Date 7 / 4 / 24

# Bayesian Decision Theory

→ Priori probabilities: $P(\omega_1), P(\omega_2)$ → calculate using $N_1/N$, $N_2/N$

→ Likelihood probability: $P(X/\omega_i)$

↳ we will calculate feature vector for each instance

e.g.


↳ decision ~~rule~~ boundary is a straight line

∴ we can calculate likelihood.

→ Posterior probability: $P(\omega_i/X) = \dfrac{P(X/\omega_i)\,P(\omega_i)}{P(X)}$

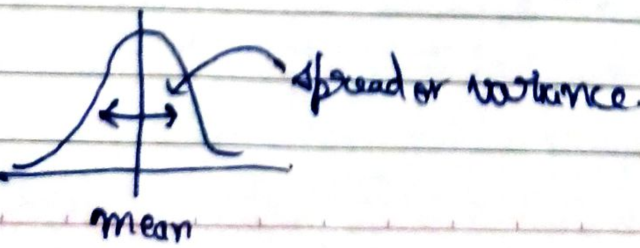
where $P(X) = \sum_{i=1}^{n} P(X/\omega_i)\,P(\omega_i)$

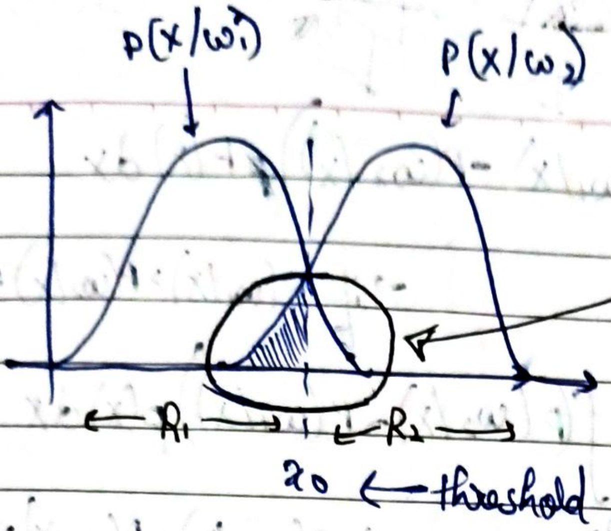
→ Bayes classification: if $P(\omega_1/X) > P(\omega_2/X)$ then new sample is $\omega_1$ else $\omega_2$

∴ $p(X/\omega_1).p(\omega_1) > p(X/\omega_2).p(\omega_2)$

↳ if $p(\omega_1) = p(\omega_2) = 1/2$, $P(X/\omega_1) > P(X/\omega_2) \Rightarrow \omega_1$

⇒ Most of examples in real world follow normal distribution






this portion has collision
∴ there is error ∵ sample belonging to $\omega_2$ is classified into $\omega_1$.

If $x < x_0$, it will classify in $\omega_1$
$x > x_0$, it will classify in $\omega_2$

∴ probability error, $P_e = \left( \int_{-\infty}^{x_0} P(x/\omega_2)\,dx + \int_{x_0}^{\infty} P(x/\omega_1)\,dx \right) \frac{1}{2}$

→ How we can minimise this error probability

↳ Bayesian classifier is an optimal classifier.

$$P_e = P(x \in R_2, \omega_1) + P(x \in R_1, \omega_2)$$

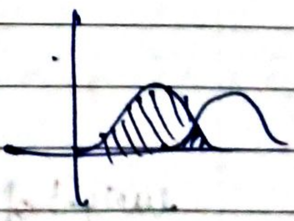
↳ i.e. $P(x \in R_2 \text{ and } x \in \omega_1)$ or $P(x \in R_2 \cap \omega_1)$

From here, $P_e = P(x \in R_2/\omega_1)P(\omega_1) + P(x \in R_1/\omega_2)P(\omega_2)$

