$P(A|B) \rightarrow$ Given that i have observed data B what is the probability that it has come from A.

$\Rightarrow$ Read about Bayesian Estimator.

- BIAS $\rightarrow$ Diff. b/w actual and the value calculated by the estimator.

- Confidence interval

① $CI = A \pm B$

A $\rightarrow$ sample property.

sample se calculated

B $\rightarrow$ margin of error.

$\rightarrow$ uncertainity

margin of error $= M \times \hat{SE}$

M $\rightarrow$ multiplier

$\hat{SE}$ $\rightarrow$ property of a sample

$\rightarrow$ std. error

$\frac{\text{std. deviation}}{\sqrt{N}}$

can be changed even for a sample.

- Hypothesis Testing →

R.V

① Uniform Distribution

$$P(X=i) = \frac{1}{b-a}$$

$$a \leq i \leq b$$

$$E(X) = \frac{a+b}{2} \qquad \text{Mode} = [a, b]$$

$$\text{median} = \frac{a+b}{2}$$

$$\text{var} = \frac{(a-b)^2}{12}$$

$$\sigma^2 =$$

Used for sampling and random thing generations

→ When there is no prior info suggesting one outcome to be more likely than other

- Bernoulli → Only two possible outcomes

$$P(X=1)=p$$

$$P(X=0)=1-p$$

$E(X)=p$ • Mode = 1

$var = p(1-p)$ • median = $1 \quad p \geq 0.5$

$0 \quad p < 0.5$

- Single learnable parameter
- Memoryless
- Coin flips, inventory, info theory

- **Binomial Distribution** →

  - N independent trials

P of k successes out of n independent bernoulli trial

$$P(X=k) = \binom{n}{k} p^k q^{n-k}$$

$$E(X) = np$$

$$Var = np(1-p)$$

- Find no. of trials

only 2 learnable params.

Use → Quality Control.

- Geometric → p of getting 1st success

$$P(X=n) = (1-p)^{n-1} p$$

p is learnable

- Discrete

$$E(X) = 1/p$$

- mode = 1

me

$$Var = \frac{1-p}{p^2}$$

- Memoryless
- No fixed no. of trails
- No upper limit
- Modeling no. of trials until the first failure.

- Gaussian →

$$P(X=x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{-(x-\mu)^2}{2\sigma^2}}$$

μ, σ

mean = mode = median.

- Poisson R.V →

$$P(X=i) = \frac{e^{\lambda} \lambda^i}{i!}$$

$E(X) = \lambda$

Var = λ

- Memoryless
- By changing the value of λ we can cover outliers.

→ To model the probability of occurrence of something in a given area, volume or measure of time

Thing per unit something

Poisson.

Rare events also.

- Exponential

Models the time b/w events in Poisson process

$$\boxed{P(X = x) = \lambda e^{-\lambda x}}$$

$$P_1 \xleftrightarrow[\text{Time}]{} P_2$$

$$E(X) = 1/\lambda$$

$$var = 1/\lambda^2$$

→ Modeling the time until a component fails or system fails.

