Bayesian decision theory

Assume , two class problem.

Example :- An automatic system for quality measurement of a product industry.

Acceptance class = w_1 , Reject class = w_2

Based on previous record,

$$\left.\begin{array}{l}\text{Probability of acceptance} = p(w_1) \text{ known} \\[1em] \text{Probability of rejection} = p(w_2) \text{ known}\end{array}\right\} \text{Prior Probability}$$

We can make simple decision rule:-

If $p(w_1) > p(w_2)$, then decide class w_1

If $p(w_2) > p(w_1)$, then decide class w_2

I can find out the probabilistic measure or probability density function (PDF) of variable x for object which belongs to class w_1 and w_2 seperately.

$$\underbrace{p(x|w_1) \text{ and } p(x|w_2)}_{\text{Class conditional PDF}}$$

Our objective is to calculate:-

$$\underbrace{p(w_1|x) \text{ and } p(w_2|x)}_{\text{Posterior probability}}$$

Joint probability density function,

$$p(w_i, x) = p(w_i|x).p(x)$$

$$= p(x|w_i).p(w_i)$$

$$\Rightarrow\ p(w_i|x).p(x) = p(x|w_i).p(w_i)$$

$$\Rightarrow\ p(w_i|x) = \frac{p(x|w_i).p(w_i)}{p(x)}$$

$$\text{Posterior} = \frac{\text{Likelihood x Prior}}{\text{Evidence}}$$

$$\boxed{p(w_i|x) = \frac{p(x|w_i).p(w_i)}{p(x)}}$$

Bayes rule

$$p(x) = \sum_{i=1}^{2} p(x|w_i).p(w_i)$$

If $p(w_1|x) > p(w_2|x)$, then decide class w_1

If $p(w_2|x) > p(w_1|x)$, then decide class w_2

By expanding,

If $p(x|w_1)p(w_1) > p(x|w_2)p(w_2)$, then decide w_1

If $p(x|w_2)p(w_2) > p(x|w_1)p(w_1)$, then decide w_2

If $p(x|w_1)\,p(w_1) = p(x|w_2)p(w_2)$, then decision will based on $p(w_1)$ and $p(w_2)$.

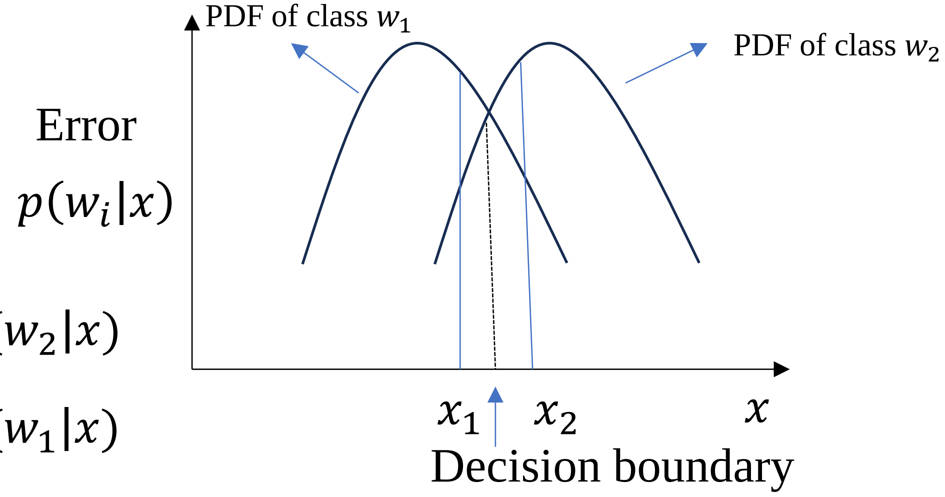
Error in this case:-

If $x_1 \in w_2$, then error $p(w_1|x)$

If $x_2 \in w_1$, then error $p(w_2|x)$

If I decide in favour of class w_1 then probability of error = $p(w_2|x)$

If I decide in favour of class w_2 then probability of error = $p(w_1|x)$



Total error = $\int_{-\infty}^{\infty} p(error, x)dx = \int_{-\infty}^{\infty} p(error|x).\, p(x)dx$

$$p(error, x) = \min\{p(w_1|x), p(w_2|x)\}$$

$$p(w_i|x) = \frac{p(x|w_i).\, p(w_i)}{\sum_{i=1}^{2} p(x|w_i).\, p(w_i)} \quad = \quad p(w_i|x) = \frac{p(x|w_i).\, p(w_i)}{p(x)}$$

