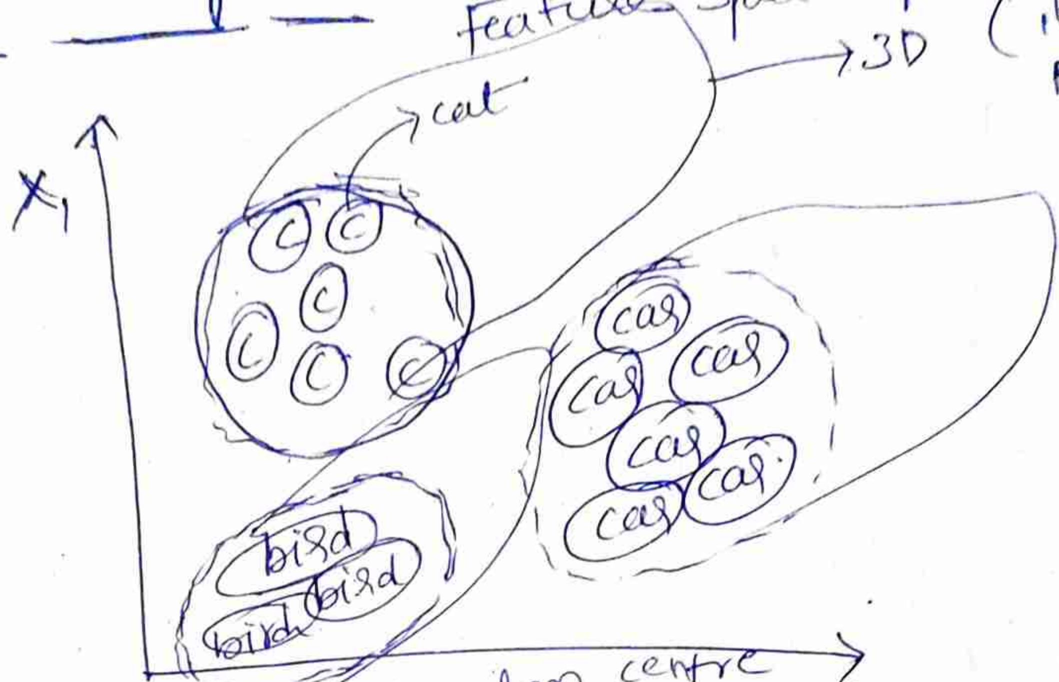
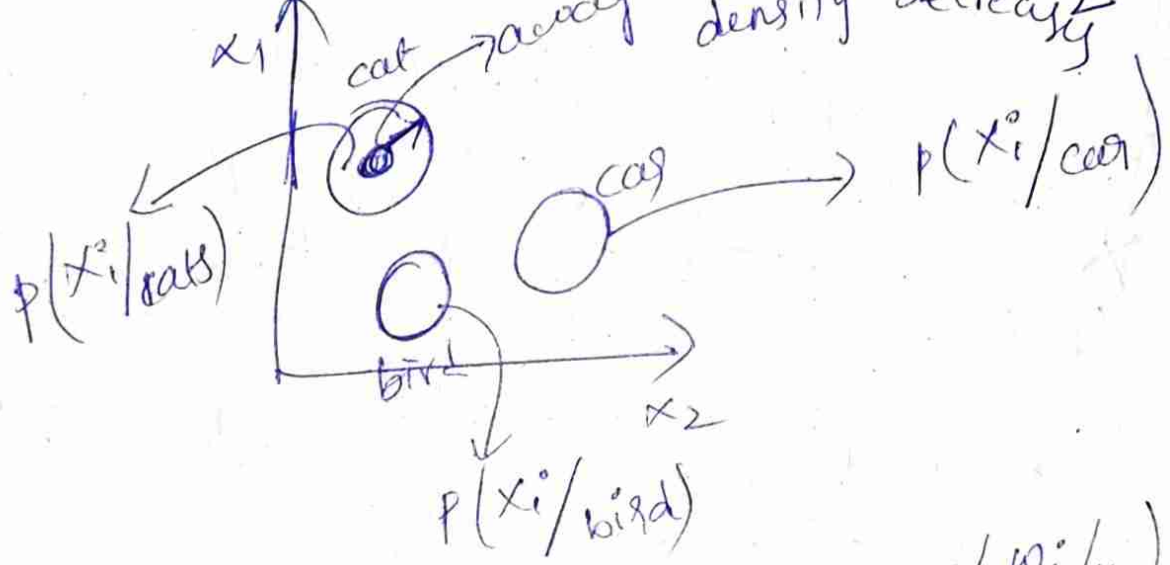
Baye's Learning - I

