Discriminant Functions I

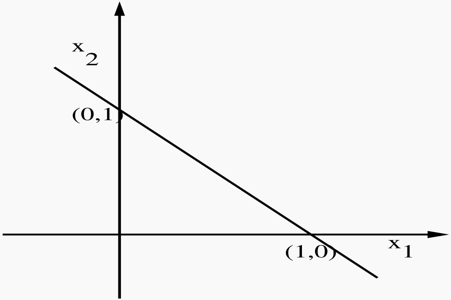
Consider a two class problem. We need a function for the decision boundary to separate the classes.
Consider a simple linear discriminant function:

$$w_2 x_2 + w_1 x_1 + w_0 = 0$$

defined over a two dimensional space, where
$\bar{w} = \begin{bmatrix} w_1 \\ w_2 \end{bmatrix}$ $\bar{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$
An example line:
$x_2 + x_1 - 1 = 0$
Clearly points about the line yield
$x_2 + x_1 - 1 > 0$
while points below the line yield
$x_2 + x_1 - 1 < 0$

Discriminant Functions II

If the points belonging to the two classes C_1 and C_2 are as shown in the Figure, they can be easily discriminated.
In general linear discriminant functions are of the form
$w_d x_d + w_{d-1} x_{d-1} + \dots + w_1 x_1 + w_0 = 0$
This represents a hyperplane in a $d-$dimensional space.
Alternatively can be written as
$\bar{w}^t \bar{x} + w_0 = 0$



Non linear discriminant functions I

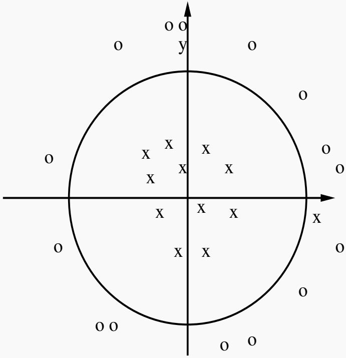
Consider two classes being separated by a circle as shown in the Figure.
$x^2 + y^2 = r^2$
$x^2 + y^2 - r^2 = 0$ is the boundary between the two classes.
Clearly $x^2 + y^2 - r^2 < 0$ inside the circle
$x^2 + y^2 - r^{>}02$ outside the circle
The boundary is clearly nonlinear in the input space.
Consider the transformation:

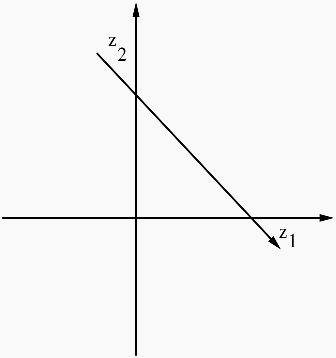
$$\begin{aligned} z_1 &= x^2 \\ z_2 &= y_2 \\ r^2 &= 1 \end{aligned}$$

This yields

Non linear discriminant functions II

$z_1 + z_2 - 1 = 0$
This leads to a linear hyperplane in $z-$space that is isomorphic to the input $x-$space.





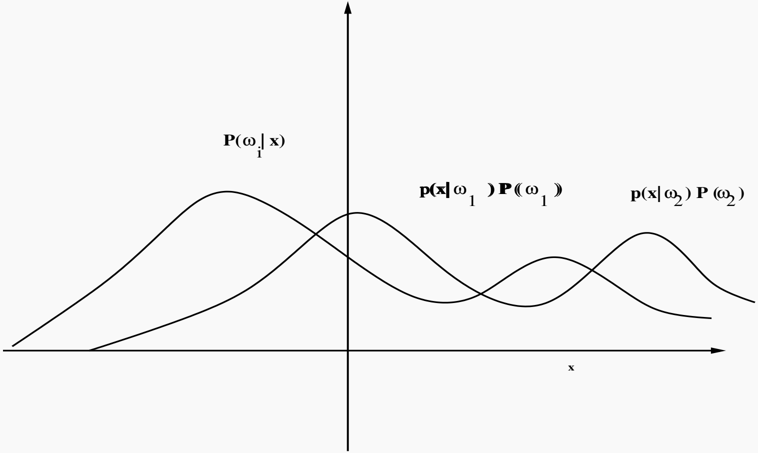
$$p(\vec{x}/\omega_i) = \mathcal{N}(\vec{x}/\vec{\mu}_i, \Sigma_i) \qquad (\text{where } \vec{x} = [x_1\ x_2 \ldots x_d]^T)$$

Classification steps:

- Training Process: We estimate $\hat{\mu}_i$ and $\hat{\Sigma}_i$ using the dataset of i^{th} class: $\mathrm{D}(\vec{x_1}\ \vec{x_2} \ldots \vec{x_N})$
- Development Process: We fix our hyperparameters in this process.
- Testing Process: We test our model using unseen data.

