MODULE 16
Combination of Classifiers

LESSON 37
AdaBoost for Classification

Keywords: Weak Learners, Boosting, Weights

ADABOOST

- *Boosting* refers to combining the class labels obtained by using weak classifiers.
- A weak classifier is a classifier formed easily which may be rough and approximate. Usually some rule-of-thumb is used to derive the weak classifier.
- A number of weak classifiers are used to classify patterns.
- This is done for a number of different classifiers. After this, the weak hypotheses are combined together to form a single classification rule.
- A popular algorithm based on boosting is the *ADABOOST* algorithm.
- The ADABOOST algorithm has a probability distribution p_t over the training set.
- The algorithm is run for a number of iterations where in each iteration, a subset of the training set is drawn with replacement according to the probability distribution $p_t(\mathrm{x})$.
- The classifier is used on this subset.
- The error in classification is used to adjust the probability distribution $p_t(\mathrm{x})$.
- The probability distribution is computed by adjusting the weights associated with every pattern wt_i.
- This works by giving more weightage for patterns which are classified wrongly by the classifier and giving less weightage to patterns which are classified correctly.
- Depending on the performance for the distribution p_t, each classifier is weighted.
- The classification of a new pattern is done by a weighted vote of individual classifiers.

- In the ADABOOST algorithm, a probability distribution p_t(x) is maintained over the training set.
- In every iteration, a subset of the training data is taken by using sampling with replacement on the training dataset.
- A weak classifier is used to classify the training set. One example of a weak classifier is a decision stump. A decision stump is a decision tree with a depth of one.
- The error in the performance of the classifier is used to adjust the probability distribution of the training set. This is done by normalizing a set of weights, where each pattern in the training set is associated with a weight. The updation of the weights is done by placing more weight on training examples which are misclassified by the classifier used and less weight on examples which are classified correctly.
- Each classifier is weighted according to its performance for the distribution that it was trained on.
- A new pattern is classified by using a weighted vote of each of the classifiers.
- The detailed algorithm is described below.

Input : Training set T, number of iterations *it*.

1. Set T'=T ; Set wt[i]=1,i=1..n; m=n;
2. For i= 1 to it {
3. Choose the classifier to be used
4. $$\alpha_i = \frac{1}{m} \sum_{x_j \in T' : C_i(x_j) \neq y_j} wt(x)$$
5. if $\alpha > \frac{1}{2}$, set T' to a bootstrap sample from T with weight 1 for every instance and go to Step 3

6. $\beta_i = \frac{\alpha_i}{1-\alpha_i}$

7. For each x_j in T', if $C_i(x_j) = y_j$ then wt(x_j)=wt(x_j).β_i
8. Normalize the weights of all the patterns so that the total weight of T' is m.
9. }
10.

- As an example, consider the data shown in Figure 1.

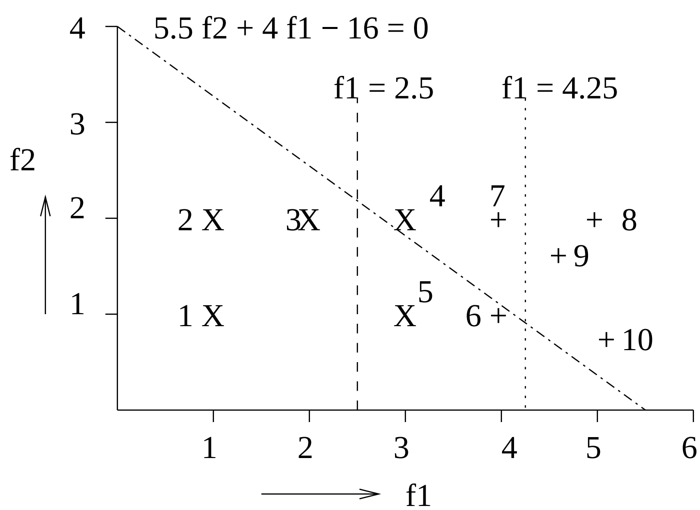

Figure 1: Two class problem

Let us consider 3 classifiers.

The first classifier:

Consider the line $x = 2.5$. The points to the left of this line belong to Class 1 and points to the right of this line belong to Class 2.

The second classifier:

Consider the line $x = 4.25$. The points to the left of this line belong to Class 1 and points to the right of this line belong to Class 2.

The third classifier:

Consider the line $5.5y + 4x - 16 = 0$. The points to the left of this line belong to Class 1 and points to the right of this line belong to Class 2.