↳ i.e. $P(x \in R_2 / x \in \omega_1) \cdot P(x \in \omega_1)$

$$\therefore P_e = P(\omega_1)\int_{R_2} P(x/\omega_1)\,dx + P(\omega_2)\int_{R_1} P(x/\omega_2)\,dx$$

[Now, for $R_1$, $P(\omega_1/x) > P(\omega_2/x)$
$R_2$ : $P(\omega_2/x) > P(\omega_1/x)$



$$\therefore P_e = \int_{R_1} P(\omega_1/x)P(x)\,dx + \int_{R_2} P(\omega_2/x)P(x)\,dx$$

[Using Bayes rule]

priori prob

$$\therefore \quad P_e = P(\omega_1) - \int_{R_1} \left(P(\omega_1/x) - P(\omega_2/x)\right) p(x)\, dx$$

→ if $P(\omega_1/x) > P(\omega_2/x) \Rightarrow$ error min

Similarly,
$$P_e = P(\omega_2) - \int_{R_2} \left(P(\omega_2/x) - P(\omega_1/x)\right) p(x)\, dx$$

→ if $P(\omega_2/x) > P(\omega_1/x) \Rightarrow$ error min

Therefore, Bayes decision rule ~~or~~ minimize the error.

---

→ $\omega$ = state of nature = class = random variable
→ $p(x)$ = probability density function = evidence
→ $p(x|\omega_i)$ = conditional probability density = likelihood

→ $P(\text{error}) = \min\left(P(\omega_1), P(\omega_2)\right)$ ⇐ prior probability decision rule

---

**Risk** :-

↳ we also consider $\lambda(\alpha_i|\omega_j)$ i.e. weight function.
where, $\alpha_i \rightarrow$ action
∴ It tells how much risk is there e.g. tumor is benign or malignant

→ risk = $R = \underbrace{\lambda_{12}}_{\text{factor}} P(\omega_1) \underbrace{\int_{R_2} p(x|\omega_1)\, dx}_{\text{original}} + \underbrace{\lambda_{21}}_{\text{factor}} P(\omega_2) \underbrace{\int_{R_1} p(x|\omega_2)\, dx}_{\text{original}}$

factor ↓ weight of class 1 but assigning it to class 2.

factor ↓ weight of class 2 but assigning it to class 1.

e.g. if tumor $\omega_1 \to$ malign
$\omega_2 \to$ benign
then $\lambda_{12} > \lambda_{21}$

→ In general for $m$ classes,

$$r_k = \sum_{i=1}^{m} \lambda_{ki} \int_{R_i} P(x/\omega_k)\, dx$$ ← belonging to class k but classified to other classes

∴ total risk, $r = \sum_{k=1}^{m} r_k\, P(\omega_k)$

→ ∴ $x \in R_i$, if $\sum_{k=1}^{m} \lambda_{ki} P(x/\omega_k) P(\omega_k) < \sum_{k=1}^{m} \lambda_{kj} P(x/\omega_k) P(\omega_k)$

$\forall\ j \neq i$

because in that case the risk is minimum.

→ ∴ $$\left.\begin{aligned} l_1 &= \lambda_{11} P(x/\omega_1) P(\omega_1) + \lambda_{21} P(x/\omega_2) P(\omega_2) \\ l_2 &= \lambda_{12} P(x/\omega_1) P(\omega_1) + \lambda_{22} P(X/\omega_2) P(\omega_2) \end{aligned}\right\}$$ for 2 class classifier

↳ 4 weights needed, where, $\lambda_{11} = \lambda_{22} = 0$.

∴ we assign $x \to \omega_1$ if $l_1 < l_2$ and vice versa

∴ loss matrix, $L = \begin{bmatrix} 0 & \lambda_{12} \\ \lambda_{21} & 0 \end{bmatrix}$ ← in our case $\lambda_{12} > \lambda_{21}$

∴ $P(X/\omega_2)\lambda_{21} < P(X/\omega_1)\lambda_{12}$ in our case.