If $p(w_1|x)$>$p(w_2|x)$,then decide class w_1

If $p(w_2|x)$>$p(w_1|x)$,then decide class w_2

Generalized bayes classifier

- Use more than two states of nature.
- More than one feature .
- More action to consider.
- Loss function.

c $\rightarrow$ No. of classes

$\{w_1, w_2, \dots \dots w_c\}$

No. of actions $= \{\alpha_1, \alpha_2, \dots \alpha_a\}$

Loss function:- $\lambda(\alpha_i | w_j)$, loss is occurred for taking action α_i when state of class is w_j.

x is d – dimensional vector

$$R(\alpha_i|x) = \sum_{j=1}^{c} \lambda(\alpha_i|w_j)p(w_j|x)$$

Risk function/conditional risk/expected loss.

Minimum Risk classifier:-

Two category case :- w_1 and w_2 and actions $\alpha_1$and α_2

For simplicity , $\lambda(\alpha_i|w_j)$ = λ_{ij}

In general,

$$R(\alpha_i|x) = \sum_{j=1}^{c} \lambda\underbrace{(\alpha_i|w_j)}_{\lambda_{ij}}p(w_j|x)$$

For two class problem.

For action $\alpha_1, R(\alpha_1|x) = \underset{\substack{\downarrow \\ \lambda(\alpha_1|w_1)}}{\lambda_{11}}\, p(w_1|x) + \underset{\substack{\downarrow \\ \lambda(\alpha_1|w_2)}}{\lambda_{12}}p(w_2|x)$

For action α_2, $R(\alpha_2|x) = \underset{\substack{\downarrow \\ \lambda(\alpha_2|w_1)}}{\lambda_{21}}\, p(w_1|x) + \underset{\substack{\downarrow \\ \lambda(\alpha_2|w_2)}}{\lambda_{22}}p(w_2|x)$

If $R(\alpha_1|x) < R(\alpha_2|x)$, then in favour of α_1

If $R(\alpha_1|x) > R(\alpha_2|x)$, then in favour of α_2

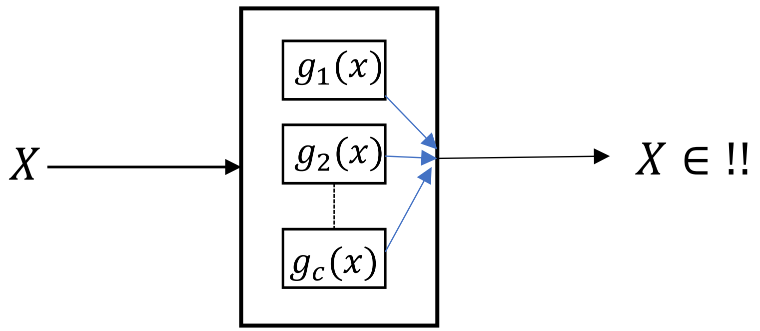
$$\lambda_{21}\, p(w_1|x) + \lambda_{22}p(w_2|x) > \lambda_{11}p(w_1|x) + \lambda_{12}p(w_2|x)$$

for decision in favour of $w_1$or action α_1

$$= (\lambda_{21} - \lambda_{11})p(w_1|x) > (\lambda_{12} - \lambda_{22})p(w_2|x)$$

If both , $(\lambda_{21}$- $\lambda_{11}) > 0$ and $(\lambda_{12} - \lambda_{22}) > 0$
And $p(w_1|x) > p(w_2|x)$, then decide class w_1.