Feature Space representation → 3D (it is a probability distribution)





but we need to find $p(w_i/x)$ → given this representation what is the class for that

$$P(A \cap B) = P(A/B)\,P(B)$$
$$= P(B/A)\,P(A)$$

$A \to x$, $B \to \omega_i$

$$P(x/\omega_i)\,P(\omega_i) = P(\omega_i/x)\,P(x)$$

$$P\left(\frac{\omega_i}{x}\right) = \frac{P(x/\omega_i)\,P(\omega_i)}{P(x)}$$

$$P(\omega_j/x) = \frac{P(x/\omega_j)\,P(\omega_j)}{P(x)}$$

same so

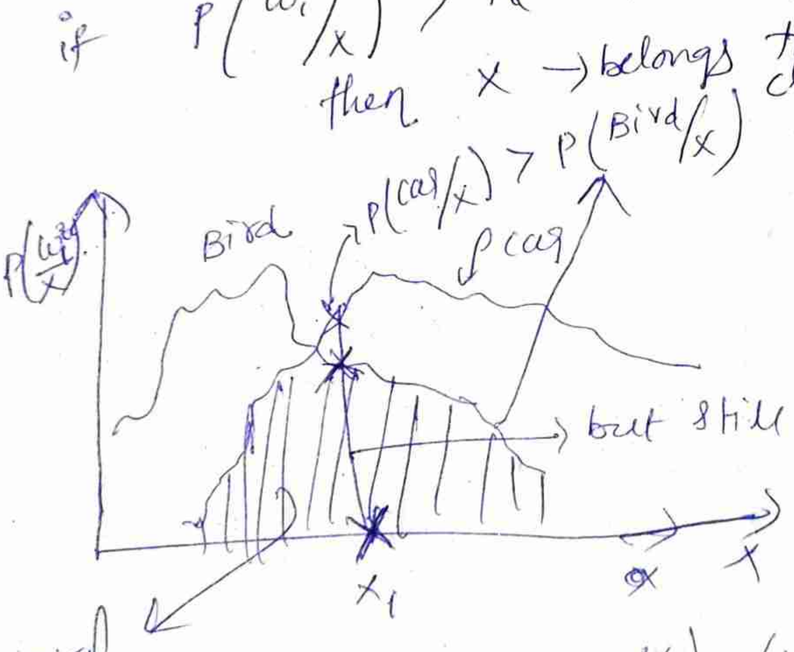
$$P(\omega_i/x) = P(x/\omega_i)\,P(\omega_i)$$
$$P(\omega_j/x) = P(x/\omega_j)\,P(\omega_j)$$

$$P(x) = \sum_{\forall i} P(x/\omega_i)\,P(\omega_i)$$

if $P(\omega_i/x) > P(\omega_j/x)$

then $x \to$ belongs to class ω_i

$P(\text{cat}/x) > P(\text{Bird}/x)$



but still there is finite probability that x belongs to bird also.

total error

$$P(\text{error}/x) = \min\left(P(\omega_i/x),\ P(\omega_j/x)\right)$$

$\hookrightarrow$ Bayes minimum error classifier

Bayes Minimum Risk classifier

ω_i : $i = 1, 2, \ldots c$

α_j : $j = 1, 2, \ldots k \to$ actions.

x : $d \to$ dimensional feature vector.

$\lambda(\alpha_i/\omega_j)$ for $i = 1, 2, \dots K$

$X \leftarrow \alpha_i$ loss function

$$R(\alpha_i/x) = \sum_{\forall j} \lambda(\alpha_i/\omega_j)\, P(\omega_j/x)$$

if we take action α_i what is the loss that it belongs to ω_j

gives overall risk for taking an action α_i on given vector x_i.

