Homework #1 solution key

Prayag
Neural networks and learning systems-I

February 25, 2019

Problem 1.

Solution. The output y is given by

$$y(n) = \phi\left(w_{O1}y_1(n) + w_{O2}y_2(n)\right) \tag{1}$$

where

$$y_1(n) = \phi\left(x_1w_{11} + x_2w_{12} + y_1(n-1)w^{''}_{11} + y_2(n)w^{'}_{12}\right) \tag{2}$$

$$y_2(n) = \phi\left(x_1w_{21} + x_2w_{22} + y_2(n-1)w^{''}_{22} + y_1(n)w^{'}_{21}\right) \tag{3}$$

■

Problem 2.

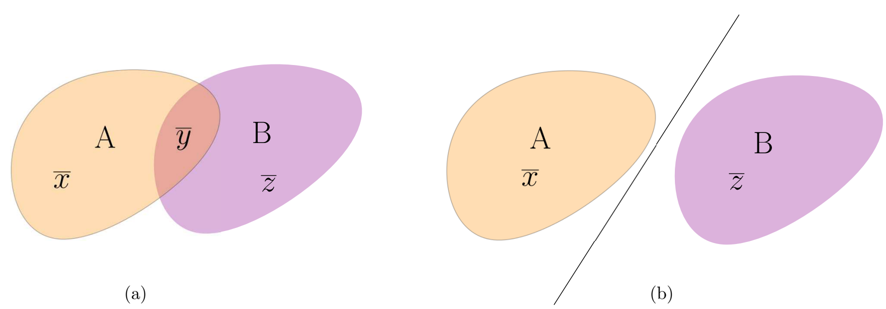


Figure 1: (a) overlapping of convex hulls A and B. (b) linearly separable convex hulls

Solution. (a) From Figure 1(a) we see that the two convex hulls are intersecting. Assume that there exist vectors $\overline{w}$ and $\overline{b}$ such that $\overline{w}^{\mathrm{T}}\overline{x}_n + \overline{b} > 0$ for all $\overline{x}_n \in \mathrm{A}$ and $\overline{w}^{\mathrm{T}}\overline{z}_n + \overline{b} < 0$ for all $\overline{z}_n \in \mathrm{B}$. We know that $\mathrm{A} \cap \mathrm{B} \neq \phi$ implies there exist at least one element $\overline{y} \in \mathrm{A} \cap \mathrm{B}$. We can write $\overline{y} = \sum_{n=1}^{\mathrm{N}_x} \alpha_n \overline{x}_n$ and $\overline{y} = \sum_{m=1}^{\mathrm{N}_z} \beta_m \overline{z}_m$ for some α and β satisfying the conditions as mentioned in the question. We look at $\overline{w}^{\mathrm{T}}\overline{y}$ which is given by

$$\overline{y} = \begin{cases} \sum_{n=1}^{\mathrm{N}_x} \alpha_n \overline{w}^{\mathrm{T}}\overline{x}_n + \overline{b} > 0, & \text{for all } \overline{x}_n \in \mathrm{A} \\ \sum_{n=1}^{\mathrm{N}_x} \alpha_n \overline{w}^{\mathrm{T}}\overline{z}_n + \overline{b} < 0, & \text{for all } \overline{z}_n \in \mathrm{B} \end{cases} \tag{4}$$

From equation (4), we see that it is contradicting the assumption of existence of vectors $\overline{w}$ and $\overline{b}$.

(b) From Figure 1(b) we see that $\exists$ vectors $\overline{w}^{\mathrm{T}}, \overline{b}$ such that $\overline{w}^{\mathrm{T}}\overline{x}_n + \overline{b} > 0$ for all $\overline{x}_n \in \mathrm{A}$ and $\overline{w}^{\mathrm{T}}\overline{z}_n + \overline{b} < 0$ for all $\overline{z}_n \in \mathrm{B}$. Now, we have to show $\mathrm{A} \cap \mathrm{B} = \phi$. We know that existence of $\overline{w}$ and $\overline{b}$ guarantees $\overline{x} = \sum_{n=1}^{\mathrm{N}_x} \alpha_n \overline{x}_n \notin \mathrm{B}$ and $\overline{z} = \sum_{m=1}^{\mathrm{N}_z} \beta_m \overline{z}_m \notin \mathrm{A}$ implying $\mathrm{A} \cap \mathrm{B} = \phi$.