Multi-category class



c = No. of classes $\quad g(x)$ = discriminant function

$\{w_1, w_2, \dots \dots w_c\}$, are c no of classes

$g_i(x)$; $i = 1,2,..,c$.

If $g_i(x) > g_j(x) \ \forall j \neq i$ decide $x \in w_i$.

<u>Minimum risk classifier</u>

We can let $g_i(x) = -R(\alpha_i|x)$

As we know , $R(\alpha_i|x) = 1 - p(w_i|x)$

Then, $g_i(x) = p(w_i|x)$

$f(\ g_i(x)) =$ monotonically increasing function.

Minimum error rate classification:-

$$g_i(x) = p(w_i|x) = \frac{p(x|w_i).p(w_i)}{\sum_{j=1}^{2} p(x|w_i).p(w_j)} \longrightarrow p(x)$$

$$g_i(x) = p(x|w_i).p(w_i)$$

Logarithmic function $\longrightarrow$ monotonically increasing function.

$$g_i(x) = \ln\ p(w_i|x) = \ln\ p(x|w_i) + \ln\ p(w_i)$$

Two category case:- Class w_1, Class w_2 $\Big\}$ We have now two discriminant function $g_1(x)$ and $g_2(x)$.

If $g_1(x) > g_2(x)$ decide class w_1
If $g_1(x) < g_2(x)$ decide class w_2 $\Big\}$ Decision boundary $g_1(x) - g_2(x) = 0$

Single discriminant function:-

$$g(x) = g_1(x) - g_2(x) = \ln\ p(w_1|x) - \ln\ p(w_2|x)$$

$$= \ln\ p(x|w_1) + \ln\ p(w_1) - \ln\ p(x|w_2) - \ln\ p(w_2)$$

$$= \ln\ \frac{p(x|w_1)}{p(x|w_2)} + \ln\ \frac{p(w_1)}{p(w_2)}$$

The Normal Density

Univariate Density:- We begin with the continuous univariate normal or Gaussian density,

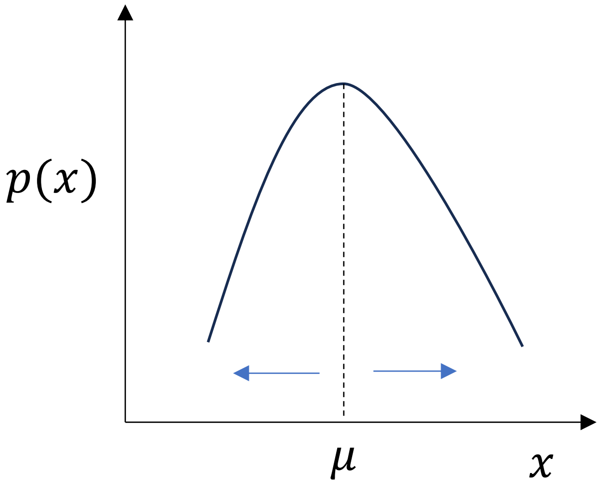
$$p(x) = \frac{1}{\sqrt{2\pi\sigma}} \exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right]$$

$\mu =$ expected value of x

$$\mu = E(x) = \int_{-\infty}^{\infty} x p(x) dx$$

$\sigma^2 =$ variance

$$\sigma^2 = E[(x-\mu)^2] = \int_{-\infty}^{\infty} (x-\mu)^2 p(x) dx = N(\mu, \sigma^2)$$



Multivariate Density:- The general multivariate normal density in d dimensions is written as

$$p(x) = \frac{1}{(2\pi)^{d/2}|\Sigma|^{1/2}} \exp\left[-\frac{1}{2}(x-\mu)^t \Sigma^{-1}(x-\mu)\right]$$

