# Machine learning

Learning = Training + Testing.

Weighted avg b/w GPA and credits,
it is called CGPA.

Moving Avg: Study of Population
Set Theoretic Approach to Probability, &
RY approach to Probability

(1. Trial and an Event:

Event is represented by A.

Event that happens for sure.

Appearing Head / Tail – Event.

Tossing a coin – Trial.

Event must have outcome.

(2. Sample Space: $\{1, 2, 3, 4, 5, 6\}$.

↓ Event of 1 appearing = $\{1\}$.

Collection of all events under consideration.

(3. $\text{Prob}(A) = \dfrac{n(A)}{n(S)}$

Axioms:

→ If $i$ is subscripted to probability $p_i \Rightarrow 0 \leq p_i \leq 1 \; \forall i$.

→ $\sum_i p_i = 1$

→ cumulative distribution function $F(x) = P[X < x]$

(4. Addition Theorem of Probability :-

$$P[A \cup B] = P[A] + P[B] - P[A \cap B]$$

$A \cap B$ : simultaneous happening

$A \cup B$ : individual happening

(5. Boole's Inequality :

(6 Baye's Theorem :

Succeeding value depends on previous values.

Statement : If $A_1, A_2, A_3 \dots A_n$ are n events & B is any event occurring along with some of $A_i$'s then :

$$P[A_i / B] = \frac{P[A_i] \, P[B/A_i]}{\sum_{j=1}^{n} P[A_j] \, P[B/A_j]}.$$

conditional probability :

let A & B be any two events :

$$P[A|B] = \frac{P(A \cap B)}{P(B)} = \frac{P[A] \, P[B/A]}{P[B]}$$ where $P(B) \neq 0$.

$$\underbrace{P(A|B)}_{\text{posterior}} \propto \underbrace{P(A)}_{\text{MAP}} \underbrace{P(B/A)}_{\text{maximum likelihood}}$$

Random variable Approach.

→ Discrete Random variable

→ continuous Random variables.

Random variable: It is a variable which can correspond to a ~~regd~~ real number in its sample space.

$X(S) \in \mathbb{R}$

$X$: random variable

Discrete random variable is a RV where the outcome space is having ~~fin~~ finite or countable no. of pts. Eg: No. of steps covered by me. from one pt to another.

Continuous Random variable is a variable whose outcomes belongs to, to have uncountably many forms.

Eg: points cover by cycle to move from one pt to another.

Axioms of Probability:

$0 \le p_i \le 1$

$\sum p_i = D$

CF $P[x<n] = E(\alpha)$

Scanned with CamScanner

Discrete Distribution

(1. Binomial Distribution

(2. Poisson Distribution

Other

(3. Normal Distribution

(4. Gaussian Distribution

What are the 3 sources of uncertainty.

(1. ~~Inherent~~ stochasticity
Inherent

(2. Incomplete observability.

(3. Incomplete Modelling.

For modelling of any physical security models.

Incomplete observability :

Cumulative Distribution function:

It is defined as $F(x) = P[X < x] = \begin{cases} \int_{-\infty}^{x} f(x)dx & \text{if } X \text{ is continuous} \\ \sum_{i=1}^{p_i} p(x_i) & \text{if } X \text{ is discrete} \end{cases}$

Marginal and Conditional probability Mass & Density functions.

Let $x, y$ be a two dimensional random variable, whose combined probability ~~random~~ mass function is defined as $p(x_i, y_j)$ in discrete case.

$$\sum_{i=1}^{m} \sum_{j=1}^{n} p(x_i, y_j) = 1$$

If $x, y$ is discrete, the marginal probability mass function of $x$ is defined as

$$p(x_i) = \sum_{j=1}^{n} p(x_i, y_j)$$

$$q(y_j) = \sum_{i=1}^{m} p(x_i, y_j)$$

Note: If $p(x_i, y_j) = p(x_i)\, q(y_j)$ then $x$ and $y$ are said to be independent.

Diff. b/w independent events & mutually exclusive events.

Train from Tricky to Thanjavur ] Independent
Train from Vpm to Chennai ] events.

If $X, Y$ is a pair of continuous random variable, $g(x) = \int_{-\infty}^{\infty} f(x, y)\, dy$.

marginal density $h(y) = \int_{-\infty}^{\infty} f(x, y)\, dx$.

Note: If $f(x, y) = g(x)\, h(y)$; $x$ and $y$ are said to be independent.

Defenition of conditional PMF's and Densities:

Let $X(x_i,\ i = 1 \ldots m)$
$Y(y_j,\ j = 1 \ldots n)$

be a discrete

$$p(x_i / y_j) = \frac{P[X = x_i,\ Y = y_j]}{q(y_j)} \text{ where } q(y_j) \neq 0.$$

$$q(y_j / x_i) = \frac{P[X = x_i,\ Y = y_j]}{p(x_i)};\ p(x_i) \neq 0$$

For continuous case:

$$g(x/y) = \frac{F(x,y)}{h(y)} \; ; \; h(y) \neq 0$$

$$h(y|x) = \frac{f(x,y)}{g(x)} \; ; \; g(x) \neq 0.$$

State chain Rule of conditional Probability

$$P[a, b, c] = P[a|b,c] \times P[b,c]$$

P (first event occurs | b, c already occured)

High = [(A+B) + ABSOLUTE (A-B)] /2

$$= P[a/b,c] \cdot P(b/c)\, P(c)$$

$$P[x_1, x_2, \ldots x_n] = P[x_1] \prod_{i=2}^{n} P(x_i / x_2, x_3 \ldots x_{n-1})$$

product (like $\Sigma$)