■

Problem 3.

Solution. (a) Let $\overline{x}$ be the input with true class $\mathcal{C}_k$. The total expected loss with $L_{kk} = 0$ for $k = 1, \ldots, N$ is given by

$$\mathbb{E}(L) = \sum_{k=1}^{N}\sum_{j=1}^{N}\int_{\mathcal{C}_j} L_{kj} P\left(\overline{x} \in \mathcal{C}_k \mid \overline{x}\right) P(\overline{x}) \mathrm{d}\overline{x}. \tag{5}$$

Equation (5)

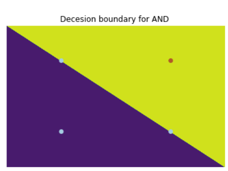
(b) The optimal rule for assigning class labels

$$j^\star = \min_j \left(\sum_{k=1}^{N} \int_{\mathcal{C}_j} L_{kj} P\left(\overline{x} \in \mathcal{C}_k \mid \overline{x}\right) P(\overline{x}) \mathrm{d}\overline{x} \right) \tag{6}$$

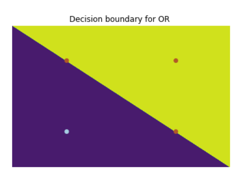
■

Problem 4.

Solution. (a) The figure 2 shows the decision boundary for the **AND** and **OR** operations.



(a)



(b)

Figure 2

(b) We know that the output of the **XOR** operation forms a non-linearly separable class as shown in Figure 3. Therefore, it is not possible to separate the two classes using perceptron.

■

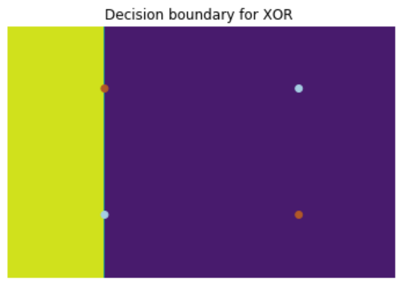


Figure 3: The two classes are non-linearly separable.

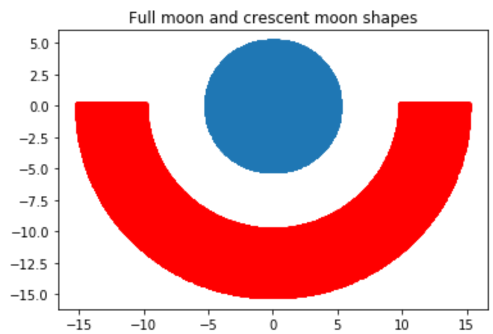


Figure 4: Full moon and crescent moon shapes with $R_{\text{m}} = 5$, $R_{\text{in}} = 10$, $R_{\text{out}} = 15$, and $D = 0$.

Problem 5.

Solution. (a) The required data points are generated as shown in Figure 4

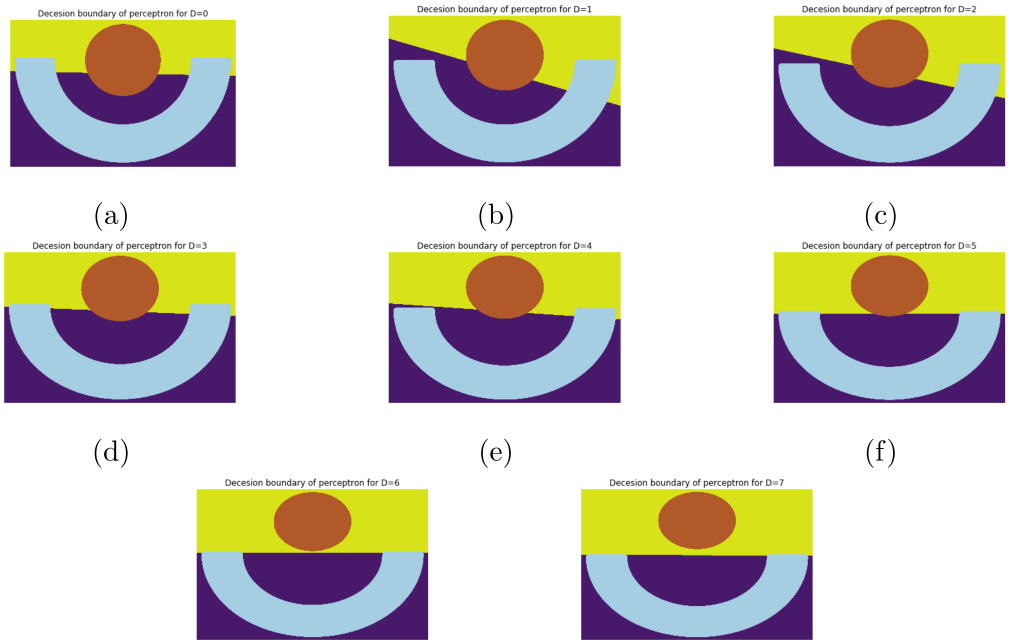
(b) The value of D is varied from 0 to 7 and the resulting perceptron decision boundary is shown in Figures 5a-5h.

