

CSI 5810 (Assignment # 3)

1. The following examples from a two-class classification problem are given:

Class1: $[2 \ 2]^T$, $[3 \ 5]^T$; Class 2 $[1 \ 3]^T$, $[-1 \ -0.5]^T$

Starting with an augmented weight vector, $[1 \ 1 \ 1]^T$, determine a solution vector for above data using the perceptron learning rule. Show first five steps of weight vector updating.

2. Suppose you are given a collection of weak learners where each learner is able to operate with 60% accuracy. You combine seven of these learners with a majority rule for final output. What will be the accuracy of the ensemble?

3. Consider the following eight records; each record is described by two quantitative attributes:

$A = (2, 10)^t$, $B = (2, 5)^t$, $C = (8, 4)^t$, $D = (5, 8)^t$, $E = (7, 5)^t$, $F = (6, 4)^t$, $G = (1, 2)^t$, $H = (4, 9)^t$.

Let records "A", "B", "G", and "H" be from class 1 and the remaining four records from class 2. Using this information, construct the Fisher's linear discriminant function for this problem and determine the class label for the point $M = (3, 3)^t$.

4. Consider the following six examples with three attributes:

Example #	Color	Shape	Size	Class
1	Red	Square	Big	+
2	Blue	Square	Big	+
3	Red	Round	Small	-
4	Green	Square	Small	-
5	Red	Round	Big	+
6	Green	Square	Big	-

Determine the best attribute for root node of a decision tree classifier for above data.

5. Let $\mathbf{x}_1 = (4 \ 5)^t$, $\mathbf{x}_2 = (1 \ 4)^t$, $\mathbf{x}_3 = (0 \ 1)^t$ and $\mathbf{x}_4 = (5 \ 0)^t$ be four items for clustering. Consider the following three partitions:

A. $P_1 = \{\mathbf{x}_1, \mathbf{x}_2\}$, $P_2 = \{\mathbf{x}_3, \mathbf{x}_4\}$

B. $P_1 = \{\mathbf{x}_1, \mathbf{x}_4\}$, $P_2 = \{\mathbf{x}_2, \mathbf{x}_3\}$

C. $P_1 = \{\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3\}$, $P_2 = \{\mathbf{x}_4\}$.

Determine the partition favored by the sum-of-square-error (SSE) clustering criterion.

6. Consider the eight records of Exercise #3 without their class labels. Apply complete link clustering to this data and produce the dendrogram. This exercise must be done by hand without clustering software.