Q1 A random process gives measurements of $x$ between 0 and 1 with PDF $f(x) = \begin{cases} 12x^3 - 21x^2 + 10x & 0 \leq x \leq 1 \\ 0 & \text{otherwise} \end{cases}$

(i) $P[X \leq 1/2]$ (ii) $P[X > 1/2)$ (iii) find $k$: $P[X \leq k] = 1/2$

Solution:

verification if $F(X,Y)$ is PDF or not

$$\int_0^1 12x^3 - 21x^2 + 10x$$

$$= \cancel{12}^{3}[1/\cancel{4}] - \cancel{21}^{7}[1/\cancel{3}] + \cancel{10}^{5}[1/\cancel{2}]$$

$$= 3 - 7 + 5 = 1$$

As $P = 1$; $f(X,Y)$ is a PDF

$$P[X \leq a] = \int_0^a 12x^3 - 21x^2 + 10x$$

$$P[X < 1/2] = \int_0^{1/2} 12x^3 - 21x^2 + 10x$$

$$= \frac{\cancel{12}^{3}x^4}{\cancel{4}} - \frac{\cancel{21}^{7}x^3}{\cancel{3}} + \frac{\cancel{10}^{5}x^2}{\cancel{2}}$$

$$= 3[1/16 - 0] - 7[1/8] + 5[1/4]$$

$$= \frac{3}{16} - \frac{7}{8} + \frac{5}{4}$$

$$= \frac{3 - 14 + 20}{16} = \frac{23-14}{16}$$

$$= 9/16.$$

$$P(X > 1/2) = \int_{1/2}^1 12x^3 - 21x^2 + 10x$$

$$= \left[3x^4 - 7x^3 + 5x^2\right]_{1/2}^{1}$$

$$= (3 - 7 + 5) - [9/16] \Rightarrow 1 - \frac{9}{16}$$

$$= 16 - 9/16 = 7/16$$

Scanned with CamScanner

$P[X \leq K] = 1/2$

$$\int_0^K f(x)\,dx = 1/2$$

$$3K^4 - 7K^3 + 5K^2 = 1/2$$

$$K^2[3K^2 - 7K + 5] = 1/2$$

The PDF of $(x, y)$ is given by $f(x,y) = \begin{cases} e^{-(x+y)} & x, y > 0 \\ 0 & \text{elsewhere} \end{cases}$

Find $P[1/2 < X < 2 ; 0 < Y < 4]$

$$\Rightarrow \int_{1/2}^{2} \int_0^4 f(x,y)\,dy\,dx = \int_{1/2}^{2} [-e^{-y}]_0^4 e^{-x} dx$$

$$= -\int_{1/2}^{2} [e^{-4} - 1]\, e^{-x} = -(e^{-4} - 1)(-e^{-x})_{1/2}^{2}$$

$$= (e^{-4} - 1)\,(e^{-2} - e^{-1/2})$$

consider PDF of $(x, y)$ as $F(x,y) = \begin{cases} 2 & 0 < x < 1, 0 < y < x \\ 0 & \text{elsewhere} \end{cases}$

Find a) Marginal density

b) conditional density.

$$g(x) = \int_0^x f(x,y)\,dy = \int_0^x 2\,dy = 2x$$

$$h(y) = \int_0^1 2\,dx = 2$$

$$g(x/y) = \frac{f(x,y)}{h(y)} = \frac{2}{2} = 1$$

$$h(y/x) = \frac{f(x,y)}{g(x)} = \frac{2}{2x} = \frac{1}{x}$$

$f(x,y) \neq g(x)\, h(y)$

$\therefore$ X & Y are <u>not</u> independent.

Two random variables X and Y

$$f(x,y) = \begin{cases} A e^{-(2x+y)} & x, y > 0 \\ 0 & \text{otherwise} \end{cases}$$

(1) Evaluate A.

(2) Find marginal PDF

(3) Find condn PDF

(4) Find cumulative distribution function

To Evaluate A:

Total Probability $= 1$

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} F(x,y)\, dx\, dy = 1$$

$$\int_{0}^{\infty} \int_{0}^{\infty} A e^{-(2x+y)}\, dx\, dy = 1$$

$$A \left[\frac{e^{-2x}}{-2}\right]_0^{\infty} \left[\frac{e^{-y}}{-1}\right]_0^{\infty} = 1$$

$$\frac{A}{2} [0-1] [0-1] = 1.$$

$$A = 2$$

$$g(x) = \int_{-\infty}^{\infty} f(x,y)\, dy = \int_{0}^{\infty} 2e^{-(2x+y)} dy$$

$$= 2e^{-2x} \left[-e^{-y}\right]_0^{\infty}$$

$$= -2e^{-2x} [0-1] = 2e^{-2x}$$

Scanned with CamScanner

$$h(y) = \int_{-\infty}^{\infty} f(x,y)\,dx$$

$$= \int_{0}^{\infty} 2e^{-(2x+y)}\,dx$$

$$= 2e^{-y}\left[\frac{e^{-2x}}{-2}\right]_0^{\infty} = -e^{-y}[0-1] = e^{-y}$$

$$g(x)\,h(y) = 2e^{-(2x+y)} = f(x,y)$$

$\therefore$ X & Y are independent

$$g(x/y) = \frac{F(X,Y)}{h(y)} = \frac{Ae^{-(2x+y)}}{e^{-y}}$$

$$h(y/x) = \frac{F(X,Y)}{g(x)} = \frac{Ae^{-(2x+y)}}{2e^{-2x}}$$

<u>CDF</u> $F(X,Y) = P[X<x,\ Y<y]$

$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} 2e^{-(2x+y)}\,dx\,dy.$$

Scanned with CamScanner

$$F(x,y) = \begin{cases} 8xy & 0 \le x \le y,\ 0 \le y \le 1 \\ 0 & \text{otherwise} \end{cases}$$

Check whether X and Y are independent of each other or not.

$$g(x) = \int_{-\infty}^{\infty} F(x,y)\,dy$$

$$= \int_0^1 8xy\,dy$$

$$= 8x \left[\frac{y^2}{2}\right]_0^1 \Rightarrow 4x[16-0]$$

$$= 4x[16] = 64x$$

$$h(y) = \int_{-\infty}^{\infty} F(x,y)\,dx$$

$$= \int_0^y 8xy\,dx \quad y\left[\frac{8x^2}{2}\right]_0^y$$

$$= y\left[\frac{8y^2}{2}\right] = \underline{\underline{4y^3}}$$

$$g(x)\,h(y) = 64x \times 4y^3 = 254\,y^3 x$$

$\therefore$ X & Y are not independent.

