MODULE 12
Linear Discriminant Functions

LESSON 25

Introduction to Discriminant Functions

Keywords: Linear Classifier, Discriminant, Separability, Positive and Negative Half-Spaces

Linear Discriminant Functions

- One popular way of separating patterns belonging to two classes is by using a simple decision boundary.
- We illustrate this using the collection of two-dimensional patterns shown in Figure 1.

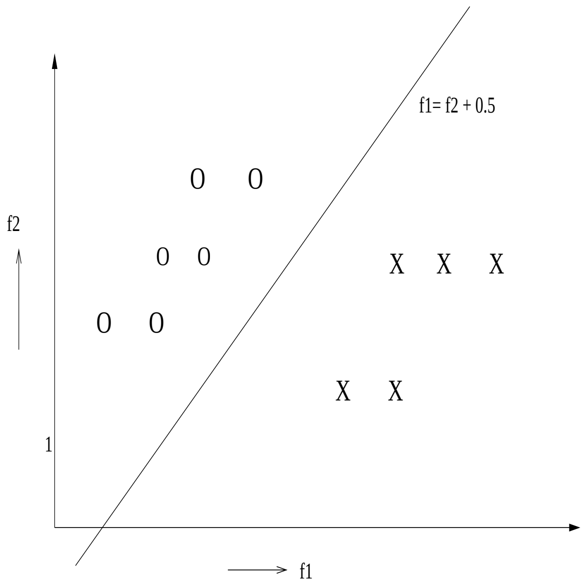

Figure 1: Classification using a Linear Discriminant Function

- **Two-dimensional example data**

  Example 1

  There are **five** patterns from one class (labelled 'X') and **six** patterns from the second class (labelled 'O'). This set of labelled patterns may be described using Table 1.

- **Linear separability**
  The first six patterns are from class 'O' and the remaining patterns

Pattern No.	$feature_1$	$feature_2$	Class
1	0.5	1.5	'O'
2	1.0	1.5	'O'
3	1.0	2.0	'O'
4	1.5	2.0	'O'
5	1.5	2.5	'O'
6	2.0	2.5	'O'
7	3.0	1.0	'X'
8	3.5	1.0	'X'
9	3.5	2.0	'X'
10	4.0	2.0	'X'
11	4.5	2.0	'X'

Table 1: Description of the patterns

are from class 'X'. Consider the line represented by $f_1 = f_2 + 0.5$ shown in Figure 1. All the patterns labelled 'O' are to the left of the line and patterns labelled 'X' are on the right side. In other words, patterns corresponding to these two classes are **linearly separable** in this example as patterns belonging to each of the classes fall on only one side of the line.

- **Positive and negative half spaces**
  One more way of separating the patterns of the two classes is by noting that all the patterns of class 'O' satisfy the property that $f_1 < f_2 + 0.5$ or equivalently f_1 - f_2 is less than 0.5. For example, pattern 1 in Table 1 has a value of 1.5 for $feature_2$ and 0.5 for $feature_1$ and so this property is satisfied (because 0.5 - 1.5 = -1.0 ($<$ 0.5)). Similarly, the sixth pattern satisfies this property because f_1 - f_2 is -0.5. In a symmetric manner, all the patterns belonging to class 'X' have the value of f_1 to be larger than that of $f_2 + 0.5$; equivalently, f_1 - $f_2 > 0.5$. For example, for the seventh pattern in the table, f_1 - f_2 is 2.0 (3.0 - 1.0) which is greater than 0.5. Based on the location of the patterns with respect to the line, we say that patterns of class 'X' are on the positive side of the line (because $f_1 - f_2 > 0.5$ for these patterns) and patterns of class 'O' are on the negative side (equivalently $f_1 - f_2 < 0.5$ for patterns labelled

'O''). Here, the line separates the two-dimensional space into two parts which can be called as **positive half space** (where patterns from class 'X' are located) and **negative half space** (which has patterns from class 'O'').

- **Variety of separating lines**
  It is easy to observe that there are possibly infinite ways of realizing the decision boundary, a line in this two-dimensional case, which can separate patterns belonging to the two classes. For example, the line $f_1 = f_2$ also separates the first six points in Table 1 from the remaining 5 points.

- **Functional form of the linear decision boundary**
  It is convenient to abstract such lines using the following functional form:

$$f(X) \quad = w_1 f_1 + w_2 f_2 + b = 0 \tag{1}$$

  Correspondingly, the line f_1 - f_2 = 0 has w_1 = 1; w_2 = -1; and b = 0. Similarly, w_1 = 1; w_2 = -1; and b = -0.5 for f_1 - f_2 = 0.5. This representation permits us to deal with patterns and decision boundaries in the multi-dimensional space. For example, in a d-dimensional space, the decision boundary is a hyperplane and it can be represented by

$$f(X) \quad = w^t X + b = 0 \tag{2}$$

  where w and X are d-dimensional vectors.

- **Separation of patterns based on w and b**
  We can use this representation to characterize linear separability. We say that two classes, say classes labelled 'X' and 'O' are linearly separable, if we can find a weight vector w and a scalar b such that

  $w^t X + b > 0$ for all patterns X belonging to one class (say class 'X') and

  $w^t X + b < 0$ for all the patterns belonging to the other class (that is class 'O').

