2
0.4
0.6
0.8
1.0
x_1
0.0
0.2
0.4
0.6
0.8
x_2

3. Use gradient descent learning on a logistic regression neuron to realize the following classification of 3-dimensional inputs.

Class 1	Class 2
$(-1.75 \quad 0.34 \quad 1.15)$	$(-0.25 \quad 0.98 \quad 0.51)$
$(0.22 \quad -1.07 \quad -0.19)$	$(-0.58 \quad 0.82 \quad 0.67)$
$(0.26 \quad -0.46 \quad 0.44)$	$(-0.1 \quad -0.53 \quad 1.03)$
$(-0.44 \quad -1.12 \quad 1.62)$	$(0.18 \quad 0.94 \quad 0.73)$
$(1.54 \quad -0.25 \quad -0.84)$	$(1.36 \quad -0.33 \quad 0.06)$

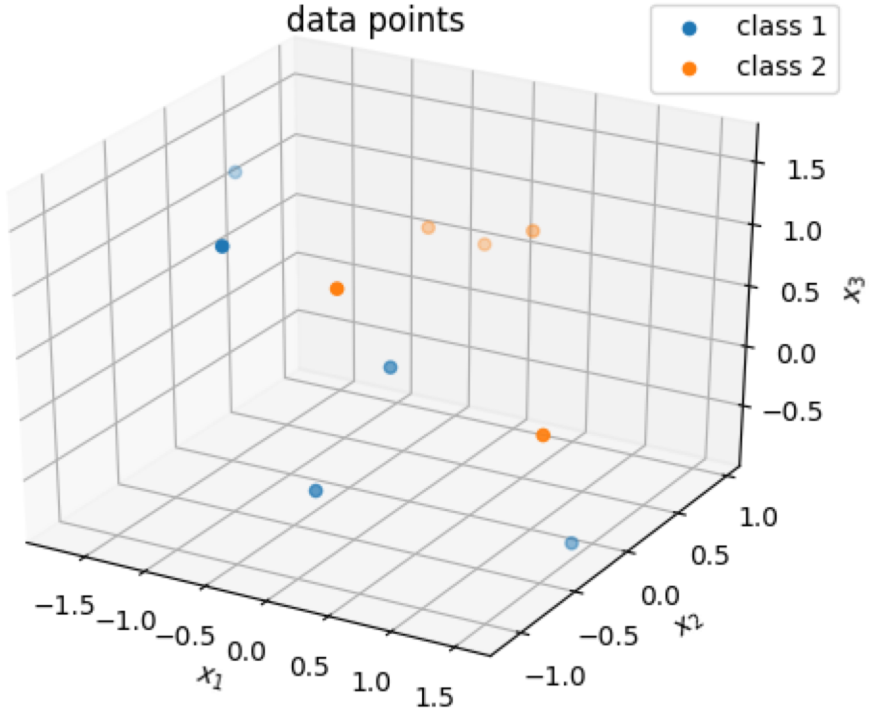
Initialize the weights randomly and bias to 0.0 and use a learning factor $\alpha = 0.4$

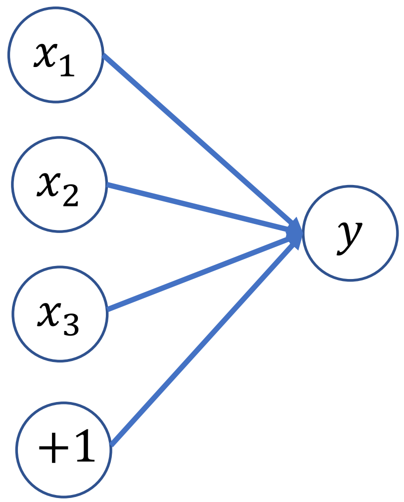
Show one iterations of learning of the neuron and plot classification error vs. iterations until convergence.

Find the probabilities that the trained network assigns to each input pattern belonging to its class label.