Two discrete RV's have PDF:

$$p(x,y) = \frac{\lambda^x e^{-\lambda} p^y (1-p)^{x-y}}{y!\,(x-y)!} ; \quad y = 0,1,2 \cdots y$$

$$\lambda > 0 ; \quad 0 < p < 1$$

Find marginal and conditional PMF's.

Solution:

$$g(x) = \sum_{y=0}^{x} \frac{\lambda^x e^{-\lambda} p^y (1-p)^{x-y}}{y!\,(x-y)!}$$

$$= \lambda^x e^{-\lambda} \sum_{y=0}^{x} \frac{p^y (1-p)^{x-y}}{y!\,(x-y)!}$$

$$= \frac{\lambda^x e^{-\lambda} x!}{x!} \left[ \frac{(1-p)^x}{x!} + \frac{p(1-p)^{x-1}}{1!\,(x-1)!} + \frac{p^2(1-p)^{x-2}}{2!\,(x-2)!} + \cdots + \frac{p^x(1-p)^0}{x!\,0!} \right]$$

$$= \frac{e^{-\lambda}\lambda^x}{x!} \left[ (1-p)^x + xp(1-p)^{x-1} + \frac{x(x-1)p^2(1-p)^{x-2}}{2!} + \cdots \frac{\cancel{x!}\, p^x (1-p)^0}{\cancel{x!}} \right]$$

$$= \frac{e^{-\lambda}\lambda^x}{x!} [p + (1-p)]^x = \frac{e^{-\lambda}\lambda^x}{x!}$$

$$h(y) = \frac{e^{-\lambda}\lambda^y}{y!}$$

Consider a random exp of tossing two honest dies. Find the marginal probability distribution. Also find $P[2 < X \leq 3,\ 3 \leq Y < 6]$

$X \leq 3$

| Y \ X | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|
| 1 | 1/36 | 1/36 | 1/36 | 1/36 | 1/36 | 1/36 |
| 2 | 1/36 | 1/36 | 1/36 | 1/36 | 1/36 | 1/36 |
| 3 | | | 1/36 | | | |
| 4 | | | 1/36 | | | |
| 5 | | | 1/36 | | | |
| 6 | | | | | | |

$P(Y=1) = 1/6$

$P(Y=2) = 1/6$

$\Sigma$.

$$P[2 < X \leq 3,\ 3 \leq Y < 6] = \frac{1}{36} + \frac{1}{36} + \frac{1}{36}$$

$$= 3/36 = 1/12$$

Are X and Y independent,

$$P(X=1, Y=1) = 1/36 = P(X=1)\, P(Y=1)$$
$$= 1/36.$$

$\therefore P(X=i, Y=j) = P(X=i) \cdot P(Y=j)$ for all $i$ and $j$ And thus

$\therefore$ X and Y are independent.

Tasks T :- particular task is to be
Experience E completed, ~~someone~~ with the
Performance P past experience set to a
particular performance level.

Experience - earlier data.
Task - present data ( Instantaneous Incoming Data or the machine data)

performance: moving towards which target.

$\{1/n\} \rightarrow$ converges at 0 (limit point)

perceptron - more than learning Alg. It can perceive through the data itself.

How a task is to be carried out, with ~~a~~ past experience ~~to~~ set to a performance

ML Alg. address both soft computing as well as Scientific computing.

Find the formula for the probability distribution of no of heads when a fair coin is tossed four times.

No. of tosses of a coin : $n = 4$

$P[X = x] = nC_x \, p^x q^{n-x}$ ↗ doubt

Let X be discrete RV denoting the no. of heads,

let P be prob of head to appear.

∴ probability of Tail to appear $= 1 - p = q$

$= 4C_x \cdot p^x q^{\cancel{4}-x}$ ; $x = 0, 1, 2, 3, 4$

$P[X=0] = 4C_0 \, p^0 q^4$

$= 1 \cdot 1 \cdot \frac{1}{2^4} = \frac{1}{2^4} = \frac{1}{64}$

$P[x=1] = 4C_1 \, p^1 \, q^3 =$

$= 4C_1 \, p^1 \frac{1}{2^3}. \quad 4C_1 \frac{1}{2} \frac{1}{2^3} = \frac{4}{2^4}$

$P[X=2] = 4C_2 \, p^2 \, q^2$

$= 4C_2 \, p^2 \frac{1}{2^2} = 4C_2 \frac{1}{4} \times \frac{1}{4}$

$P[X=3] = 4C_3 \, p^3 \frac{1}{2^1}$

$P[X=4] = 4C_4 \, p^4 \frac{1}{2^0}$

Consider the experiment of tossing a coin, the two event of the space being the occurrence of head or tail. Assign prob. P and Q for head & Tail resptly. Define RV X by $X(h) = 1$ and $X(t) = 0$. Determine probability fn $f(x)$ and $F(x) \to$ CDF (cumulative Distribution function).

