MODULE 16
Combination of Classifiers

LESSON 36
Introduction to Combining Classifiers

Keywords: Ensemble, Sampling, Bagging

Combination of Classifiers

- When a new pattern is to be classified, the decision on the class label is made by looking at the decisions of different classifiers and combining them.
- The classifiers used here can be either different classification algorithms, the same classifier used on different training data or the same classifier used on the training data using different features.
- The combination of the class labels of the different classifiers can also be done in several ways which will be discussed later in the module.
- Quite often, the combination of classifiers works better than using a single classifier because of the following reasons.
    1. If the training data does not have discriminative examples, whichever classifier is used, the classification accuracy will suffer. In such cases, use of combination of classifiers may help.
    2. If the new pattern to be classified is such that it is an outlier or it is on the boundary of classes and is difficult to classify, combination of classifiers may give a better classification.
    3. For the particular classification problem, there may not be a single classifier which works well and the combination of classifiers may compensate for this. Even though each classifier used is a weak classifier, the combination may be finding the right hypothesis.

Methods of Constructing Ensemble of Classifiers

- The different methods generally vary the input given to the classifier and then combine the decision of each classifier.
- These methods include

    1. Subsampling the training set
    2. Varying features used in the training set

3. Varying the classes by combining the classes and partitioning into two or more blocks.
4. Varying the classifier by using randomness
5. Some methods specific to certain type of classifiers.

Subsampling the Training Set

- The classifier is run on different datasets obtained from the training set.
- Each dataset used for classification of the new pattern is a subset of the training set.
- The class labels from each such subset are combined together to get the final classification.
- There are different techniques depending on how the subsets are found.
- Some of these techniques are
    1. Bagging
    2. Leaving out disjoint subsets
    3. Boosting
- **Bagging** refers to forming a subset of the training set by taking patterns at random from the training set.
- Each subset so formed is called a bootstrap replicate of the training set.
- This technique is also called *bootstrap aggregation*.
- Since the patterns are drawn at random with replacement, each bootstrap replicate may contain more than one copy of the same pattern.
- Generally each bootstrap replicate has a size which is two-thirds of the size of the original training data.
- Consider the set of patterns in Figure 1.

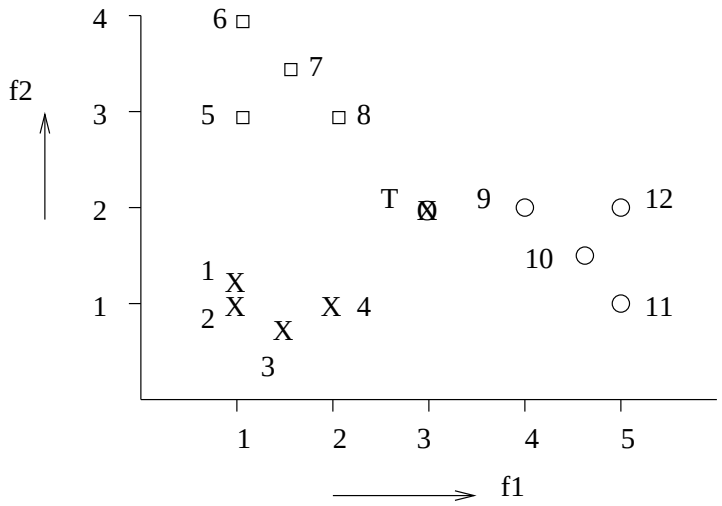

Figure 1: Dataset of three classes

- This consists of the patterns

Pattern 1 : (1,1.25,1)	Pattern 2 : (1,1,1)
Pattern 3 : (1.5,0.75,1)	Pattern 4 : (2,1,1)
Pattern 5 : (1,3,2)	Pattern 6 : (1,4,2)
Pattern 7 : (1.5,3.5,2)	Pattern 8 : (2,3,2)
Pattern 9 : (4,2,3)	Pattern 10 : (4.5,1.5,3)
Pattern 11 : (5,1,3)	Pattern 12 : (5,2,3)

Each triplet consists of feature 1, feature 2 and the class label.
Let us have a test pattern at (3,2).
To generate a bootstrap replicate, let us draw at random two patterns from each class for each classifier.