$x =$ feature vector with dimension d

$\mu =$ expected value of dimension d

$$\mu = E(x) = \int_{-\infty}^{\infty} xp(x)dx$$

$\Sigma =$ covariance matrix

$$\Sigma = E[(x-\mu)(x-\mu)^t] = \int_{-\infty}^{\infty} (x-\mu)\,(x-\mu)^t p(x)dx$$

$$(x-\mu) = (d \times 1)$$
$$(x-\mu)^t = (1 \times d)$$
$$\Sigma = E(x) = (d \times d)$$

ith component $\mu_i = E[x_i]$

$(i,j)^{th}$ component $\sigma_{i,j} = E[(x_i-\mu_i)(x_j-\mu_j)^t]$

Diagonal component, $\sigma_{i,i} = E[(x_i-\mu_i)^2] = \sigma_i^2$

Bivariate normal density function:-

$X =$ two dimensional feature vector

$$X = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

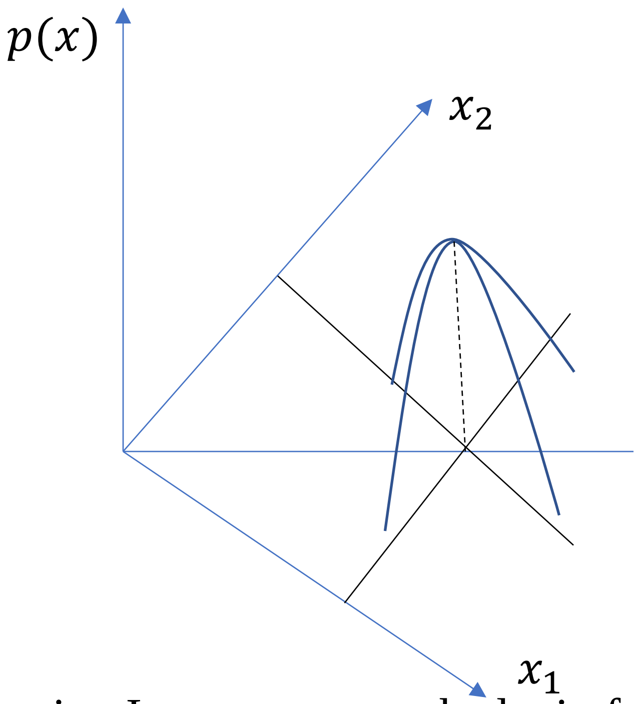
$$p(x) = \frac{1}{(2\pi)|\Sigma|^{1/2}} \exp\left[-\frac{1}{2}\{(\frac{x_1 - \mu_1}{\sigma_1})^2 + (\frac{x_2 - \mu_2}{\sigma_2})^2\}\right]$$

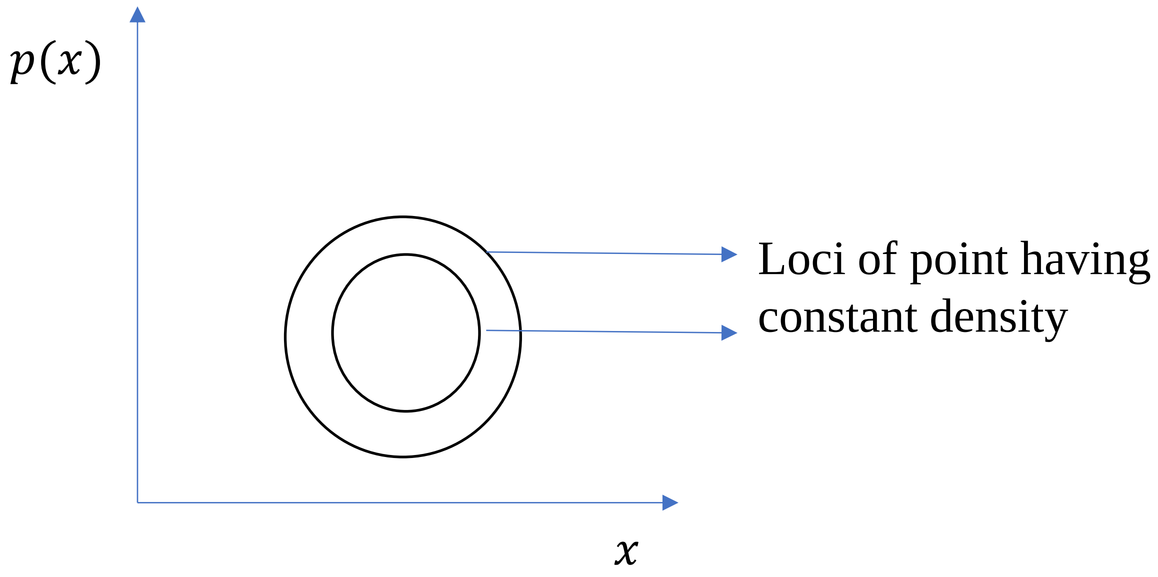
$$\Sigma = \begin{bmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{bmatrix} \qquad \mu = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}$$