Occurrence of head $P(1) = 1/2$

Occurrence of Tail $P(0) = 1/2$

$P(X = 0) = \frac{1}{2}$ $\quad P[X = 1] = \frac{1}{2}$. $\Rightarrow$ Probability Distribution.

$$F(x) = \begin{cases} 0 & x \notin \{0, 1\} \\ 1/2 & x \leq 0 \\ 1 & x \leq 1 \end{cases} \Rightarrow \text{Cumulative Distribution.}$$

2 Dice are tossed. X denotes RV representing the sum of numbers appearing on the dice. Find the probability distribution of X. Also find Cumulative Distribution of X.

Range of for RV X / RV space $\Rightarrow \{2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12\}$.

Sample space $S = \{(1,1), (1,2) \cdots (1,6)$

$n(S) = 36$.

$P[X=2] = 1/36$ $P[X=6] = 5/36$

$P[X=3] = 2/36$ $P[X=7] = 6/36$

$P[X=4] = 3/36$ $P[X=8] = 5/36$

$P[X=5] = 4/36$ $P[X=9] = 4/36$

$P[X=10] = 3/36$ $P[X=11] = 2/36$

$P[X=12] = 1/36.$

4 (2,2 / 1,3 / 3,1)

CD: $P[X] = \begin{cases} 0 & X \leq 2 \\ 1/36 & X \leq 2 \\ 3/36 & X \leq 3 \\ 6/36 & X \leq 4 \\ 10/36 & X \leq 5 \\ \vdots & \\ 1 & X \leq 12 \end{cases}$

consider the expt of tossing a fair four coin four times. Define $X=0$ if 0 or 1 head appear. $X=1$ if two heads appear. $X=2$ if 3 or 4 heads appear. Find the probability Function, mean & variance of X.

$S = \{(H,T,T,T) \ldots \}$ $n(S) = 2^4 = 16$

$X: \{0 \quad 1 \quad 2\}$

0 → 0 or 1; 1 → 2H; 2 → 3 or 4 H

$P[X=0] = 4C_0 (1/2)^0 (1/2)^4$

$+ 4C_1 (1/2)^1 (1/2)^3 = 5/16$

$P[X=1] = 4C_2 (1/2)^2 (1/2)^2 = 6/16$

$P[X=2] = [4C_3 (1/2)^3 (1/2)^1] + [4C_4 (1/2)^4 (1/2)^0]$

$5/16$

$=$

Mean:

$\text{Mean} = E(X) = \sum x\,P(x)$

$= 0 \times 5/16 + 1(6/16) + 2 \times \frac{5}{16}$

$= \frac{6}{16} + \frac{10}{16} = \frac{16}{16} = 1$

Variance: $E(X)^2 = \sum x^2 P(x)$

$= (0)^2 \times 5/16 + (1)^2\, 6/16 + 4 \times \frac{5}{16}$

$= \frac{6}{16} + \frac{20}{16} = \frac{26}{16} = \frac{13}{8}$ //

$Var(X) = E(X^2) - [E(X)]^2$

$= \frac{13}{8} - 1 = 5/8$ //

$\sigma = STD(X)$

$= \sqrt{var\,x}$

$\sigma = \sqrt{5/8}$ //

11/8/22

Define Expectation of a RV:

Let X be a random variable

E(X) is defined as $E(X) = \begin{cases} \sum_i x_i p(x_i) & \text{if X is discrete} \\ \int_{-\infty}^{\infty} x f(x)\,dx & \text{if X is continuous} \end{cases}$

Variance :

Variance of X is defined as

$V(X) = E(X^2) - [E(X)]^2$ // For Pblms solving //

By defenition variance is given by

$E[(X - E(X))^2] = Var(X).$

Covariance : // need to be defined w.r.to PCA //

List out common probability distributions and write a short note on each one of them

(1. Bernoulli Distribution:

[B.T ⇒ [Has only 2 outcomes]]

$X \in \{0,1\}$ Let X be a binary RV with probability of success be '$\theta$'. Let X follows Bernoulli distribution of parameter $\theta$ which is a PMF

$$X \sim \text{Ber}(\theta)$$

PMF : $\text{Bern}[x/\theta] = \begin{cases} \theta & x=1 \\ 0 & x=0 \end{cases}$

Note: This is a special case of ~~Binomi~~ ~~Bernoullian~~ Binomial distribution with $n=1$

$1 p \cdot q^{1-x}$  $n \Rightarrow 0, 1$

If $n=0$ we get only q
$n=1$ we get only p.

Binomial Distribution $\text{Bin}[k/n,\theta] = {}^nC_k \, \theta^k (1-\theta)^{n-k}$

Bin. Dist. is the generalised for Bernoullian distribution

(2. Multinomial Distribution:

$$\text{Mu}(x/n,\theta) = \binom{n}{x_1, x_2 \cdots x_k} \prod_{j=1}^{k} \theta_j^{x_j}$$

(multinomial coeff)

where $\theta_j$ : probability that side j source up

and $\binom{n}{x_1, x_2 \cdots x_k} = \dfrac{n_j}{x_1!\, x_2! \cdots x_k!}$

Gaussian Distribution

$$N[x \mid \mu, \sigma^2] = \frac{1}{\sigma\sqrt{2\pi}} e^{-1/2\left[\frac{x-\mu}{\sigma}\right]^2}$$

$\mu = E(x)$

$\sigma$ std deviation, $\sigma^2 = var(x)$

CDF of Gaussian Random variable

$$\int_{-\infty}^{x} N[z/\mu ; \sigma^2]\, dz$$

In terms of Error function:

$$\phi(x, \mu, \sigma) = \frac{1}{2}(1 + erf(z/\sqrt{2})]$$

$$\text{where } erf(z) = \sqrt{\frac{2}{\pi}} \int_0^z e^{-t^2} dt$$

Error fn is defined for ill posed data

Ill posed data

Well posed data $n < 29$.

$$\Gamma_n = \int_0^\infty e^{-x} x^{n-1} dx \Rightarrow$$

When erf is integrated over $-\infty$ to $\infty$, the probability is 1.