Classifier 1 :

Let the two random patterns drawn from each class be

Class 1 : 1 and 2
Class 2 : 8 and 7
Class 3 : 11 and 12

Now, if we use nearest neighbour classifier to find the class of the test pattern T using the 6 patterns above, T is closest to 8 and is therefore classified as belonging to Class 2.

Classifier 2 :

Let the two random patterns from each class be

Class 1 : 1 and 4
Class 2 : 5 and 6
Class 3 : 9 and 11

Now, if we use the nearest neighbour classifier to find the class of the test pattern T using the 6 patterns above, T is closest to 9 and is therefore classified as belonging to Class 3.

Classifier 3 :

Let the two random patterns from each class be

Class 1 : 3 and 4
Class 2 : 5 and 8
Class 3 : 9 and 12

Now, if we use the nearest neighbour classifier to find the class of the test pattern T using the 6 patterns, above, T is closest to 9 and is therefore classified as belonging to Class 3.

We now have to combine the class label obtained in the three classifiers to decide on the classification of T. If we use majority vote, then T is

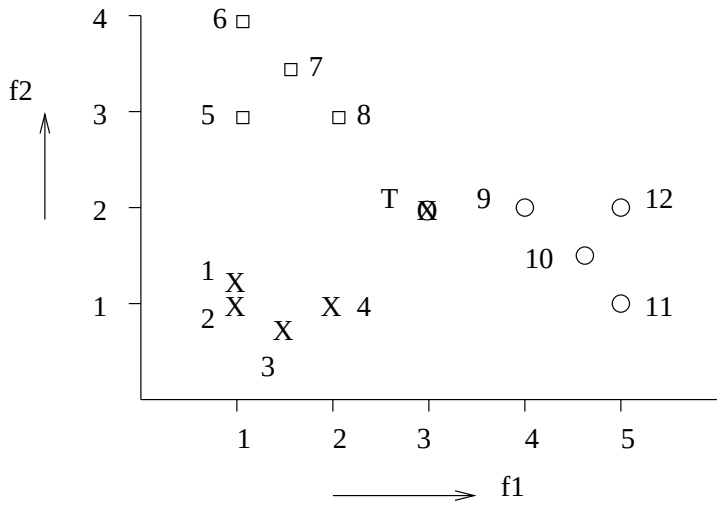

Figure 2: Dataset of three classes

given the class label 3.

Since for each classifier, the patterns are drawn at random, bagging is being carried out.

- **Leaving out one disjoint subset** refers to the process where the training dataset for each case is constructed by leaving out one disjoint subset every time.
- The training data is divided into a number of disjoint subsets.
- Each time one subset is left out which will result in different training subsets which are overlapping.
- As an example, let us take the same dataset as the one for bagging. This is shown in Figure 2.

  Let us have 6 subsets S_1, S_2, S_3, S_4, S_5 and S_6.

  S_1=(1,2), S_2=(3,4), S_3=(5,6), S_4=(7,8), S_5=(9,10), S_6=(11,12)

  If two subsets are left out each time,

1. If S_2 and S_5 are left out, the set of training examples is (1,2,5,6,7,8,11,12). If T is assigned the class label of the closest pattern, T is classified as belonging to Class 2.
2. If S_1 and S_3 are left out, the set of training examples is (3,4,7,8,9,10,11,12). If T is assigned the class label of the closest pattern, T is classified as belonging to Class 3.
3. If S_4 and S_5 are left out, the set of training examples is (1,2,3,4,5,6,11,12). If T is assigned the class label of the closest pattern, T is classified as belonging to Class 1.
4. If S_3 and S_4 are left out, the set of training examples is (1,2,3,4,9,10,11,12). Note that in this case there are no training patterns of Class 2. If T is assigned the class label of the closest pattern, T is classified as belonging to Class 3.

Taking these four cases, and the majority vote for classifying T, it would be classified as belonging to Class 3. In this way, it is run for a number of runs and the class of the test pattern decided.