Confusion matrix :

| prediction \ actual | T | F |
|---|---|---|
| T | T.P. | F.P |
| F | F.N. | T.N. |

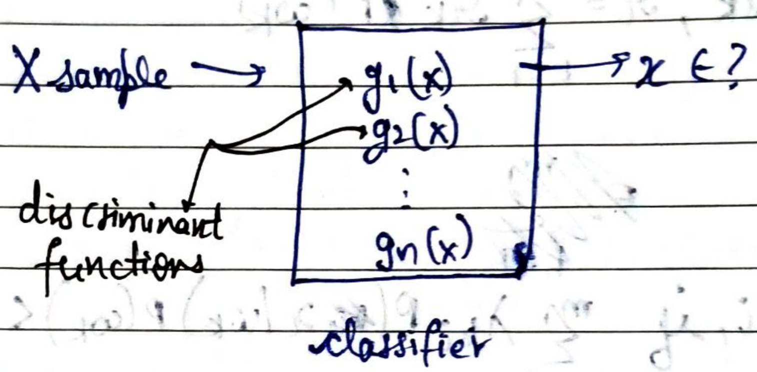
true positive

false positive (type-1 error)

false negative (type-2 error) → more dangerous

true negative ~~false positive~~

→ Discriminant function




→ It is opposite to risk function.

→ It is monotonically increasing function.

∴ if $g_i(x) > g_j(x)$, $x \in \omega_i$

$\boxed{g_i(x) = -R(\alpha_i/x)}$ ~~where $R(\alpha_i/x) = 1 - P(\omega_i/x)$~~

DOMS Page No.
Date 10/9/24

# Discriminant function

→ It is a function which is used to create decision boundary / decision surface.

as we know, risk, $R(\alpha_i/x) = \sum_{i=1}^{c} \lambda(\alpha_i/\omega_i)\, P(\omega_j/x)$ (no. of classes ↓ $c$; action ↓ $\alpha_i$)

$\approx 1 - P(\omega_i/x)$

→ ∴ discriminant function $g_i(x) = \max \{R(\alpha_i/x)$
$= -R(\alpha_i/x)$
$= P(\omega_i/x)$ ↳ density function.

↳ it can be any F s.t. $F(\cdot)$ is monotonically function then $F(g_i(x))$ is also monotonically incr. function.

→ $g_i(x) = P(\omega_i/x)$
take F = log function

∴ $F(g_i(x)) = \ln P(\omega_i/x)$

now, $P(\omega_i/x) = \dfrac{P(x/\omega_i)\,P(\omega_i)}{\boxed{P(x)}}$ ← constant

∴ $F(g_i(x)) = \ln\left[P(x/\omega_i)P(\omega_i)\right]$

$= \ln P(x/\omega_i) + \ln P(\omega_i)$ ← Discriminant function

↑ prob. distr. functions

Prob. density function → Parametric ← distribution curve known
→ Non-parametric ← distribution curve unknown.

## Discrete PDF :-

① Bernoulli distribution

→ Denoted by $x \sim Ber(p)$ ← known parameter

→ There is one element which can have only 2 outputs

→ Prob. of success $= p$
Prob. of failure $= 1-p$

② Binomial distribution → n independent elements
or n identical Bernoulli trials

$$P(X=x) = {}^{n}C_{x}\, p^{x} q^{n-x}$$

③ Poisson distribution

→ $P(X=x) = \dfrac{e^{-\lambda}\lambda^{x}}{x!}$ → when $n \to \infty$
$p$ is small
$np = \text{constant } (\lambda)$

④ Geometric distribution
→ Keep trying until you get success.

$$P(X=x) = pq^{x-1}$$

e.g. if $x$ = toss of coin & head is success
T → TT → TTT → TTTH.

∴ Success occurs only one time in all the trials.