Prop of Normal Distribution:

(1. Bell shaped curve - normal Distri

(2. Mean: $\mu$, Std Deviation = $\mu$

$(\mu - \sigma)(\mu + \sigma) \Rightarrow$ Area covered by the two ordinates $x = \mu - \sigma$ & $\mu + \sigma$

Normal distion is symmetrical over $x = \mu - \sigma$ & $x = \mu + \sigma$.

Laplace transforms are used in Control Algorithm.

$L[f'(t)] = sF(s) - f(0)$ // Derivative is converted into ydt //

Human perception

nyc perception

$L\left[\int_0^t f(t)\,dt\right] = \frac{F(s)}{s}$.

$L[t\,f(t)] = -F'(s)$

$L\left[\frac{f(t)}{t}\right] = \int_s^\infty F(s)\,ds$.

Reasons of wide use of Gaussian Distribution.

Scanned with CamScanner

Exponential Distribution:

$$p(x, \lambda) = \lambda e^{-\lambda x} \quad ; \quad x \geq 0$$

It is to have a sharp point at $x = 0$

Laplace Distribution:

$$\text{Lap}[x, \mu, \gamma] = \frac{1}{2\gamma} e^{-\left[\frac{(x-\mu)}{\gamma}\right]}$$

To accomodate a sharp peak of probability mass at an arbitrary point $U$.

Direction Distribution & Empirical Distribution:

(1. To accomodate all probability mass around a single pt, direct delta function is defined as

$$p(x) = \delta(x-\mu) \; ; \quad \delta(x = \begin{cases} \infty & x = 0 \\ 0 & x \neq 0 \end{cases}$$

* Shifted by $-U$ to obtain infinitely narrow and
* Infinite peak of probability mass when $x = \mu$

$$\delta(x-\mu) = \begin{cases} \infty & ; \quad x = \mu \\ 0 & ; \quad x \neq \mu \end{cases}$$

Scanned with CamScanner

eg: Hammering
convergence of rays in convex lens

Direct $\delta$ function: to focus the mass tot at a single point.

Empirical Distribution:

Given a data set $\{x_1, x_2, x_3 \ldots x_n\}$.

EMR $P_{emp}(A) = \frac{1}{N} \sum_{1}^{\infty} \delta_{x_i}(A)$

$$\delta_x(A) = \begin{cases} 0 & x \notin A \\ 1 & x \in A \end{cases}$$

$$p(x) = \sum_{1}^{\infty} w_i \, \delta_{x_i}(x) \text{ where } 0 \le w_i \le 1, \; \sum_{i=1}^{N} w_i = 1$$

How to mix two Distributions:

Mixtures of Distributions.

Let $k$ be scalar categorical random variable
Let $X$ be a dummy encoding.

$$X = [(x=1)\;(x=2)\;\ldots\;(x=k)]$$

If $k = 3$, it encodes states 1, 2, & 3 as $(1,0,0)$ $(0,1,0)$ $(0,0,1)$

This is called One-hot encoding

with these notations, multinoulli distribution is defined as

$$M_u(x \mid 1, \theta) = \prod_{j=1}^{K} \theta_j^{(x_j = 1)}$$

Difference b/w multinomial & multinoulli distribution

It is the absence of multinomial coefficient in multinoulli Distribution.

(1. A mixture distribution is made up of several component Distribution

(2. On each trial, the choice of component distribution generates a sample determined by a sampling component identity from multinoulli Distribution.

Training Eg

Probability

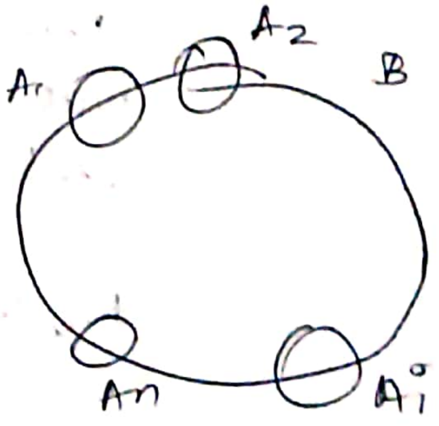
$$P(x) = \sum_i P[c=i] \; P[x|c=i]$$

where $P(c)$ is multinoulli Distribution.

In x dist, we donot know $A_1, A_2, A_3 \dots A_n$

Baye's theorem:

$$P[A_i/B] = \frac{P(A_i)\,P[B/A_i]}{\sum_{i=1}^{n} P(A_i)\,P(B/A_i)}$$



(X) 6M Explain the properties of common functions used in Deep Learning Models.