$$p(x) = \frac{1}{(2\pi)|\Sigma|^{1/2}} \exp\left[-\frac{1}{2}(\frac{x_1-\mu_1}{\sigma_1})^2\right] \exp\left[(\frac{x_2-\mu_2}{\sigma_2})^2\right]$$

Physical Interpretation:-

(i) First case:- $\sigma_1 = \sigma_2$





For bivariate function I want to trace the loci of constant density i.e. all value of x for which $p(x)$ is constant , those loci is nothing but circle.

Along with these circles , I have more probability of occurrence of set of points which are drawn from a single population arbitrary.

(ii) Second case:- ${\sigma_1}^2 \neq {\sigma_2}^2$

$$\Sigma = \begin{bmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{21} & \sigma_2^2 \end{bmatrix}$$

If all samples are statistically independent,

$$\sigma_{12} = \sigma_{21} = 0$$

then $$\Sigma = \begin{bmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{bmatrix}$$

(iii) Third case:-Data are not statistically independent .

$e_1 =$ eigen vectors of the covariance matrix Σ

Discriminant Functions for the Normal Density

We know that the discriminant functions given as:-

$$g_i(x) = lnp(w_i|x) = lnp(x|w_i) + lnp(w_i)$$

Multivariate Density:-

$$p(x|w_i) = \frac{1}{(2\pi)^{d/2}|\Sigma_i|^{1/2}} \exp\left[-\frac{1}{2}(x-\mu_i)^t \Sigma_i^{-1} (x-\mu_i)\right]$$

Discriminant Functions:-

$$g_i(x) = -\frac{1}{2}\left[(x-\mu_i)^t \Sigma_i^{-1} (x-\mu_i)\right] - \frac{d}{2} ln2\pi - \frac{1}{2}\ln|\Sigma_i| + lnp(w_i)$$

Let us examine the discriminant function and resulting classification for a no. of special cases.

<u>Case :-</u>

$\Sigma_i = \sigma^2 I$ $\qquad\qquad$ I = Identity matrix $(d \times d)$

$\sigma_{i,j} = 0,$ different components are statistically independent.

$|\Sigma_i| = \sigma^{2d}$

$$\left|\Sigma_i^{-1}\right| = \frac{1}{\sigma^2} I$$

$$g_i(x) = -\frac{1}{2}\left[(x-\mu_i)^t \Sigma_i^{-1}\,(x-\mu_i)\right] - \underbrace{\frac{d}{2} ln 2\pi - \frac{1}{2} \ln}_{\text{constant}} |\Sigma_i| + \ln\ p(w_i)$$

$\longrightarrow$ Independent of i,so they are ignored.