Let class-1, d = 1, and class-2 d = 0

$\boldsymbol{x = (x_1, x_2, x_3)}$	$\boldsymbol{d}$
$(-1.75, 0.34, 1.15)$	1
$(-0.25, 0.98, 0.51)$	0
$(0.22, -1.07, -0.19)$	1
$(0.26, -0.46, 0.44)$	1
$(-0.58, 0.82, 0.67)$	0
$(-0.1, -0.53, 1.03)$	0
$(-0.44, -1.12, 1.62)$	1
$(1.54, -0.25, -0.84)$	1
$(0.18, 0.94, 0.73)$	0
$(1.36, -0.33, 0.06)$	0


data points
class 1
class 2
x_1
x_2
x_3
−1.5
−1.0
−0.5
0.0
0.5
1.0
1.5



$$u = \boldsymbol{x}^T\boldsymbol{w} + b$$

$$f(u) = P(y = 1|\boldsymbol{x}) = \frac{1}{1 + e^{-u}}$$

$$y = 1(f(u) > 0.5)$$

Cost function is given by cross-entropy.

Given a training dataset $\{(\boldsymbol{x}_p, d_p)\}_{p=1}^{P}$ where $\boldsymbol{x}_p \in \boldsymbol{R}^n$ and $d_p \in \{0,1\}$, the cross-entropy for GD is given by

$$J = -\sum_{p=1}^{P} d_p\, log\left(f(u_p)\right) + (1 - d_p) log\left(1 - f(u_p)\right)$$

GD for logistic regression:

Given training data $(\boldsymbol{X}, \boldsymbol{d})$
Set learning rate α
Initialize $\boldsymbol{w}$ and b
Iterate until convergence:

$$\boldsymbol{u} = \boldsymbol{X}\boldsymbol{w} + b\mathbf{1}_P$$

$$f(\boldsymbol{u}) = \frac{1}{1+e^{-\boldsymbol{u}}}$$

$$\boldsymbol{w} \leftarrow \boldsymbol{w} + \alpha\boldsymbol{X}^T\big(\boldsymbol{d} - f(\boldsymbol{u})\big)$$

$$b \leftarrow b + \alpha\mathbf{1}_P^T\big(\boldsymbol{d} - f(\boldsymbol{u})\big)$$

$$\boldsymbol{X} = \begin{pmatrix} -1.75 & 0.34 & 1.15 \\ -0.25 & 0.98 & 0.51 \\ 0.22 & -1.07 & -0.19 \\ 0.26 & -0.46 & 0.44 \\ -0.58 & 0.82 & 0.67 \\ -0.1 & -0.53 & 1.03 \\ -0.44 & -1.12 & 1.62 \\ 1.54 & -0.25 & -0.84 \\ 0.18 & 0.94 & 0.73 \\ 1.36 & -0.33 & 0.06 \end{pmatrix} \text{ and } \boldsymbol{d} = \begin{pmatrix} 1 \\ 0 \\ 1 \\ 1 \\ 0 \\ 0 \\ 1 \\ 1 \\ 0 \\ 0 \end{pmatrix}$$

Learning factor $\alpha = 0.4$.

Weights are initialized randomly and biases to zero:

$$\boldsymbol{w} = \begin{pmatrix} 0.77 \\ 0.02 \\ 0.63 \end{pmatrix}, b = 0.0$$

Iteration 1:

$$\boldsymbol{u} = \boldsymbol{X}\boldsymbol{w} + b\mathbf{1} = \begin{pmatrix} -1.75 & 0.34 & 1.15 \\ -0.25 & 0.98 & 0.51 \\ 0.22 & -1.07 & -0.19 \\ 0.26 & -0.46 & 0.44 \\ -0.58 & 0.82 & 0.67 \\ -0.1 & -0.53 & 1.03 \\ -0.44 & -1.12 & 1.62 \\ 1.54 & -0.25 & -0.84 \\ 0.18 & 0.94 & 0.73 \\ 1.36 & -0.33 & 0.06 \end{pmatrix} \begin{pmatrix} 0.77 \\ 0.02 \\ 0.63 \end{pmatrix} + 0.0 \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{pmatrix} = \begin{pmatrix} -0.62 \\ 0.15 \\ 0.03 \\ 0.47 \\ -0.01 \\ 0.56 \\ 0.66 \\ 0.65 \\ 0.62 \\ 1.08 \end{pmatrix}$$

$$f(\boldsymbol{u}) = \frac{1}{1 + e^{(-\boldsymbol{u})}} = \begin{pmatrix} 0.35 \\ 0.54 \\ 0.51 \\ 0.62 \\ 0.50 \\ 0.64 \\ 0.66 \\ 0.66 \\ 0.65 \\ 0.75 \end{pmatrix}$$

$$\boldsymbol{y} = 1(f(\boldsymbol{u}) > 0.5) = \begin{pmatrix} 0 \\ 1 \\ 1 \\ 1 \\ 0 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{pmatrix}$$

