Homework #1 solution key

Prayag
Neural networks and learning systems-I

February 25, 2019

Problem 1.

Solution. The output \( y \) is given by

\[
y(n) = \phi \left( w_{O1} y_1(n) + w_{O2} y_2(n) \right)
\]

where

\[
y_1(n) = \phi \left( x_1 w_{11} + x_2 w_{12} + y_1(n - 1) w'_{11} + y_2(n) w'_{12} \right)
\]

\[
y_2(n) = \phi \left( x_1 w_{21} + x_2 w_{22} + y_2(n - 1) w'_{22} + y_1(n) w'_{21} \right)
\]

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 \( \overline{w}^{T} \overline{x}_n + \overline{b} > 0 \) for all \( \overline{x}_n \in A \) and \( \overline{w}^{T} \overline{z}_n + \overline{b} < 0 \) for all \( \overline{z}_n \in B \). We know that \( A \cap B \neq \emptyset \) implies there exist at least one element \( \overline{y} \in A \cap B \). We can write \( \overline{y} = \sum_{n=1}^{N_x} \alpha_n \overline{x}_n \) and \( \overline{y} = \sum_{m=1}^{N_z} \beta_m \overline{z}_m \) for some \( \alpha \) and \( \beta \) satisfying the conditions as mentioned in the question. We look at \( \overline{w}^{T} \overline{y} \) which is given by

\[
\overline{y} = \begin{cases} 
\sum_{n=1}^{N_x} \alpha_n \overline{w}^{T} \overline{x}_n + \overline{b} > 0, & \text{for all } \overline{x}_n \in A \\
\sum_{n=1}^{N_x} \alpha_n \overline{w}^{T} \overline{z}_n + \overline{b} < 0, & \text{for all } \overline{z}_n \in B 
\end{cases}
\]
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 \( \overline{w}^T, \overline{b} \) such that \( \overline{w}^T \overline{x}_n + \overline{b} > 0 \) for all \( \overline{x}_n \in A \) and \( \overline{w}^T \overline{z}_m + \overline{b} < 0 \) for all \( \overline{z}_m \in B \). Now, we have to show \( A \cap B = \phi \). We know that existence of \( \overline{w} \) and \( \overline{b} \) guarantees \( \overline{x} = \sum_{n=1}^{N} \alpha_n \overline{x}_n \notin B \) and \( \overline{z} = \sum_{m=1}^{N} \beta_m \overline{z}_m \notin A \) implying \( A \cap B = \phi \).

\[ \text{Problem 3.} \]

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

\[ \mathbb{E} (L) = \sum_{k=1}^{N} \sum_{j=1}^{N} \int_{C_j} L_{kj} P (\overline{x} \in C_k \mid \overline{x}) P(\overline{x}) d\overline{x}. \]  

Equation (5)

(b) The optimal rule for assigning class labels

\[ j^* = \min_{j} \left( \sum_{k=1}^{N} \int_{C_j} L_{kj} P (\overline{x} \in C_k \mid \overline{x}) P(\overline{x}) d\overline{x} \right) \]  

(6)

\[ \text{Problem 4.} \]

\textbf{Solution.} (a) The figure 2 shows the decision boundary for the \textbf{AND} and \textbf{OR} operations.

![Decision boundary for AND](image1)

(a)

![Decision boundary for OR](image2)

(b)

Figure 2

(b) We know that the output of the \textbf{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.
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 $[0, 0]^T$. The error threshold as stopping criterion is set to $10^{-6}$. The number
of epochs required for convergence is 3 for all initial learning rates. This indicates that
the convergence is independent of the learning rate.
Figure 6

Figure 7

Figure 8