MODULE 12
Linear Discriminant Functions

LESSON 26
Characterization of the Decision Boundary

Keywords: Weight Vector, Threshold, Location, Orientation

Linear Discriminant Functions

- **Cosine of the angle between w and any pattern vector**
  For example, if α is the angle between $w = (1, -1)^t$ and $X = (3, 1)^t$, a vector in the positive half space (refer to Table 1 of lesson 25), the cosine of the angle between these vectors is given by

$$cos(\alpha) = \frac{w^t X}{|| w |||| X ||} = \frac{2}{\sqrt{10}\sqrt{2}} = \frac{1}{\sqrt{5}} > 0$$

  Similarly, if β is the angle between w and a vector in the negative half space, for example $(1, 2)^t$ (refer to Table 1 of lesson 25), then

$$cos(\beta) = \frac{-1}{\sqrt{2}\sqrt{5}} = \frac{-1}{\sqrt{10}} < 0$$

- So, the value of b *characterizes the location of the decision boundary.* Similarly, the weight vector w *decides the orientation of the decision boundary.*

- **Shortest distance between a point and the boundary**
  Another useful notion is the distance between a point X and the decision boundary. The shortest distance between a point X and a point X_p on the decision boundary $f(X_p) = w^t X_p + b = 0$ is given by
  Minimize $|| X_p - X ||^2$ (we consider the squared Euclidean norm for the sake of simplicity in calculus)
  such that $w^t X_p + b = 0$.

- **Constrained optimization**
  This constrained optimization problem can be solved by minimizing the corresponding Lagrangian $L(X_p, \lambda)$ given by

$$L(X_p, \lambda) = || X_p - X ||^2 + \lambda(w^t X_p + b) \tag{1}$$

  By taking the gradient of the Lagrangian with respect to X_p and λ and equating to zero (0), we have

$$2(X_p - X) + \lambda w = 0 \tag{2}$$
$$w^t X_p + b = 0 \tag{3}$$

from equation (2), we get

$$X_p \;=\; X \;-\; \frac{\lambda}{2} w \tag{4}$$

Also, we can get

$$\begin{aligned} \lambda w \;&=\; -\,2\,(X_p \;-\; X) \Rightarrow \\ \lambda w^t w \;&=\; -\,2 w^t X_p \;+\; 2 w^t X \Rightarrow \\ \lambda w^t w \;&=\; -\,2 w^t X_p \;+\; 2 w^t X \;-\; 2b \;+\; 2b \end{aligned}$$

In the above, the second equation is obtained by premultiplying with w^t; the third equation is obtained by subtracting and adding $2b$ from the right hand side of the second equation. Noting that $w^t X_p + b = 0$ (from (3)) and $f(X) = w^t X + b$, we get

$$\lambda \;=\; \frac{2 f(X)}{||\; w\; ||^2}$$

By substituting this value of λ in equation (4), we get

$$X_p \;=\; X \;-\; \frac{f(X) w}{||\; w\; ||^2} \tag{5}$$

- **Representation of X in terms of two components**
  Any point X can be written as a sum of two vectors; one vector along the decision boundary and another orthogonal to it. So,

  $$X = X_p + X_o \tag{6}$$

  where X_p is the projection of X along the decision boundary and X_o is the orthogonal component. Further, we know that w is also orthogonal to the decision boundary and is oriented towards the positive half space.

- **Distance, d_n, of X from the decision boundary**
  So, X_o is of the form $d_n \frac{w}{||w||}$ where d_n is a real number and it is positive if X is from class 'X'' and negative if it is from class 'O''. Based on these observations and equation (6), we have

  $$X_o \;=\; d_n \frac{w}{||\; w\; ||} \;=\; \frac{f(X) w}{||\; w\; ||^2}$$

From the above equation we have,

$$d_n = \frac{f(X)}{\| w \|} = \frac{w^t X + b}{\| w \|} \tag{7}$$

Example 1

Consider the data shown in Figure 1 and the decision boundary given by the line f_1 - f_2 - 0.5 = 0. Note that in this case, $\| w \|^2$ is 2 and b = - 0.5. So, distance from the origin ($X = (0,0)$) is $\frac{-0.5}{\sqrt{2}}$.

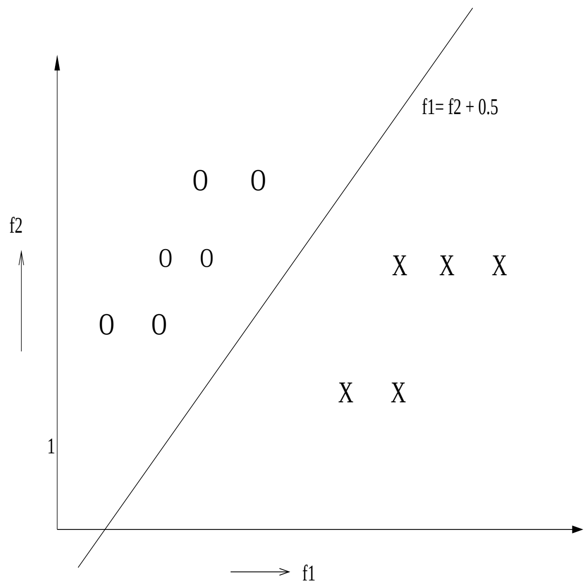

Figure 1: Classification using a Linear Discriminant Function

Similarly, distance from the point (1,1) is also $\frac{-0.5}{\sqrt{2}}$. In general, it can be shown that the distance from any point of the form (γ, γ) where γ is a real number is $\frac{-0.5}{\sqrt{2}}$.