Q1. $\mu = 75 \quad \sigma = 10$

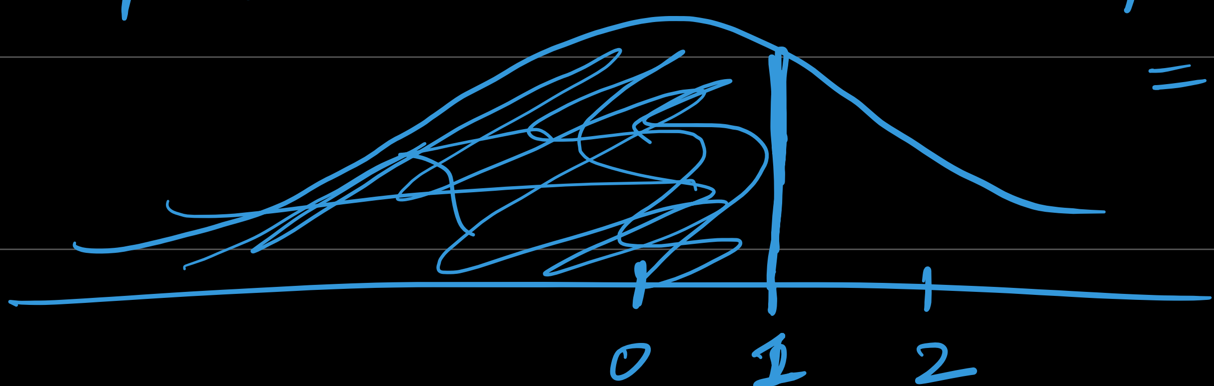
$$\frac{85-75}{10} = 1$$

$$SE = \frac{10}{\sqrt{10}}$$

$$\frac{70-75}{10/\sqrt{10}}$$

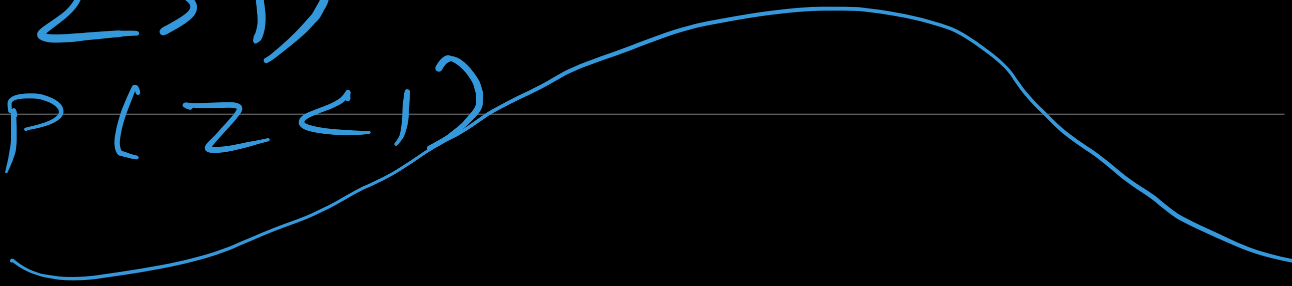
Q1. $n = 30 \quad \mu = 75 \quad \sigma = 10$

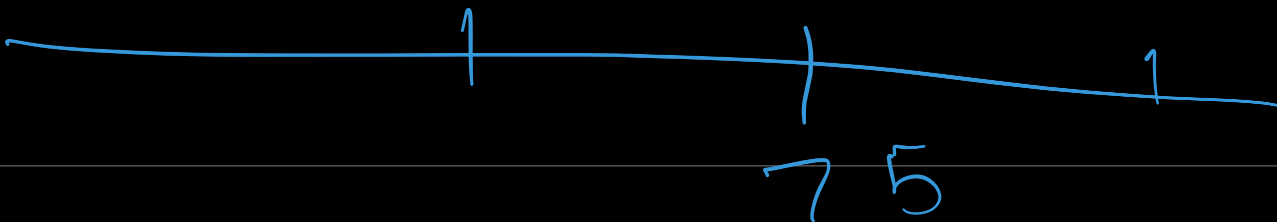
$P(X > 85)$ $\quad z = \frac{85-75}{10} = 1$



$P(Z > 1)$

$1 - P(Z < 1)$





b) $n = 10$

$\hat{SE} = \frac{10}{\sqrt{10}}$ $\quad z = \frac{70 - 75}{10/\sqrt{10}}$

$P(Z < \quad) =$

Q2. $p = 0.95$ $\quad q = 0.05$

n a) $^{10}C_3 (0.05)^3 (0.95)^7$

b) $np^{\rightarrow 0.05}$ $\quad \sqrt{npq}$

Q3.

a) $\dfrac{180-170}{10} = 1$

$1 - P(Z<1)$

b) ① $\left(1 - P(X<160)\right)^3$

$\dfrac{160-170}{10} = -1$

Q4. $\lambda = 4$

a) $\displaystyle\int_0^{1/6} \lambda e^{-\lambda x}\,dx$

b) $E(X) = 1/\lambda$ 15 min.