Logistics Sigmoid :

$$\sigma(x) = \frac{1}{1+e^{-x}}$$

25/3/22

(1) $\sigma(x) = \dfrac{1}{1+e^{-x}}$

To prove $\sigma'(x) = \sigma(x)[1-\sigma(x)]$

$$\sigma'(x) = \frac{-1}{(1+e^{-x})^2}[-e^{-x}]$$

$$= \frac{e^{-x}}{(1+e^{-x})^2}$$

$$= \frac{(1+e^{-x})-1}{(1+e^{-x})^2}$$

$$= \frac{1}{(1+e^{-x})} - \frac{1}{(1+e^{-x})^2}$$

$$= \frac{1}{1+e^{-x}}\left[1-\frac{1}{1+e^{-x}}\right] \Rightarrow \sigma(x)[1-\sigma(x)]$$

Soft Plus:

$$\gamma(x) = \log[1+e^{-x}]$$

$$\gamma'(x) = -\sigma(-x)$$

$$\gamma^{-1}(x) = \log[e^{x}-1]$$

$$\gamma(x) = \int_{-\infty}^{x} \sigma(y)\,dy$$

$$\gamma(x) - \gamma(-x) = x$$

$$\sigma(x) = \frac{1}{1+e^{-x}} \qquad \sigma^{-1}(x) = \log\frac{x}{1-x} \quad \forall x \in (0,1)$$

let $y = \sigma(x)$

$$= \frac{1}{1+e^{-x}}$$

$$1+e^{-x} = \frac{1}{y}$$

$$e^{-x} = \frac{1}{y} - 1$$

$$= \frac{1-y}{y}$$

$$-x = \log\left(\frac{1-y}{y}\right)$$

$$x = \log\left(\frac{y}{1-y}\right)$$

As $x$ is dummy

$$\sigma^{-1}(x) = \log\left[\frac{x}{1-x}\right]$$

(2) $1-\sigma(x) = \sigma(-x)$

$$\sigma(x) = \frac{1}{1+e^{-x}} \rightarrow (1)$$

$$\sigma(-x) = \frac{1}{1+e^{x}} \rightarrow (2)$$

$$1-\sigma(x) = 1 - \frac{1}{1+e^{-x}}$$

$$= \frac{\cancel{1}+e^{-x}-\cancel{1}}{1+e^{-x}} = \frac{e^{-x}}{1+e^{-x}}$$

multiply & divide it by $e^{x}$

$$\frac{e^{-x} \cdot e^{x}}{e^{x} + e^{-x} \cdot e^{x}} = \frac{1}{1+e^{x}} = \sigma(-x) \text{ from } ②$$

③ $$\frac{d}{dx} \gamma(x) = \sigma(x)$$

$$\gamma(x) = \log[1+e^{-x}]$$

$$\gamma'(x) = \frac{1}{1+e^{-x}}[-e^{-x}]$$

$$= \frac{-e^{-x} e^{x}}{(1+e^{-x})\, e^{x}}$$

$$= \frac{-1}{e^{x}+1} = -\sigma(-x)$$

④ $$\log \sigma(x) = -\gamma(-x)$$

$$\sigma(x) = \frac{1}{1+e^{-x}}$$

$$\log \sigma(x) = \log(1) - \log(1+e^{-x})$$

$$= 0 - \gamma(x) \rightarrow ①$$

$$\cancel{\gamma(-x) = \frac{h}{1+e^{+x}}}$$

$$\gamma(-x) = \log[1+e^{x}]$$

$$-\gamma(-x) = -\log[1+e^{x}]$$

$$= \log(1+e^{x})^{-1}$$

$$= \log\left(\frac{1}{1+e^{x}}\right)$$

$$= \log 1 - \log(1+e^{x}) = -\gamma(-x)$$

$-\gamma(x) = \log \sigma(x)$

$\log \sigma(x) = \log \left[\frac{1}{1+e^{-x}}\right]$

$= \log 1 - \log(1+e^{-x})$

$= \log(1+e^{-x})^{-1}$

$= \log(1/1+e^{-x})$

$= \log 1 - \log(1+e^{-x})$

$= 0 - \gamma(x)$

To prove: $\gamma^{-1}(x) = -\log[e^{x}-1]$

$\gamma(x) = \log[1+e^{-x}] = y$

$1+e^{-x} = e^{y}$

$e^{-x} = e^{y}-1$

$-x = \log(e^{y}-1)$

$x = -\log(e^{y}-1)$

As $y$ is dummy

$\gamma^{-1}(x) = -\log(e^{x}-1)$

To prove: $\gamma(x) - \gamma(-x) = x$

$$= \log(1+e^{-x}) - \log(1+e^{x})$$

$$= \log\left[\frac{1+e^{-x}}{1+e^{x}}\right]$$

$$= \log\left[\frac{(1+e^{-x})e^{x}}{(1+e^{x})e^{x}}\right]$$

$$= \log\left[\frac{\cancel{e^{x}+1}}{e^{x}\cancel{(1+e^{x})}}\right]$$

$$= \log\left[\frac{1}{e^{x}}\right]$$

$$= \log e^{-x}$$

$$= -x \log e$$

$$= -x.$$

Explain Baye's rule and poor conditioning in numerical computations:

Baye's Rule:

If X and Y are 2 given events

$$P[X/Y] = \frac{P(X)\ P[Y/X]}{P(Y)}$$

$P(x)$: known
$P(Y)$: not known
But $P[Y] = \sum_x P(Y|x) P(x)$

($P(Y)$: total probability)

We donot need to begin with knowledge of $P(Y)$

Poor conditioning:

Conditioning refers to the rapid change in the function behaviour with respect to small changes in the input such condn can be probleamatic because rounding errors in i/p can result in large change in o/p.

Let $f(x) = A^{-1}(x)$ where $A \in R^{n \in R}$ has an eigen value decomposition its condn no. defined by $\max_{i=j} \left| \frac{\lambda_i}{\lambda_j} \right| \rightarrow$ 1

It is the ratio of magnitude of largest & smallest eigen value when 1 is large. matrix inversion is particularly sensitive to error in the i/p.

Poorly conditioned matrices amplify pre existing errors when we multiply by true matrix inverse

A Problem on Total Probability theorem:

An urn contains 10 white balls and 3 Black balls. Another urn contains 3 white balls & 5 black balls. Two balls are drawn at random from the first urn and are placed in second urn. And then one ball is taken at random from the latter. What is the probability that it is a white ball.