(c) With $D = 7$, five different initial weight vectors were chosen and the resulting perceptron decision boundary is shown in Figures 6a-6e. The decision boundaries were different with different initial conditions.

(d) With $D = 7$, the sequence in which the input is presented is randomized and the resulting perceptron decision boundary is shown in Figures 7a-7e. We do not observe any significant difference in the decision boundaries.

(e) A Gaussian noise with mean 0 and standard deviation ranging from 1 to 3.1 is added to the dataset generated as shown in Figure 4 and the resulting perceptron decision boundary is shown in Figures 8a-8d. The stopping criterion is error threshold set to 10^{-6}. The linear classification is not possible with the increase in standard deviation.

(f) The initial learning rate is varied from 0.1 to 1 in steps of 0.1. The initial weight vector was set to $[0, 0]^{\text{T}}$. The error threshold as stopping criterion is set to 10^{-6}. The number



(a) (b) (c)

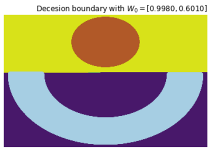
(d) (e) (f)

(g) (h)

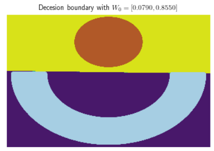
Figure 5

of epochs required for convergence is 3 for all initial learning rates. This indicates that the convergence is independent of the learning rate.

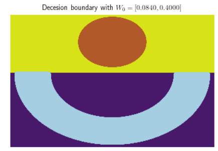
■



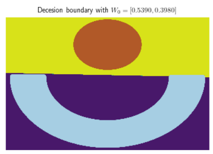
(a)



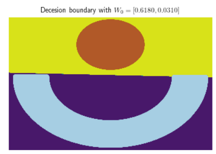
(b)



(c)

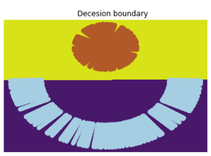


(d)

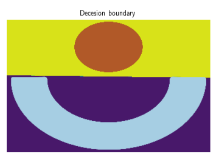


(e)

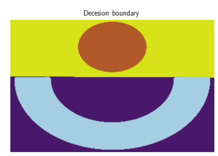
Figure 6



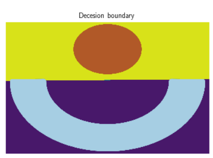
(a)



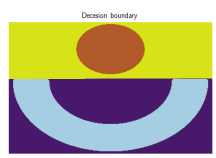
(b)



(c)

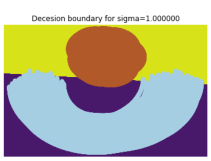


(d)

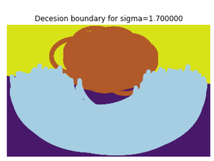


(e)

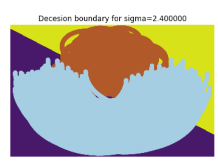
Figure 7



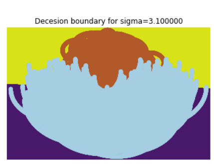
(a)



(b)



(c)



(d)

Figure 8