Q5.

Supervised Learning → Classification

↳ A prior info of that domain i.e class labels in ML.

↳ A set of patterns with class label information (which we call as a training pattern available to us).

↳ Make a classification model with the help of training pattern.

↳ Assign the class label of a new pattern to a predefined classes by classifier through calculation of Similarity / Dissimilarity measures.

→ Probablistic Model

↳ Bayes Classifier.

→ Similarity Based

MDC, KNN, LDC, QDC.

→ Neural Network

→ Graph Based

Decision Tree

→ Support Vector Machine

→ Random Forest.

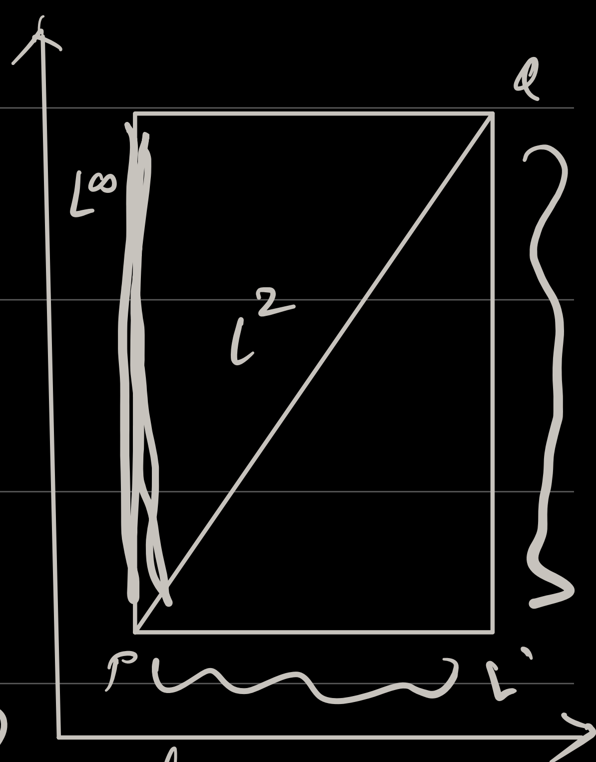
→ L^p norm

$p=1$, city - block

$p=2$, Euclidean

⋮

$p=\infty$, $L^\infty = \underset{\text{over all } i}{\mathrm{Max}} (x_i - y_i)$



Distance.

$L^1 > L^2 > L^3 > \ldots > L^\infty$

Euclidean distance $= \sqrt{\sum_{i=1}^{n} (x_i - y_i)^2}$

→ A pattern is a point in n-dimensional Space, where n is no. of features present in the data set.

Each feature represent one dimension.

$$x = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \qquad y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}$$

$$z = x - y = \begin{bmatrix} x_1 - y_1 \\ x_2 - y_2 \\ \vdots \\ x_n - y_n \end{bmatrix}$$

$$d_2^{\,2} = \sum_{i=1}^{n} (x_i - y_i)^2$$

$$= \begin{bmatrix} x_1 - y_1 & x_2 - y_2 & \cdots & x_n - y_n \end{bmatrix}_{1\times n} \begin{bmatrix} x_1 - y_1 \\ x_2 - y_2 \\ \vdots \\ x_n - y_n \end{bmatrix}_{n\times 1}$$

$$= Z' \cdot Z$$

$$= Z' I_{n\times n} Z$$

$$I_{n\times n} = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & 0 \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ 0 & 0 & \cdots & & 1 \end{bmatrix}_{n\times n}$$

$$d_2^{\,2} = \begin{bmatrix} x_1 - y_1 & x_2 - y_2 & \cdots & x_n - y_n \end{bmatrix} \begin{bmatrix} x_1 - y_1 \\ x_2 - y_2 \\ \vdots \\ x_n - y_n \end{bmatrix}$$

$$d_2^2 = ((150,4),(130,3))$$

$$= 400 + 1$$ one feature dominates even if they aren't that important for classification

$$= 401$$

$$= Z' I_{n \times n} Z$$

$$I_{n \times n} = \begin{bmatrix} \lambda_1 & 0 & 0 & \cdots & 0 \\ 0 & \lambda_2 & 0 & 0 \cdots & 0 \\ 0 & 0 & \lambda_3 & \cdots & 0 \\ 0 & 0 & \cdots & \cdots & \lambda_n \end{bmatrix}_{n \times n}$$ weights to different dimensions.