- **Homogeneous form**
  We can transform the patterns from the d-dimensional space to $d+1$-dimensional space by adding 1 as the $d+1^{th}$ component to each pattern,

that is $f_{d+1} = 1$, and by considering $w_{d+1} = \text{b}$. This mapping permits us to represent the decision boundary in the **homogeneous** form as

$$f(X') \quad = z^t X' = 0 \tag{8}$$

where z and X' are $d+1$-dimensional vectors; z is obtained by adding b as the $d+1^{th}$ value to w and X' is obtained by adding to X, a value of 1 as the $d+1^{th}$ entry. So, z and X' are given as follows:

$$\begin{aligned} z^t &= (w_1, w_2, \cdots, w_d, b) \\ X'^t &= (f_1, f_2, \cdots, f_d, 1) \end{aligned}$$

- **Conversion into homogeneous form**
  Using this representation, the line $f_1 = f_2 + 0.5$ can be represented as $z^t X' = 0$, where $z^t = (1, -1, -0.5)$ and $X'^t = (f_1, f_2, 1)$. From now onwards, we use w and X instead of z and X' respectively. The distinction between the homogeneous case and non-homogeneous case would be obvious from the context.

Example 2

Consider the two-class patterns in two dimensions shown in Table 1.

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

Table 1: Description of the patterns

- **Classification of patterns based on homogeneous function**
  The corresponding patterns, in the three dimensional space, after transforming into the homogeneous form are shown in Table 2. Now consider the vector w in the three-dimensional space, which is given by

Pattern No.	$feature_1$	$feature_2$	$feature_3$	Class
1	1.0	1.5	1.0	'O'
2	1.5	2.0	1.0	'O'
3	1.5	2.5	1.0	'O'
4	2.0	2.5	1.0	'O'
5	3.0	1.0	1.0	'X'
6	3.5	2.0	1.0	'X'
7	4.0	2.0	1.0	'X'

Table 2: Description of the patterns in three dimensions

$w = (1, -1, -0.5)^t$ and a pattern from class 'O'', say the first pattern, P_1, given by $(1.0, 1.5, 1.0)^t$. Note that

$$w^t P_1 = 1 * 1.0 - 1 * 1.5 - 0.5 * 1.0 = -1.0 < 0$$

Note that for all the patterns from class 'O'', shown in Table 2, $w^t X < 0$. Similarly, for any pattern from class 'X'', say pattern numbered 5, $P_5 = (3.0, 1.0, 1.0)^t$, we have

$$w^t P_5 = 1 * 3.0 - 1 * 1.0 - 0.5 * 1.0 = 1.5 > 0$$

It can be observed that for every pattern from class 'X'', $w^t X > 0$.

- **Converting the negative patterns to fall in the positive space**
  In the binary classification problem, it is possible to affect normalization so that all the patterns lie on the positive side of the line in the two-dimensional case and hyperplane in the general d-dimensional case. This is achieved by replacing every feature value of each pattern from class 'X'' by its negation including the $d + 1^{th}$ entry. For example, the first pattern in Table 1, $(1.0, 1.5)^t$ in the homogeneous form is $(1.0, 1.5, 1)^t$ and after converting to fall in the positive space is $(-1.0, -1.5, -1)$. So, by transforming and normalizing the data in Table 1 in this manner, we get the data shown in Table 3.

- $w^t X > 0$ **for every normalized pattern**
  Note that, using the homogeneous w, $(1, -1, -0.5)^t$, every normalized pattern, X, shown in Table 3 satisfies the property that $w^t X >$

Pattern No.	$feature_1$	$feature_2$	$feature_3$
1	- 1.0	- 1.5	- 1.0
2	- 1.5	- 2.0	- 1.0
3	- 1.5	- 2.5	- 1.0
4	- 2.0	- 2.5	-1.0
5	3.0	1.0	1.0
6	3.5	2.0	1.0
7	4.0	2.0	1.0

Table 3: Description of the normalized patterns

0. For example, for the first pattern after normalization, given by $(-0.5, -1.5, -1.0)^t$, we have

$$w^t X \;=\; (1, -1, -0.5)(-0.5, -1.5. - 1.0)^t \;=\; 1.5 \;>\; 0$$

Similarly, for the fifth pattern given by $(3.0, 1.0, 1.0)^t$, we have

$$w^t X \;=\; (1, -1, -0.5)(3.0, 1.0.1.0)^t \;=\; 1.5 \;>\; 0$$