We classify the feature vector $\vec{x}$ to the class for which $P(\omega_i/\vec{x})$ is the highest and the rest becomes the error. Example: If we have M classes and max is the answer, then error $= 1 - P(\omega_{max}/\vec{x})$

Bayes' decision Theory



As $p(\mathbf{x})$ does not affect the decision process,
$P(\omega_i|\mathbf{x}) \sim p(\mathbf{x}|\omega_i)P(\omega_i)$

Unimodal Multivariate Gaussian Distribution

$$p(\vec{x}/\omega_i) = \frac{1}{(\sqrt{2\pi})^d \ \left|\Sigma_i\right|^{\frac{1}{2}}} e^{-\frac{1}{2}(\vec{x}-\vec{\mu_i})\Sigma_i^{-1}(\vec{x}-\vec{\mu_i})^T}$$

where,

$$\Sigma_i = E[(\vec{x}-\vec{\mu_i})(\vec{x}-\vec{\mu_i})^T]$$
$$\mu_i = E[\vec{x_i}]$$

$$\ln g_1(\vec{x}) = \ln p(\vec{x}/\omega_1)$$
$$= -\frac{1}{2}\ln(\sqrt{2\pi}) - \frac{d}{2}\ln\left|C_1\right|^{\frac{1}{2}} - \frac{1}{2}(\vec{x}-\vec{\mu})^T C_1^{-1}(\vec{x}-\vec{\mu}) + \ln P(w_1)$$

Similarly for $g_2(x) = p(\vec{x}/\omega_2)$

Discriminating Function: $g(\vec{x}) = \ln g_1(\vec{x}) - \ln g_2(\vec{x})$

CASE-1: $\mathbf{C_1 = C_2 = \sigma^2 I}$ (less parameters $\Longrightarrow$ less data reqd.)

The quadratic term $(x^T x)$ is same in both $g_1(\vec{x})$ and $g_2(\vec{x})$, so we can assume linear equation. Hence,

$$g_i(\vec{x}) = \vec{\omega_i}^t \vec{x} + \omega_{io} \quad \text{(assumed)}$$

Now, neglecting terms which gets cancelled out in $\ln g_1(\vec{x}) - \ln g_2(\vec{x})$, we get:

$$g_1(\vec{x}) = \frac{-1}{2\sigma^2}(\vec{\mu_1}\vec{\mu_1}^T - 2\vec{\mu_1}^T \vec{x}) + \ln P(\omega_1)$$

Now comparing our assumed equation and this equation, we get:

$$\omega_i = \frac{\vec{\mu_i}}{\sigma_i^2} \text{ and } \omega_{io} = \frac{-1}{2\sigma^2}\vec{\mu_1}\vec{\mu_1}^T + \ln P(\omega_i)$$

Decision Boundary: $g(\vec{x}) = \vec{\omega_1}^T \vec{x} + \omega_{1o} - \vec{\omega_2}^T \vec{x} - \omega_{2o} = 0$

So, $g(\vec{x}) = \vec{\omega}^T \vec{x} + \omega_o$ (because Straight Line)

where,

$$\vec{\omega} = \vec{\omega_1} - \vec{\omega_2} = \frac{1}{\sigma^2}(\mu_1 - \mu_2)^t \vec{x}$$

$$\omega_o = \omega_{1o} - \omega_{2o} = \frac{-1}{2\sigma^2}(\vec{\mu_1}\vec{\mu_1}^T - \vec{\mu_2}\vec{\mu_2}^T) + \ln \frac{P(\omega_1)}{P(\omega_2)}$$

Now since,

$$\vec{\mu_1}\vec{\mu_1}^T - \vec{\mu_2}\vec{\mu_2}^T = \|\vec{\mu_1}\|^2 - \|\vec{\mu_2}\|^2 = (\vec{\mu_1} - \vec{\mu_2})^t(\vec{\mu_1} + \vec{\mu_2})$$

$$\begin{aligned} g(\vec{x}) &= \frac{1}{\sigma^2}(\vec{\mu_1} - \vec{\mu_2})^t \vec{x} - \frac{1}{2\sigma^2}\left[(\vec{\mu_1} - \vec{\mu_2})^t(\vec{\mu_1} + \vec{\mu_2})\right] + \ln \frac{P(\omega_1)}{P(\omega_2)} \\ &= \frac{1}{\sigma^2}(\vec{\mu_1} - \vec{\mu_2})^t \left[\vec{x} - \frac{1}{2}(\vec{\mu_1} + \vec{\mu_2}) + \frac{\sigma^2(\vec{\mu_1} - \vec{\mu_2})}{\|\vec{\mu_1} - \vec{\mu_2}\|^2} \ln \frac{P(\omega_1)}{P(\omega_2)}\right] \\ &= \vec{\omega}^t(\vec{x} - x_o) = 0 \quad \text{(i.e The separating plane passes through } x_o) \end{aligned}$$

Now, if $P(\omega_1) = P(\omega_2)$ then the boundary perpendicularly bisects the line joining $\vec{\mu_1}$ and $\vec{\mu_2}$