$$d_2^2 = \sum_{i=1}^{n} \lambda_i (x_i - y_i)^2$$

$$d = Z' A_{n \times n} Z$$

A is symmetric, positive definite matrix

- all eigenvalues > 0

- $|A| > 0$

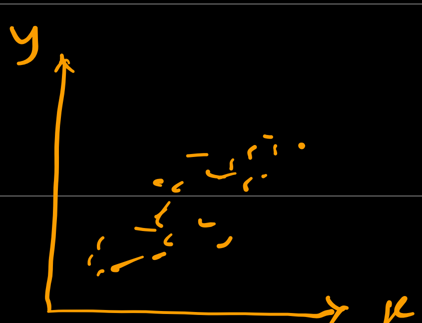
- A^{-1} also be positive definite matrix.

$\Sigma \rightarrow$ Var/Covariance Matrix → padhlena

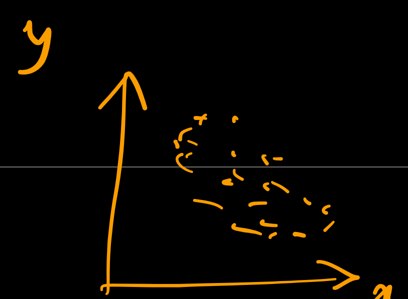
Dispersion Matrix.

$$Var(x) = \frac{1}{n} \sum_{i=n}^{n} (x_i - \mu_x)^2$$

$$Cov(x, y) = \frac{1}{n} \sum_{i=1}^{n} (x_i - \mu_x)(y_i - \mu_y)$$



$Cov(x, y) > 0$



$Cov(x, y) < 0$

$$\begin{array}{c|cccc} & f_1 & f_2 & \cdots & f_D \\ \hline P_1 & & & & \\ P_2 & & & & \\ P_3 & & & & \\ \vdots & & & & \end{array}$$

- Dispersion Matrix / Variance-Covarience Matrix.

$$\Sigma = \begin{array}{c} \\ f_1 \\ f_2 \\ f_3 \\ \vdots \\ f_D \end{array} \begin{array}{c} \begin{array}{ccccc} f_1 & f_2 & f_3 & \cdots & f_D \end{array} \\ \begin{bmatrix} cov(f_1, f_1) & cov(f_1, f_2) & \cdots & & \cdots \\ & & & & \\ & & & & \\ & & & & \\ & & & & \end{bmatrix} \end{array}$$

$$cov(f_1, f_1) = var(f_1)$$

$$cov(f_1, f_2) = cov(f_2, f_1)$$

- Symmetric positive definite matrix
- Classwise dispersion matrix was used in mahalanobolis distance

ex: → Only class 1 elements are used in the dispersion matrix.

ex:

$$X = \begin{array}{cc} f_1 & f_2 \\ \left[\begin{array}{cc} 150 & 55 \\ 160 & 60 \\ 170 & 65 \end{array}\right] \\ 160 & 60 \end{array}$$

$$\Sigma = \begin{bmatrix} 200/3 & 100/3 \\ 100/3 & 50/3 \end{bmatrix}_{2\times 2}$$

$$\Sigma = \frac{1}{3}\begin{bmatrix} 200 & 100 \\ 100 & 50 \end{bmatrix}$$

Y = mean subtraction of X

$$= \begin{bmatrix} -10 & -5 \\ 0 & 0 \\ 10 & 5 \end{bmatrix}$$

$$\Sigma = Y^T Y$$

$$= \begin{bmatrix} -10 & 0 & 10 \\ -5 & 0 & 5 \end{bmatrix} \begin{bmatrix} -10 & -5 \\ 0 & 0 \\ 10 & 5 \end{bmatrix}$$

