MODULE 16
Combination of Classifiers

LESSON 38

Combination of Homogeneous Classifiers

Keywords: Varying Features, Classes, Randomness

Varying Features used in the Training set

- In this method, a subset of the features are chosen and only these features in the training set are used for classification.
- By taking different subsets, classification of the new pattern is carried out.
- The classification made by the classifiers generated by using different feature subsets are combined to get the final classification of a new pattern.
- Let us consider the following set of patterns.

  Class 1 :

  $X_1 = (1, 2, 1); X_2 = (2, 2, 1); X_3 = (3, 2, 6);$
  $X_4 = (2, 1, 4); X_5 = (2, 3, 5);$

  Class 2 :

  $X_6 = (5, 6, 1); X_7 = (5, 5, 4); X_8 = (7, 5, 3);$
  $X_9 = (6, 6, 5); X_{10} = (5, 7, 2);$

  Let us consider a point P=(4,3,4)

  Classifier 1

  Let us take the first two features. Considering only the first two features in point P, P is closest to X_3 and is therefore classified as belonging to Class 1.

  Classifier 2

Let us take the first and third features. Considering only the first and third features, P is closest to X_7 and is therefore classified as belonging to Class 2.

Classifier 3

Let us consider the second and third features. Considering only the second and third feature, P is closest to point X_5 which belongs to Class 1.

P is therefore classified as belonging to Class 1.

In this case, since there are only three features, different combinations of two features are chosen. In general, a subset of the feature set is chosen for every classifier.

Varying the Output Classes by Combining Classes by Partitioning the Classes

- This technique is generally used if there are a large number of classes.
- The technique is called error-correcting output coding.
- At each trial, the classes are partitioned into two blocks resulting in two subsets of the data.
- A new pattern will get the new label l_1 if it is classified as belonging to the first block and the label l_2 if it is classified as belonging to the second block.
- For the next trial, the classes are again partitioned into two blocks, another partitioning being used here.
- This will result in two subsets of the data.The new pattern is again classified as belonging to Class l_1 or Class l_2.
- Everytime the new pattern is classified as belonging to a block, all the classes in that block will be given a vote.

- After a number of trials, the pattern is classified as belonging to the class which has the most number of votes.
- Let us consider an example. Consider the following set of patterns of 6 classes.

  Class 1 :

  $X_1 = (1,2,3); X_2 = (3,1,2); X_3 = (2,2,1); X_4 = (4,1,2)$

  Class 2 :

  $X_5 = (6,7,2); X_6 = (8,6,1); X_7 = (7,7,3); X_8 = (8,8,3)$

  Class 3 :

  $X_9 = (3,1,7); X_{10} = 1,1,8); X_{11} = (2,3,6); X_{12} = (2,2,7)$

  Class 4 :

  $X_{13} = (6,8,7); X_{14} = (8,6,6); X_{15} = (7,8,7); X_{16} = (8,8,8)$

  Class 5:

  $X_{17} = (1,7,2); X_{18} = (2,6,1); X_{19} = (3,8,2); X_{20} = (1,8,3)$

  Class 6 :

  $X_{21} = (6,1,3); X_{22} = (7,2,1); X_{23} = (8,2,2); X_{24} = (7,1,3)$

  Let us consider classification of the point P (4,4,4) using the nearest neighbour algorithm. If two points are equidistant from P, let us consider the pattern which appears earlier in the list of patterns shown above.

First partitioning :

Let the patterns of Classes 1,2 and 3 form one block and patterns of Class 4, 5 and 6 form the second block.

P is closest to pattern X_5 which appears in the first block. Therefore, the vote for each classifier is as follows :

C_1	C_2	C_3	C_4	C_5	C_6
1	1	1	0	0	0

Second Partitioning :

Let the patterns of Classes 1,3 and 5 be in one block and patterns of Class 2,4 and 6 be in the other block.

In this case, P is closest to pattern X_5 which appears in the first block. Therefore, the vote for Class 1,3 and 5 is incremented by one. Therefore, the current status of the vote for each class is :

C_1	C_2	C_3	C_4	C_5	C_6
2	1	2	0	1	0

Third Partitioning :

Let the patterns of Classes 2,5 and 6 be in one block and patterns of Class 1,3 and 4 be in the other block.

In this case, P is closest to a pattern in the second block. Therefore, the current status of the vote for each class is :

C_1	C_2	C_3	C_4	C_5	C_6
3	1	3	1	1	0

Fourth Partitioning :

Let the patterns of Class 1,2 and 6 be in one block and the patterns of Class 3,4 and 5 be in the other block.

In this case, P is closest to a pattern in the first block. Therefore, the current status of the vote for each class is :

C_1	C_2	C_3	C_4	C_5	C_6
4	2	3	1	1	1

If we stop at this point, we can see that Class 1 has the highest vote and therefore P is classified as belonging to Class 1.

Varying the Classifier by using Randomness

- The learning algorithm used in the classifier can be varied by using randomness in each of the algorithms considered earlier.
- One possibility is to change the decision tree by changing the order in which the features are chosen at the decision nodes. The change in the order of features will result in different classifiers.
- Other possibilities of creating randomness would include bootstrapping of the training data, injecting noise into the training data and creating a number of classifiers trained on a subset of the training examples.
- The Markov Chain Monte Carlo(MCMC) method can be used to decide which of the random classifiers generated is to be retained and which are to be discarded. This starts with a particular hypothesis. It then generates the next hypothesis. If the performance of this new hypothesis is good it is retained, otherwise it is discarded and another hypothesis is tried out. In this way, K classifiers are selected and used to classify a new pattern. These classifiers are combined by weighted vote depending on their likelihood values. Since each hypothesis depends only on the previous hypothesis and the entire process is run for

a discrete set of time values, it is a Markov chain created using randomness and hence the name Markov Chain Monte Carlo method.

To illustrate this, consider a training set being used for classification. Let us divide this training data into a training set and a validation set. Let us start with a classifier Cl_1. Suppose the validation set has 100 patterns. The classifier classifies 80 patterns correctly and misclassifies 20 patterns. Then the likelihood is 0.8 . Now using Cl_1 and injecting randomness, we get a classifier Cl_2. If this classifier classifies 90 patterns correctly and misclassifies 10 patterns of the validation set. This classifier has a likelihood of 0.9 which is satisfactory and therefore this classifier is retained. Injecting randomness into Cl_2, we get Cl_3. If this classifier classifies 60 patterns of the validation set correctly and misclassifies 40 patterns, the likelihood is 0.6 . This is not satisfactory and Cl_3 is discarded and another classifier is generated by infusing randomness into Cl_2. In this way a set of K classifiers are generated which give satisfactory likelihood values. These classifiers are combined by giving a weightage of 0.8 to the first classifier, 0.9 to the second classifier and so on.