8

Combination of Classifiers

Learning Objectives

At the end of this chapter you will

- Know why an ensemble of classifiers are more often used than just one classifier. There are different methods for constructing ensembles. They are
  - Sub-sampling the training examples using
    - Bagging
    - Leaving out disjoint sub-sets
    - ADABOOST
  - Manipulating input features
  - Manipulating output targets
  - Injecting randomness into the learning algorithm
- Learn the different methods for combining classifiers

8.1 Introduction

A combination or an ensemble of classifiers is a set of classifiers whose individual decisions are combined to classify new examples. A combination of classifiers is often much more accurate than the individual classifiers that make them up. One reason for this could be that the training data may not provide sufficient information for choosing a single best classifier and a combination is the best compromise. Another reason could be that the learning algorithms used may not be able to solve the difficult search problem posed. Since solving the search problem may be difficult, suitable heuristics may be used in the search. As a consequence of this, even though with the training examples and prior knowledge, a unique best hypothesis exists, we may not be able to find it. A combination of classifiers is a way of compensating for imperfect classifiers. The learning algorithms we use may give us good approximations to the true value but may not be the right hypothesis. By taking a weighted combination of these approximations, we may be able to represent
the true hypothesis. In fact, the combination could be equivalent to very complex decision trees.

**Example 1**

![Figure 8.1](image.png)

**Figure 8.1** Training data set and the test pattern

Figure 8.1 shows a data set consisting of two classes, the class X and the class O. The patterns are:

\[
X_1 = (0.5, 1, X); 
X_2 = (1, 1, X); 
X_3 = (0.5, 0.5, X); 
X_4 = (1, 0.5, X); 
X_5 = (2, 2.5, X); 
X_6 = (2, 2, X); 
X_7 = (4, 1.25, O); 
X_8 = (5, 1.25, O); 
X_9 = (4, 0.5, O); 
X_{10} = (5, 0.5, O);
\]

If different classifiers are formed by taking a different sub-set of the training set, any training set which contains either or both the patterns \(X_5\) and \(X_6\) will classify the test pattern \(P\) at \((3, 2)\) as belonging to Class X using the nearest neighbour algorithm. But if these patterns are not in the sub-set, then \(P\) is classified as belonging to Class O. If the majority of the classifiers do not have either \(X_5\) or \(X_6\), if the combination of the classifiers is done according to a majority vote, \(P\) will be classified as belonging to Class O. If a majority of the classifiers contain either \(X_5\) or \(X_6\) or both, then \(P\) will be classified as belonging to Class X.