Classification error = 5

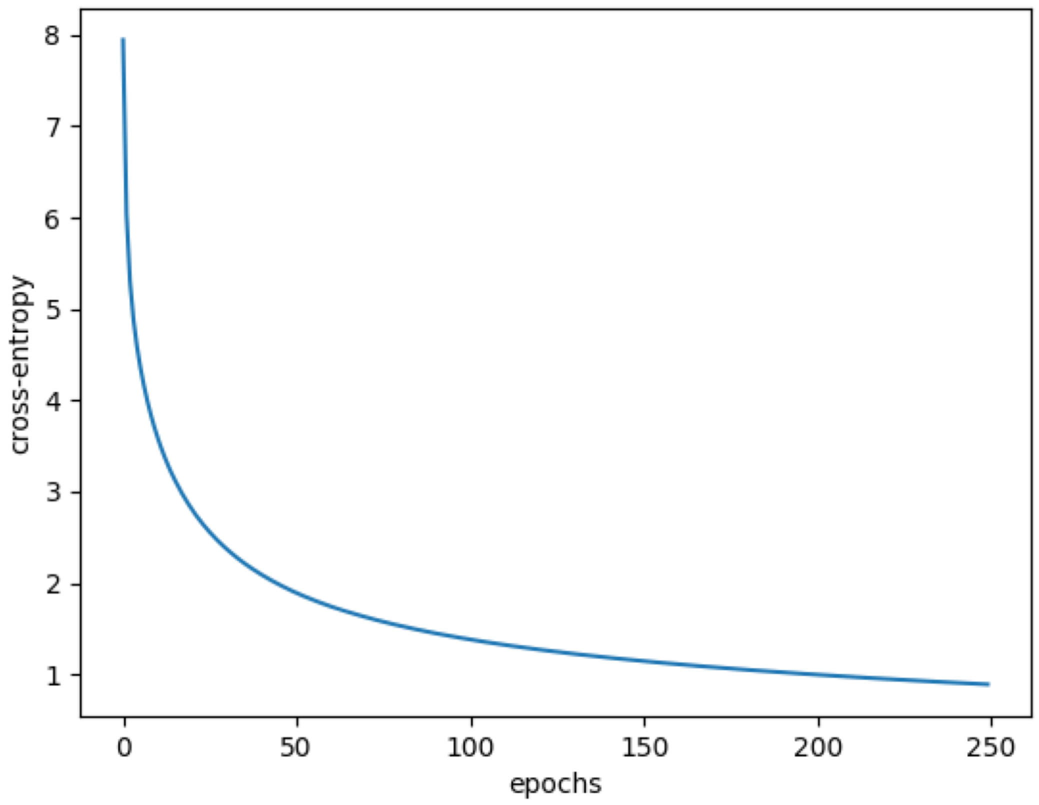
$$\begin{aligned} J &= -\sum_{p=1}^{P} d_p \, log\left(f(u_p)\right) + (1 - d_p) log\left(1 - f(u_p)\right) \\ &= -(log(0.35) + log(1 - 0.54) + \cdots + \log(1 - 0.75)) \\ &= 7.95 \end{aligned}$$