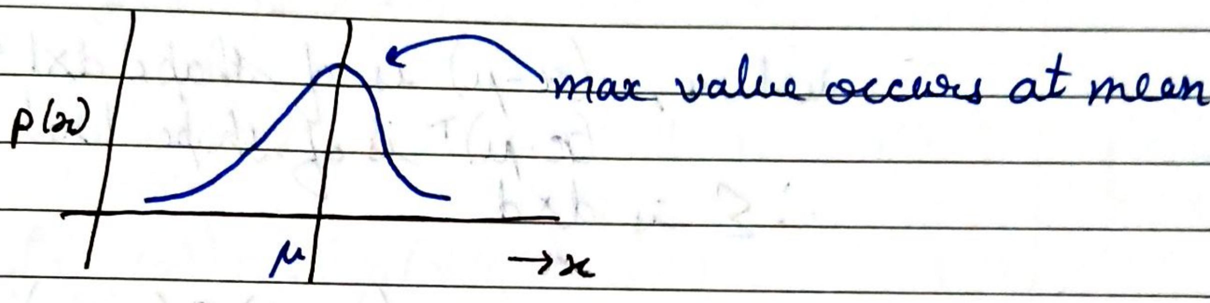
## Normal distribution / Gaussian distribution

→ Represented as $X \sim N(\mu, \sigma^2)$

$\mu$ ↑ mean, $\sigma^2$ ↑ variance

→ PDF, $p(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$

→ Curve :




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

$= \int_{-\infty}^{\infty} (x-\mu)^2 p(x)\, dx$

---

## Multivariate Normal Density

⇒ univariate → $x$ i.e. there is only one feature

→ But we cannot take only single feature to decide which feature belong to which class.

→ Instead we have feature vector.

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

where, $\mu$ = expected value = $E[x] = \int_{-\infty}^{\infty} x\, p(x)\, dx$

$x$ is d-dimensional feature vector

$\Sigma$ is covariance matrix

$|\Sigma|$ = is determinant of covarience matrix

$\Rightarrow$ $\Sigma = E[(x-\mu)(x-\mu)^T]$

where, $(x-\mu)$ is of shape $d \times 1$
$(x-\mu)^T$ is of shape $1 \times d$
$\therefore$ $\Sigma$ is $d \times d$

$\Rightarrow$ Inner product = scalar = $(x-\mu)^T (x-\mu)$

$\therefore$ $\Sigma = \int_{-\infty}^{\infty} (x-\mu)(x-\mu)^T p(x)\, dx$

$\sigma_{ij} = E[(x_i-\mu_i)(x_j-\mu_j)]$ → co variance

$\sigma_i^2 = E[(x-\mu_i)^2]$ → variance

$\therefore$ Diagonals represent variance and rest are covariance

e.g. Bivariate, $x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$

Lets assume $x_1$ and $x_2$ are statistically independent

$\therefore$ $\sigma_1^2$ & $\sigma_2^2$ only exist while $\sigma_{12} = \sigma_{21} = 0$

DOMS Page No.
Date / /

$\rightarrow \therefore \Sigma = \begin{bmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{bmatrix}$ & $p(x) = \dfrac{1}{(2\pi)|\Sigma|^{1/2}} e^{\left(\frac{-1}{2}\left(\left(\frac{x_1-\mu_1}{\sigma_1}\right)^2 + \left(\frac{x_2-\mu_2}{\sigma_2}\right)^2\right)\right)}$

$\rightarrow \therefore \cancel{p(x/\omega_i)}$

Final pdf

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

$g_i(x) = p(x/\omega_i)$

$\therefore f(g_i(x)) = \ln p(x/\omega_i)$

Final discriminant function

$= \dfrac{-1}{2}\left[(x-\mu_i)^T \Sigma_i^{-1} (x-\mu_i)\right] - \dfrac{d}{2}\ln(2\pi) - \dfrac{1}{2}\ln \Sigma_i + \ln p(\omega_i)$