Let us consider one classifier at a time. Using the first classifier

If the x-coordinate of the point is less than 4.25, classify the point as belonging to Class 1 otherwise classify the point as belonging to Class 2.

Let us set the weights of all the points to be 1. So we get $wt[i] = 1, i = 1, .., 10$.

Let us consider the first classifier. This classifier misclassified the points X_4 and X_5 of Class 1. So,

$\alpha_1 = \frac{2}{10} = 0.8$

$\beta_1 = \frac{0.2}{1-0.2} = \frac{0.2}{0.8} = 0.25$

The weights of all the points except X_4 and X_5 is carried out. So weight of X_1 is

$wt(1) = wt(1) * \beta_1 = 1.0$ * 0.25 0= 0.25
Similarly,

$wt(2) = 0.25; wt(3) = 0.25; wt(6) = 0.25; wt(7) = 0.25$
$wt(8) = 0.25; wt(9) = 0.25; wt(10) = 0.25$

The weights of X_4 and X_5 remain unchanged.

$wt(4) = 1.0; wt(5) = 1.0$

These weights should be normalized so that they sum to 10.

This can be done by using the formula

$$w_i = \frac{w_i}{\sum_i w_i} * 10$$

So we get

$wt(1) = \frac{0.25}{4} * 10 = 0.625$
Similarly, we get

$wt(2) = 0.625; wt(3) = 0.625; wt(6) = 0.625; wt(7) = 0.625$
$wt(8) = 0.625; wt(9) = 0.625; wt(10) = 0.625$

$wt(4) = \frac{1}{4} * 10 = 2.5$
$wt(5) = \frac{1}{4} * 10 = 2.5$

Now let us consider the second classifier

This misclassifies patterns X_6 and X_7.

So

$\alpha_2 = \frac{1}{10}(0.625 + 0.625) = 0.125$

$\beta_2 = \frac{0.125}{1-0.125} = 0.1429$

The weights are

$wt(1) = 0.625 * 0.1429 = 0.0893$

$wt(2) = 0.0893; wt(3) = 0.0893$

$wt(4) = 0.3573; wt(5) = 0.3573$
$wt(8) = 0.0893; wt(9) = 0.0893; wt(10) = 0.0893$
$wt(6) = 0.625; wt(7) = 0.625$ (unchanged)

The sum of these weights is 2.5004.

Normalizing these weights, we get

$wt(1) = \frac{0.0893}{2.5004} * 10 = 0.3571$

$wt(2) = 0.3571; wt(3) = 0.3571$
$wt(4) = 1.429; wt(5) = 1.429$
$wt(8) = 0.3571; wt(9) = 0.3571; wt(10) = 0.3571$
$wt(6) = 2.5; wt(7) = 2.5$

We now consider the third classifier.
This classifier misclassified patterns X_4 and X_6.

$\alpha_3 = \frac{1}{10}(1.429 + 2.5) = 0.3929$

$\beta_3 = \frac{0.3929}{1-0.3929} = 0.6472$

The weights are

$wt(1) = 0.3571 * 0.6472 = 0.2311$
$wt(2) = 0.2311; wt(3) = 0.2311$
$wt(5) = 0.9248$

$wt(7) = 1.618$
$wt(8) = 0.2311; wt(9) = 0.2311; wt(10) = 0.2311$
The weights of X_4 and X_6 remain unchanged.

$wt(4) = 1.429; wt(6) = 2.5$
Total weight is 7.8584

Normalizing the weights, we get

$wt(1) = \frac{2.311}{7.8584} * 10$=0.2941
$wt(2) = 0.2941; wt(3) = 0.2941$
$wt(5) = 1.1768$
$wt(7) = 2.0589$
$wt(8) = 0.2941; wt(9) = 0.2941; wt(10) = 0.2941$
$wt(4) = 1.8184; wt(6) = 3.1813$

Consider a point (3,2). According to the first classifier, it is classified as belonging to Class 2. According to the second classifier, it is classified as belonging to Class 1. According to the third classifier, it is classified as belonging to Class 2.

For Class 1, we get

$log\frac{1}{\beta_2} = log\frac{1}{0.1429}$=0.843

For Class 2, we get

$\Sigma log\frac{1}{\beta_i} = log\frac{1}{0.25} + log\frac{1}{0.6472}$=0.791

The pattern (3,2) is therefore classified as belonging to Class 1.