Solution:

Urn 1: 10 W, 3 B

Urn 2: 3 W, 5 B

Mode of Transfer:

Transferred ball may be 2W, 2B, 1W 1B.

Let $B_1, B_2, B_3$ be the events of drawing ~~two~~ 2W, 2B, 1W, 1B from urn 1.

$$P(B_1) = \frac{10C_2}{13C_2} = \frac{\frac{10 \cdot 9}{2 \cdot 1}}{\frac{13 \cdot 12}{2 \cdot 1}} = \frac{45}{78}$$

$$P(B_2) = \frac{3C_2}{13C_2} = \frac{3}{\frac{13 \cdot 12}{2 \cdot 1}} = \frac{3}{78}$$

$$P(B_3) = \frac{10C_1 \cdot 3C_1}{13C_2} = \frac{10 \cdot 3}{\frac{13 \cdot 12}{2 \cdot 1}} = \frac{30}{78}$$

Let A be the event of drawing one white ball from urn 2 after transfer

Total Balls in urn 2 = $8 + 2 = 10$

Urn 2: 5 W, 5 B.

$$\therefore P[A/B_1] = \frac{5C1}{10C_1} = \frac{5}{10}$$

$P[A/B_2] = \frac{3}{10}$ | Urn 2 : 3W 7B

$P[A/B_3] = 4/10$ | Urn 2 : 4W, 6B.

$P[A]$: Prob. of drawing 1W ball from Urn 2 after transfer

$$P[A] = P[A/B_1] \cdot P(B_1) + P(A/B_2)\, P(B_2) + P(A/B_3) \cdot P(B_3)$$

$$= \frac{5}{10} \times \frac{45}{78} + \frac{3}{10} \times \frac{3}{78} + \frac{4}{10} \times \frac{30}{78}$$

$$= \frac{225 + 9 + 120}{780} = \frac{354}{780} = \frac{59}{130}.$$

225
129
354

4 | 78
38

59 × 6.
354

A sample problem to use Baye's theorem:

There are three true coins and one false coin with head on both sides. A coin is chosen at random and tossed four times. If head occurs all the four times, what is the probability that false coin has been chosen and used.

Solution:

3 T coins   1 F coins.

Let T be the event of choosing True coin.

$P[T] = 3/4$

Let F be the event of choosing a False coin

$P[F] = 1/4$

Let A be the event ~~of~~ denoting 4 Bernoullian trials (tossing a coin)

$P(A|T) = (1/2)^4$

$P(A|F) = 1$

$$P[F/A] = \frac{P(F)\,[P(A/F)]}{P(A)}$$

$$= \frac{P(F)\,P(A|F)}{P(F)\cdot P(A/F) + P(T)\cdot P(A/T)}$$

$$= \frac{1/4(1)}{1/4(1) + \frac{3}{4}(1/2^4)} = \frac{1/4}{\frac{16+3}{64}} = \frac{16}{19}$$ //

The joint PDF of 2 dimensional random variable $f(x, y) = xy^2 + \frac{x^2}{8}$ determine

(i) $P[X<1]$ (ii) $P[Y<1/2]$ (iii) $P[X>1 | Y<1/2]$

(iv) $P[Y<1/2 | X>1]$ (v) $P[X<Y]$ (vi) $P[X+Y<1]$

Solution:

$P[X<1]$

Two dimensional is given, if the probability is asked for 1 variable, then marginal prob. has to be calculated given that $x \to$ continuous.

$$g(x) = \int_{-\infty}^{\infty} f(x, y)\, dy$$

$$= \int_0^1 \left(xy^2 + \frac{x^2}{8}\right) dy$$

$$= \left[x \cdot \frac{y^3}{3} + \frac{x^2}{8} \cdot y\right]_0^1$$

$$= x\left[1/3 - 0\right] + \frac{x^2}{8}(1) = \frac{x}{3} + \frac{x^2}{8}$$

$$P[X<1] = \int_0^1 \left(\frac{x}{3} + \frac{x^2}{8}\right) dx = \left[\frac{x^2}{6} + \frac{x^3}{24}\right]_0^1$$

$$= 1/6 + 1/24 = 5/24$$

To Find $P[Y<1/2]$:

$$h(y) = \int_{-\infty}^{\infty} f(x, y)\, dx$$

$$= \int_0^1 \left(xy^2 + \frac{x^2}{8}\right) dx$$

$$= \left[\frac{y^2 x^2}{2} + \frac{x^3}{24}\right]_0^1 = \frac{y^2}{2} + \frac{1}{24}$$

$$P[y < 1/2] = \int_0^{1/2} \left(\frac{y^2}{2} + \frac{1}{24}\right) dy$$

$$= \left[\frac{y^3}{6} + \frac{y}{24}\right]_0^{1/2}$$

$$= \frac{1}{6}\left(\frac{1}{8}\right) + \frac{1}{24}(1/2) = \quad 2/48 = 1/24$$