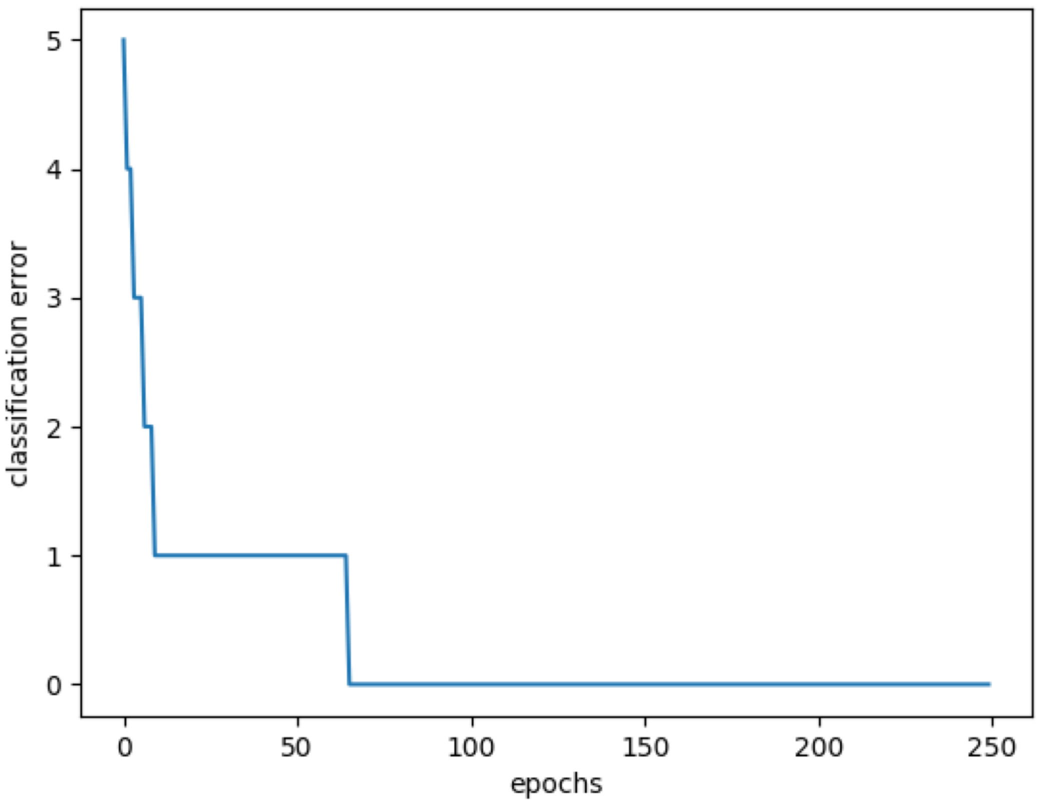
$$\boldsymbol{w} = \boldsymbol{w} + \alpha \boldsymbol{X}^{\boldsymbol{T}}\left(\boldsymbol{d} - f(\boldsymbol{u})\right)$$

$$= \begin{pmatrix} 0.77 \\ 0.02 \\ 0.63 \end{pmatrix} + 0.4 \begin{pmatrix} -1.75 & -0.25 & 0.22 & 0.26 & -0.58 & -0.1 & -0.44 & 1.54 & 0.18 & 1.36 \\ 0.34 & 0.98 & -1.07 & -0.46 & 0.82 & -0.53 & -1.12 & -0.25 & 0.94 & -0.33 \\ 1.15 & 0.51 & -0.19 & 0.44 & 0.67 & 1.03 & 1.62 & -0.84 & 0.73 & 0.06 \end{pmatrix} \left(\begin{pmatrix} 1 \\ 0 \\ 1 \\ 1 \\ 0 \\ 0 \\ 1 \\ 1 \\ 0 \\ 0 \end{pmatrix} - \begin{pmatrix} 0.35 \\ 0.54 \\ 0.51 \\ 0.62 \\ 0.50 \\ 0.64 \\ 0.66 \\ 0.66 \\ 0.65 \\ 0.75 \end{pmatrix} \right)$$

$$= \begin{pmatrix} 0.29 \\ -0.74 \\ 0.35 \end{pmatrix}$$

$$b = b + \alpha \mathbf{1}^{\boldsymbol{T}}\left(\boldsymbol{d} - f(\boldsymbol{u})\right) = 0.0 + 0.4\begin{pmatrix}1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1\end{pmatrix}\left(\begin{pmatrix}1\\0\\1\\1\\0\\0\\1\\1\\0\\0\end{pmatrix} - \begin{pmatrix}0.35\\0.54\\0.51\\0.62\\0.50\\0.64\\0.66\\0.66\\0.65\\0.75\end{pmatrix}\right) = -0.34$$


cross-entropy
epochs
8
7
6
5
4
3
2
1
0
50
100
150
200
250



classification error
epochs
0
1
2
3
4
5
0
50
100
150
200
250

At convergence:

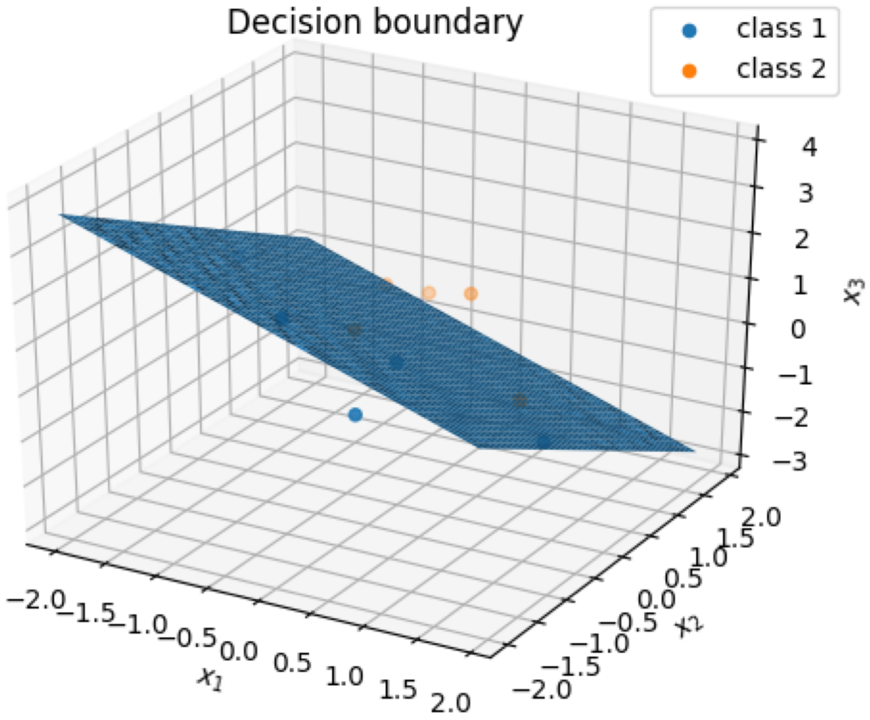
$$\boldsymbol{w} = \begin{pmatrix} -7.78 \\ -9.35 \\ -10.07 \end{pmatrix} \text{ and } b = 4.07$$

Decision boundary:

$$\boldsymbol{w}^T \boldsymbol{x} + b = 0$$

$$(-7.78 \quad -9.35 \quad -10.07) \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} + 4.07 = 0$$

$$-7.78x_1 - 9.35x_2 - 10.07x_3 + 4.07 = 0$$



Decision boundary
class 1
class 2
x_1
x_2
x_3
−2.0
−1.5
−1.0
−0.5
0.0
0.5
1.0
1.5
2.0
4
3
2
1
0
−1
−2
−3

At convergence:

$$f(\boldsymbol{u}) = \begin{pmatrix} 0.95 \\ 0.00 \\ 1.0 \\ 0.87 \\ 0.00 \\ 0.36 \\ 0.84 \\ 0.95 \\ 0.00 \\ 0.02 \end{pmatrix}$$

At convergence:

$$f(\boldsymbol{u}) = \begin{pmatrix} P(y=1|\boldsymbol{x}_1) \\ P(y=1|\boldsymbol{x}_2) \\ P(y=1|\boldsymbol{x}_3) \\ P(y=1|\boldsymbol{x}_4) \\ P(y=1|\boldsymbol{x}_5) \\ P(y=1|\boldsymbol{x}_6) \\ P(y=1|\boldsymbol{x}_7) \\ P(y=1|\boldsymbol{x}_8) \\ P(y=1|\boldsymbol{x}_9) \\ P(y=1|\boldsymbol{x}_{10}) \end{pmatrix} = \begin{pmatrix} 0.95 \\ 0.00 \\ 1.0 \\ 0.87 \\ 0.00 \\ 0.36 \\ 0.84 \\ 0.95 \\ 0.00 \\ 0.02 \end{pmatrix}$$

$$\begin{pmatrix} P(y=0|\boldsymbol{x}_1) \\ P(y=0|\boldsymbol{x}_2) \\ P(y=0|\boldsymbol{x}_3) \\ P(y=0|\boldsymbol{x}_4) \\ P(y=0|\boldsymbol{x}_5) \\ P(y=0|\boldsymbol{x}_6) \\ P(y=0|\boldsymbol{x}_7) \\ P(y=0|\boldsymbol{x}_8) \\ P(y=0|\boldsymbol{x}_9) \\ P(y=0|\boldsymbol{x}_{10}) \end{pmatrix} = \boldsymbol{1} - f(\boldsymbol{u}) = \begin{pmatrix} 0.05 \\ 1.00 \\ 0.0 \\ 0.13 \\ 1.00 \\ 0.64 \\ 0.16 \\ 0.05 \\ 1.00 \\ 0.98 \end{pmatrix}$$

$$\boldsymbol{y} = 1(f(\boldsymbol{u}) > 0.5) = \begin{pmatrix} 1 \\ 0 \\ 1 \\ 1 \\ 0 \\ 0 \\ 1 \\ 1 \\ 0 \\ 0 \end{pmatrix}$$