- **Non-linear decision boundary**
  Another possibility is to use a non-linear decision boundary to characterize the separation between the two classes. For example, the non-linear decision boundary

$$f_1 \;=\; -\, f_2^2 \;+\; 5 f_2 \;-\; 3$$

  is depicted in Figure 2.

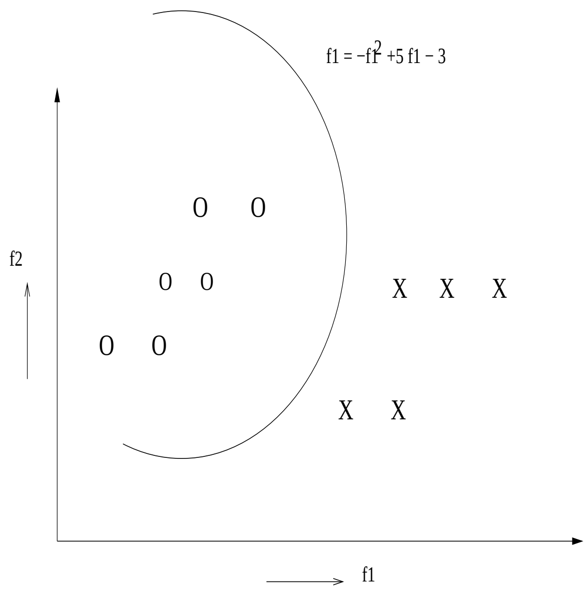

Figure 2: Classification using a Nonlinear Discriminant Function

- **Location of the decision boundary**
  Here, the value of b plays a role in deciding the location of the decision boundary. Note that the decision boundary between the two classes is characterized by $w^t X + b = 0$. Based on the location of the origin (the zero vector), we can make the following observations with respect to the value of b.

  1. Consider the situation where $b = 0$. In such a case, the origin lies on the decision boundary. This is because $w^t X + b = 0$ for $X = 0$ (origin) and $b = 0$, for any value of w.

2. When the value of b is negative, that is when $b < 0$, the origin lies in the negative side; this can be observed from the fact that $w^t X + b < 0$ when $X = 0$ and $b < 0$, for any value of w.

3. The above observations hold in general in a d-dimensional space. We examine them with respect to the two-dimensional data set ($d = 2$) shown in the example below.

   Example 2

   Consider the data shown in Figure 1 and Table 1. Let us consider the line $f_1 = f_2$; here, $w^t = (1,-1)$ and $b = 0$. The origin ((0,0)) is on the line $f_1 = f_2$ which is the decision boundary.

   Now consider the line $f_1 = f_2 + 0.5$; this can also be written as $2f_1 - 2f_2 - 1 = 0$. This line is obtained by shifting $f_1 - f_2 = 0$ down appropriately. Note that the points (0.5,0) and (0,-0.5) are on this line. Also, all the points labelled 'X' are on one side (positive side) and those labelled 'O' are located on the negative side of the line. Note that the origin is in the negative part.

- **Solving for the vector w**
  Now consider the decision boundary charcaterized by

$$f(X) \quad = w_1 f_1 + w_2 f_2 + b = 0$$

As noted earlier, (0.5,0) and (0,-0.5) are on the decision boundary. We can get the values of w_1 and w_2 by solving the equations obtained by substituting these points in the equation of the decision boundary.

By substituting (0.5,0), we get

$$0.5w_1 + 0.0w_2 + b = 0 \Rightarrow 0.5w_1 + b = 0 \tag{3}$$

Similarly, by substituting (0,-0.5), we get

$$0.0w_1 - 0.5w_2 + b = 0 \Rightarrow -0.5w_2 + b = 0 \tag{4}$$

By subtracting (4) from (3), we get

$$0.5w_1 + 0.5w_2 = 0 \Rightarrow w_1 + w_2 = 0 \Rightarrow w_1 = -w_2 \tag{5}$$

- **The role of w**
  One possible instantiation is $w_1 = 1\,; w_2 = -1$; so, the value of b is -0.5 from both (3) and (4). The other possibility is $w_1 = -1\,; w_2 = 1$; so, the value of b is 0.5. Out of these two possibilities, only the first is acceptable; the second is not. We can verify this by considering a positive example from Table 1. Let us consider the example $(3,1)^t$. In this case, we need $w^tX + b > 0$; that is by using the first set of values, we have

  $$1 * f_1 - 1 * f_2 - 0.5 = 3 - 1 - 0.5 = 1.5 > 0$$

  Now, by using the second set of values, we have an inconsistency as

  $$-1 * X_1 + 1 * X_2 + 0.5 = -3 + 1 + 0.5 = -1.5 < 0$$

- **w is orthogonal to the decision boundary**
  So, we accept the first set of values; this means that the vector w is orthogonal to the decision boundary as exemplified by equations (3), (4), and (5). In general, if p and q are two points on the decision boundary, then we have

  $$w^tp + b = 0 = w^tq + b \Rightarrow w^t(p-q) = 0$$

  So, the vectors w and $(p-q)$ are orthogonal to each other; observe that $(p-q)$ characterizes the direction of the decision boundary.

- **w points towards the positive half space**
  Further w points towards the positive half space. This can be verified by noting that the angle, θ, between w and any positive pattern vector is such that $-90 < \theta < 90$. As a consequence $cos(\theta)$ is positive.