$$= \begin{bmatrix} 200 & 100 \\ 100 & 50 \end{bmatrix}$$

Classification

Supervised

- → A training set
- → Prior probability and pdf. of classes are given
- →

- **Prior Probability** →

- **Probability Density Function** →

$$G_D = \frac{1}{(2\pi)^{D/2} |\Sigma|^{1/2}} \exp^{-\frac{1}{2}\{(x-\mu)^T \Sigma^{-1}(x-

$$G_1 = \frac{1}{\sqrt{2\pi}\,\sigma} \exp -\frac{1}{2} \frac{(x-\mu)^2}{\sigma^2}$$

P_i , $p(x/w_i)$

↳ 1) Bayes' Classifier →

↳ Let there be M classes $(M \geq 2)$

- The prior probability be $P_1, P_2, \ldots, P_m$ where $P_i > 0$ and $\sum_{i=1}^{n} P_i = 1$

- The probability density function be $p_1(x), p_2(x), \ldots, p_n(x)$ are given.

Assign new test pattern x_0 to class i

$$P_i \, p_i(x_0) \geq P_j \, p_j(x_0) \quad \forall \; i, j \quad i \neq j$$

Resolve ties arbitrarily.

$P_1 \quad p_1(x_0)$

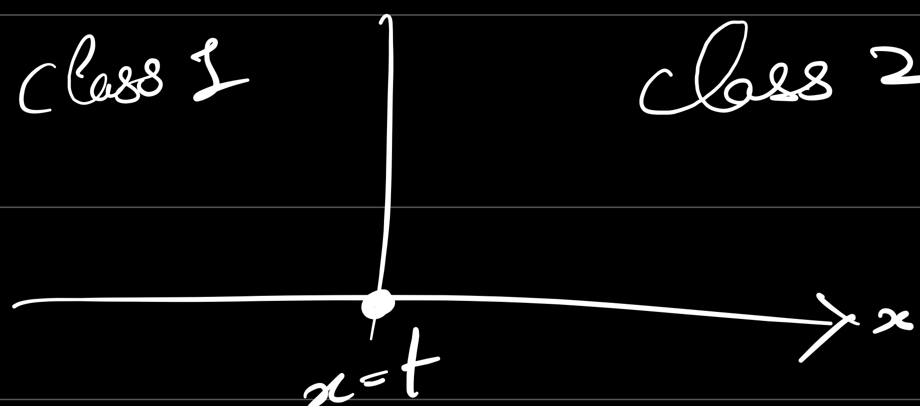
$P_2 \quad p_2(x_0)$

$\boxed{P_3 \times p_3(x_0)}$ → if highest then assign it in class 3.