$\lambda(\alpha_i/\omega_j) = \lambda_{ij}$

$\lambda_{ij} = 0$ for $i = j$ → i am taking correct action i.e, it actually belongs to ω_i class and i am also felling that it belongs to class ω_i → taking actinon.

so loss 0.

$$R(\alpha_i/x) = \sum_{\forall j} \lambda_{ij}\, P(\omega_j/x)$$

$\lambda_{ij} = 1$ for $i \neq j$

$$= \sum_{j \neq i} P(\omega_j/x) = \underbrace{1 - P(\omega_i/x)}_{\text{for minimazion of this}}$$

we need to minimize this

Bayes minimum error classifier — what is the $P(\omega_i/x)$ maximum

$$R(\alpha_i/x) = \sum_{\forall j} \lambda(\alpha_i/\omega_j)\, P(\omega_j/x)$$

$\alpha_i \quad i = 1, 2, \dots K$.

$\lambda(\alpha_i/\omega_j) = \lambda_{ij}$

$\alpha_1 \quad \alpha_2$

$\omega_1 \quad \omega_2 \quad j = 1, 2$

$$R(\alpha_1/x) = \lambda_{11} P(\omega_1/x) + \lambda_{12} P(\omega_2/x)$$

$$R(\alpha_2/x) = \lambda_{21} P(\omega_1/x) + \lambda_{22} P(\omega_2/x)$$

if $R(\alpha_1/x) < R(\alpha_2/x) \longrightarrow$ take action α_1

$$\lambda_{11} P(\omega_1/x) + \lambda_{12} P(\omega_2/x) < \lambda_{21} P(\omega_1/x) + \lambda_{22} P(\omega_2/x)$$

$$(\lambda_{21} - \lambda_{11}) P(\omega_1/x) > (\lambda_{12} - \lambda_{22}) P(\omega_2/x)$$

$$\Rightarrow \frac{P(\omega_1/x)}{P(\omega_2/x)} > \frac{\lambda_{12} - \lambda_{22}}{\lambda_{21} - \lambda_{11}}$$

loss function for taking action α_2 when actual true nature (class) is ω_1

$$P(\omega_1/x) = P(x/\omega_1) P(\omega_1)$$

$$P(\omega_2/x) = P(x/\omega_2) P(\omega_2)$$

$$\frac{P(x/\omega_1) P(\omega_1)}{P(x/\omega_2) P(\omega_2)} > \frac{\lambda_{12} - \lambda_{22}}{\lambda_{21} - \lambda_{11}}$$

$$\frac{P(x/\omega_1)}{P(x/\omega_2)} > \frac{\lambda_{12} - \lambda_{22}}{\lambda_{21} - \lambda_{11}} \times \frac{P(\omega_2)}{P(\omega_1)}$$

Discriminant Function

$R(\alpha_i/x) \rightarrow$ minimum then it is the good thing to take

so $-R(\alpha_i/x)$ is maximum then it is good to take

$g_i(x) \longleftrightarrow \omega_i$

if $g_i(x)$ is maximum then ω_i is the prediction

$$g_i(x) = P(\omega_i/x) = P(x/\omega_i) P(\omega_i)$$
$$= -R(\alpha_i/x)$$

$g_i(x) = f\{P(\omega_i/x)\}$

↑ monotically increasing function

if $P(\omega_i/x) > P(\omega_j/x)$

$\xleftrightarrow{\text{monotonically increasing}} f\{P(\omega_i/x)\} > f(P\{\omega_j/x\})$

$\ln(x) \rightarrow$ monotinically increasing so it is also better option to take as f

$$g_i(x) = \ln\{P(\omega_i/x)\}$$
$$= \ln\{P(x/\omega_i)\,P(\omega_i)\}$$
$$= \ln\left(P\left(\frac{x}{\omega_i}\right)\right) + \ln(P(\omega_i))$$
$$= \ln P(x/\omega_i) + \ln P(\omega_i)$$

lets assume $P(x/\omega_i)$ follows multivariate Normal distribution.

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

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

because for different classes different covariance matrix

$$g_i(x) = \ln P(\omega_i) \underbrace{- \frac{d}{2}\ln(2\pi)}_{} - \frac{1}{2}\ln|\Sigma_i| - \frac{1}{2}(x-\mu_i)^t \Sigma_i^{-1}(x-\mu_i)$$

independent of class ω_i as there is no subscript i. so for comparing different classes it does not matter.

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