Thus we obtain the simple discriminant functions:-

$$g_i(x) = -\frac{\left||x-\mu_i|\right|^2}{2\sigma^2} + lnp(w_i)$$

where $||.||$ is the Euclidean norm that is ,

$$\left||x-\mu_i|\right|^2 = (x-\mu_i)^t\,(x-\mu_i)$$

Expansion of the quadratic form $(x - \mu_i)^t \ (x - \mu_i)$ yields

$$g_i(x) = -\frac{1}{2\sigma^2}\left[x^t x - 2\mu_i^t x + \mu_i^t \mu_i\right] + \ln\ p(w_i)$$

which appears to be quadratic function of x. However ,the quadratic term $x^t x$ is independent for all i, making it an ignorable additive constant. then,

$$g_i(x) = -\frac{1}{2\sigma^2}\left[-2\mu_i^t x + \mu_i^t \mu_i\right] + \ln\ p(w_i)$$

Thus we obtain the equivalent linear discriminant functions

$$g_i(x) = w_i^t x + w_{i0}$$

where

$$w_i = \frac{1}{\sigma^2}\mu_i \ , w_{i0} = -\frac{1}{2\sigma^2}\mu_i^t \mu_i + \ln\ p(w_i)$$

If $g_i(x) > g_j(x)$,then $x \in$ class i

If $g_j(x) > g_i(x)$,then $x \in$ class j

$$g_i(x) = g_j(x) \text{ or } g_i(x) - g_j(x) = 0$$

$$g_i(x) = w_i^t x + w_{i0}$$

$$g_j(x) = w_j^t x + w_{j0}$$

$$g_i(x) - g_j(x) = 0$$

$$\Rightarrow (w_i - w_j)^t x + w_{i0} + w_{j0} = 0$$

$$\Rightarrow \frac{1}{\sigma^2}(\mu_i - \mu_j)^t x - \frac{1}{2\sigma^2}\mu_i^t \mu_i + \ln\ p(w_i) + \frac{1}{2\sigma^2}\mu_j^t \mu_j - \ln\ p(w_j) = 0$$

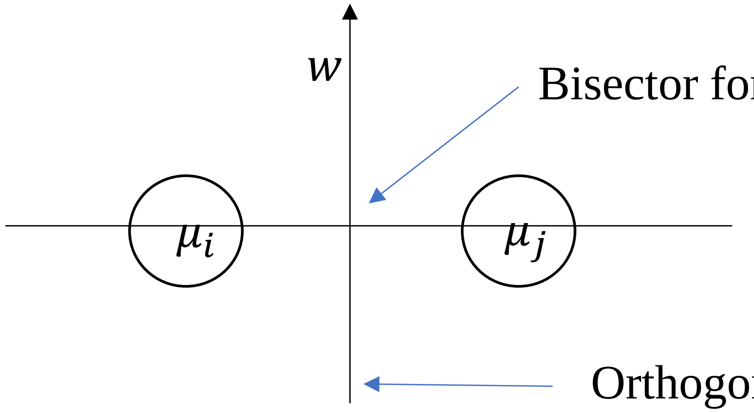
$$\Rightarrow (\mu_i - \mu_j)^t x - \frac{1}{2}(\mu_i^t \mu_i - \mu_j^t \mu_j) + \sigma^2 \frac{\ln\ p(w_i)}{\ln\ p(w_j)} = 0$$

$$\Rightarrow (\mu_i - \mu_j)^t \left[x - \left\{ \frac{1}{2}(\mu_i + \mu_j) - \frac{\sigma^2}{||\mu_i - \mu_j||^2} \left[\frac{\ln\ p(w_i)}{\ln\ p(w_j)} \right] (\mu_i - \mu_j) \right\} \right] = 0$$

$$\Rightarrow w_t(x - x_0) = 0$$

$$w = \frac{1}{2}(\mu_i - \mu_j) \qquad x_0 = \frac{1}{2}(\mu_i + \mu_j) - \frac{\sigma^2}{||\mu_i - \mu_j||^2}\left[\frac{\ln\ p(w_i)}{\ln\ p(w_j)}\right](\mu_i - \mu_j)$$

If $p(w_i) = p(w_j)$; $\qquad x_0 = \frac{1}{2}(\mu_i + \mu_j)$



Bisector for the line joining between μ_i and μ_j

Orthogonal to the line joining between μ_i and μ_j