CASE-2: $\mathbf{C_1} = \mathbf{C_2} = \mathbf{C} \begin{bmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{bmatrix} \qquad (\sigma_{jk} = 0 \text{ for } j \neq k)$

$$
\begin{aligned}
g_i(x) &= \frac{-1}{2}(\vec{x} - \vec{\mu_i})^t C_i (\vec{x} - \vec{\mu_i}) + \ln P(w_i) \\
&= \frac{-1}{2}\vec{x}^t C^{-1}\vec{x} + \frac{1}{2}\vec{\mu_i^t} C^{-1}\vec{x} + \frac{1}{2}\vec{x}^t C^{-1}\vec{\mu_i} - \frac{1}{2}\vec{\mu_i}^t C^{-1}\vec{\mu_i} + \ln P(\omega_i)
\end{aligned}
$$

Ignoring the terms that do not depend on i i.e they will cancel out.

$$
\begin{aligned}
g_i(\vec{x}) &= (C^{-1}\mu_i)^t \vec{x} - \frac{1}{2}\vec{\mu_i}^t C^{-1}\vec{\mu_i} + \ln P(\omega_i) \\
&= \vec{\omega_i}^t \vec{x} + \omega_{io}
\end{aligned}
$$

Now, the discriminating boundary can be given by:

$$
\begin{aligned}
g(x) &= (C^{-1}\vec{\mu_1} - C^{-1}\vec{\mu_2})\vec{x} - \vec{\mu_1} C^{-1}\vec{\mu_1} + \vec{\mu_2} C^{-1}\vec{\mu_2} + \ln \frac{P(\omega_1)}{P(\omega_2)} \\
&= C^{-1}(\vec{\mu_1} - \vec{\mu_2})\vec{x} - \frac{C^{-1}}{2}(\vec{\mu_1} - \vec{\mu_2})^t(\vec{\mu_1} + \vec{\mu_2}) + \ln \frac{P(\omega_1)}{P(\omega_2)}
\end{aligned}
$$

On comparing with equation of plane $g(x) = \vec{\omega}^t(\vec{x} - x_o)$

$$\vec{\omega} = C^{-1}(\vec{\mu_1} - \vec{\mu_2})$$
$$x_o = \frac{1}{2}(\vec{\mu_1} + \vec{\mu_2}) - \frac{\ln P(\omega_1)/P(\omega_2)}{(\vec{\mu_1} - \vec{\mu_2})^t C^{-1}}$$

Notice that C^{-1} will apply an affine rotation on $(\vec{\mu_1} - \vec{\mu_2})$ so $\vec{\omega}$ will not be in direction of $(\vec{\mu_1} - \vec{\mu_2})$. Also, if priors are equal, $x_o = \frac{1}{2}(\vec{\mu_1} + \vec{\mu_2})$ i.e it still passes through the midpoint but the direction is transformed.

Note that the contours have same probability density in a Gaussian because it is symmetric about the mean.

CASE-3: $\mathbf{C_1 \neq C_2}$ (They can be diagonal)

Again, neglecting the terms that don't affect.

$$g_i(\vec{x}) = \frac{1}{2}\vec{x}^t C_i^{-1}\vec{x} + \vec{\mu_i}^t C_i^{-1}\vec{x} + \vec{\mu_i}^t C_i^{-1}\vec{\mu_i} + \ln P(\omega_i) - \frac{1}{2}\ln|C_i|$$

$$g(\vec{x}) = \vec{x}^t W\vec{x} + \vec{\omega}^t\vec{x} + \omega_o = 0 \quad \text{where,}$$

$$W = \frac{-1}{2}(C_1^{-1} - C_2^{-1})$$

$$\omega = (C_1^{-1}\vec{\mu_1} - C_2^{-1}\vec{\mu_2})$$

$$\omega_o = \frac{-1}{2}(\vec{\mu_2}^t C_2^{-1}\vec{\mu_2} + \vec{\mu_1}^t C_1^{-1}\vec{\mu_1}) - \frac{1}{2}\ln|\frac{C_1}{C_2}| + \ln\frac{P(\omega_1)}{P(\omega_2)}$$

Summary of footprint of density function

Covariance	2D	3D	nD	EigenVectors parallel to axis?
$C = \sigma^2 I$	Circle	Sphere	Hypersphere	Yes
C = Diagonal	Ellipse	Ellipsoid	Hyperellipsoid	Yes
C = Full	Ellipse	Ellipsoid	Hyperellipsoid	No

An extended quadratic discriminant function in two dimensions is $Ax^2 + By^2 + Cx + Dy + E = 0$

$$A = B \implies \text{circle}$$

$$AorBiszero \implies \text{Parabola}$$

$$A.B > 0 \implies \text{Ellipse}$$

$$A.B < 0 \implies \text{Hyperbola}$$

$$A = 0 \implies \text{Hyperplane}$$