$\vdots$

$P_M \quad p_M(x_0)$

We are looking for threshold value



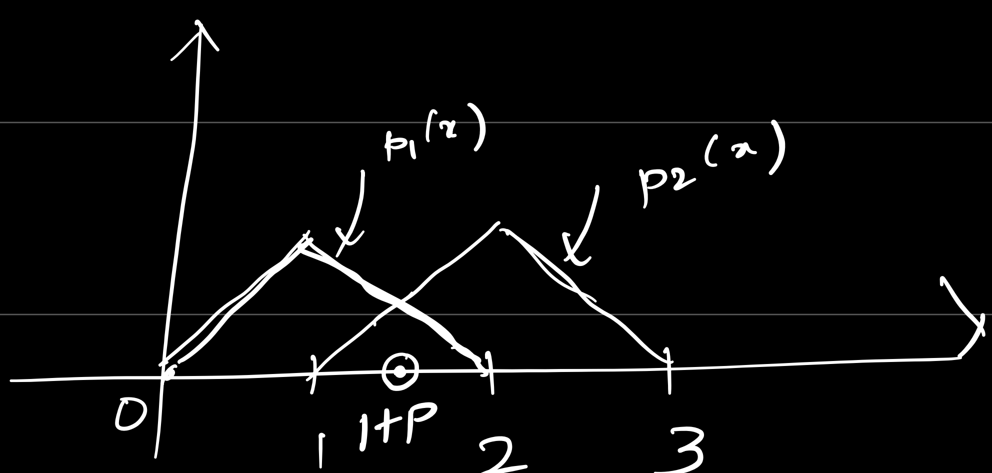
Example

$M = 2$

$P_1 = P \qquad P_2 = 1 - P$

$$p_1(x) = \begin{cases} x & 0 \le x \le 1 \\ 2 - x & 1 < x \le 2 \\ 0 & \text{otherwise} \end{cases}$$

$$p_2(x) = \begin{cases} x + 1, & 1 \le x < 2 \\ 3 - x & 2 \le x \le 3 \\ 0 & \text{otherwise} \end{cases}$$



- Bayes' Classifier

M classes $\quad M \geq 2$

Prior prob $P_i \quad \forall\, i = 1$ to M

$$P_i > 0 \text{ and } \sum_{i=1}^{M} P_i = 1$$

pdf

$p_i(x) : \forall\, i = 1$ to M

x_0 be in class i

$$P_i\, p_i(x_0) \geq P_j\, p_j(x_0)$$

$\forall\, i, j$

$i \neq j$

Case I $\quad 0 \leq x \leq 1$

$$P_1 = p \quad P_2 = 1 - p$$

$$p_1(x) = x \qquad p_2(x) = 0$$

$$P_1\, p_1(x) > P_2\, p_2(x)$$

x be in class 1.

Case II $\quad 2 \leq x \leq 3$

$$p_2(x) = 0 \qquad p_2(x) = 3 - x$$

$$P_2\, p_2(x) \geq P_1\, p_1(x)$$ | x be in Class 2.

Case III $\quad 1 \leq x \leq 2$

x be in class 1

$$P_1\, p_1(x) \geq P_2\, p_2(x)$$

$$P(2 - x) \geq (1-P)(x-1)$$

$$\Rightarrow \quad x \leq 1 + P$$

---

x be in class 1 if $x \leq 1+P$
class 2 is $x > 1+P$.

$\Omega = 0$ to 3.

$\Omega_1 = 0$ to $1+p$

$\Omega_2 = 1+p$ to 3.

misclassification probability / Error.

$$= \sum_{i=1}^{M} \int_{\Omega_i^c} P_i \, p_i(x)$$

$$E = \int_{\Omega_1^c} P_1 \, p_1(x)\,dx + \int_{\Omega_2^c} P_2 \, (p_2(x))\,dx$$

$$= p \int_{1+p}^{3} p_1(x)\,dx + (1-p) \int_{0}^{1+p} p_2(x)\,dx.$$

$$\int_{1+p}^{2} p(2-x)\,dx + \int_{1}^{1+p} (1-p)(3-x)\,dx$$

$$= p\left[2x - x^2/2\right]_{1+p}^{2} + (1-p)\left[\tfrac{x^2}{2} - x\right]_{1}^{1+p}$$

$$E_B = p(1-p)/2$$

Error for any other classifier will be high ex.

$$E_2 = p\int_{1.25}^{2}(2-x)\,dx + (1-p)\int_{1}^{1.25}(x-1)\,dx$$

Baye's Classifier gives minimum missclassification probability.

- uci - repository

IRIS

150 patterns

4 attributes

3 classes

Assume Gaussian Distribution

$$P_1 \propto p_1 \Rightarrow \mu_1 ,\ \Sigma_1$$

$$P_2 \propto p_2 \Rightarrow \mu_2 \ \Sigma_2$$

$$P_3 \propto p_3 \text{ — } \mu_3 \ \Sigma_3$$

$x_0 = [\_ \ \_ \ \_ \ \_] \ [\ ]$

class label