(iii) $P[X > 3/4 \mid Y < 1/2]$

$$P[X > 3/4 \mid Y < 1/2] = \frac{P[X > 3/4 ; Y < 1/2]}{P[Y < 1/2]}$$

$$= \frac{1}{P[Y < 1/2]}\left[\int_{3/4}^{1}\int_0^{1/2}\left(xy^2 + \frac{x^2}{8}\right).dy\ dx\right.$$

$$= \frac{1}{1/24}\left[\int_{3/4}^{1}\left[\frac{xy^3}{3} + \frac{x^2 y}{8}\right]_0^{1/2} dx\right.$$

$$= 1/1/24\left[\int_{3/4}^{1}\left(\left(\frac{x}{24}\right) + \frac{x^2}{16}\right]dx\right. \qquad 9/16 + \frac{27}{64}$$

$$= 1/1/24\left[\frac{x^2}{24\times 2} + \frac{x^3}{16\times 3}\right]_{3/4}^{1} \qquad \begin{matrix} 16\times 4 \\ 64 \\ \frac{36+27}{63} \end{matrix}$$

$$= 1/1/24\left[\left[\frac{1}{48} + \frac{1}{48}\right] - \left[\frac{9/16}{24\times 2} + \frac{27/64}{48}\right]\right.$$

$$= 1/1/24\left[\frac{2}{48} - \left[\frac{63/64}{48}\right]\right.$$

$$= \frac{1}{1/24} \times \frac{1}{48}\left[2 - \frac{63}{64}\right]$$

$$= \frac{128 - 63}{64} = \frac{65}{64}$$

$$= \cancel{24} \times \frac{1}{4\cancel{8}_2} \times \frac{65}{64}$$

(v) $$P[X<Y] = \int_0^y \left(\frac{x}{24} + \frac{x^2}{16}\right)dx$$

$$= \left[\frac{x^2}{48} + \frac{x^3}{48}\right]_0^y = \frac{y^2 + y^3}{48}$$

(vi) $$P[X+Y<1] = \int_0^1 \int_0^{1-y} \left(xy^2 + \frac{x^2}{8}\right) dx\,dy$$

$$= \int_0^1 \left(y^2x^2 + \frac{x^3}{24}\right)\Big|_0^{1-y}$$

$$= \int_0^1 \left[\frac{y^2(1-y)^2}{2} + \frac{(1-y)^3}{24}\right]dy$$

$$= \int_0 \frac{y^2 - 2y^3 + y^4}{2} + \frac{(1-y)^2}{24}\,dy$$

$$= \frac{1}{2}\left(\frac{y^3}{3} - \frac{2y^4}{4} + \frac{y^5}{5}\right) + \frac{(1-y)^4}{-4(24)}\Bigg|_0^1$$

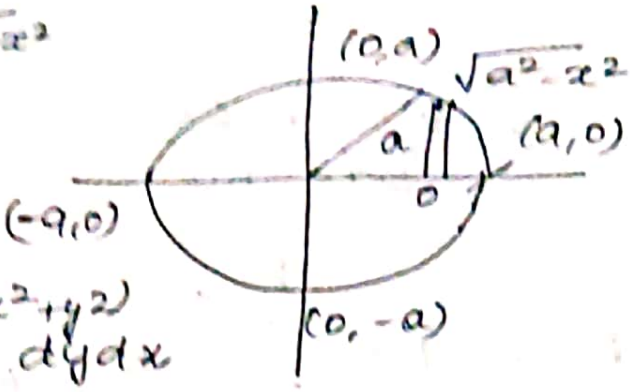
Scanned with CamScanner

$$f(x,y) = \frac{1}{2\pi\sigma^2} e^{-\frac{(x^2+y^2)}{2\sigma^2}}$$

$-\infty < x, y < \infty$ Find $P[X^2+Y^2 \le a^2]$

Soln:

$$P[X^2+Y^2 \le a^2] = 4\int_0^a \int_0^{\sqrt{a^2-x^2}}$$




$$= 4\int_0^a \int_0^{\sqrt{a^2-x^2}} \frac{1}{2\pi\sigma^2} e^{-1/2\sigma^2(x^2+y^2)} dy\,dx \rightarrow (1)$$

Put $x = r\cos\theta$ $y = r\sin\theta$.

$x^2+y^2 = r^2$

$dx\,dy = |J|\,dr\,d\theta$ where $J = \begin{vmatrix} \frac{\partial x}{\partial r} & \frac{\partial y}{\partial r} \\ \frac{\partial x}{\partial \theta} & \frac{\partial y}{\partial \theta} \end{vmatrix}$

$r$.

$$= \begin{vmatrix} \cos\theta & \sin\theta \\ -r\sin\theta & r\cos\theta \end{vmatrix} = r$$

$$(1) \Rightarrow P = \frac{2}{\pi r^2} \int_0^{\pi/2} \int_0^a e^{-r^2/2\sigma^2}\, r\,dr\,d\theta$$

$$P = \frac{2}{\pi\sigma^2} \int_0^{\pi/2} \left( e^{-\frac{r^2}{2\sigma^2}} d\left(\frac{-r^2}{2\sigma^2}\right)(-\sigma^2)\right)$$

$$= \frac{2}{\pi\sigma^2} \int_0^{\pi/2} \left[ e^{r^2/2\sigma^2}(-\sigma^2) \right]_0^a d\theta$$

$$= \frac{2}{\pi}\int_0^{\pi/2}\left[ e^{a^2/2\sigma^2} - 1 \right] d\theta$$

$$= \cancel{\frac{2}{\pi}\left(1 - e^{a^2/2\sigma^2}\right] Q\left(\pi/2\right)}$$

$$= 1 - e^{a^2/2\sigma^2}$$

In a binary commn. s/m a zero or one, is transmitted. Because of noise in the s/m, 0 can be received as 1 with prob. P and 1 can be received as 0 with same prob. P. Assuming that 0 is transmitted with prob. $P_0$ and 1 as $Q_0$ $q_0(1-P_0)$. Find the prob that one was transmitted when 1 is received.

$$P[A/B] = \frac{P(A)\; P(B/A)}{P(A)\, P(B/A) + P(\bar{A}) \cdot P(B/\bar{A})}$$

let A, be an event denoting 1 being transmitted. $\bar{A}$ : 0 being transmitted.

let B be an event denoting 1 being received. $\bar{B}$ : 0 being recieved

<u>Given:</u> $P[B/\bar{A}] = p$ $\quad P[\bar{B}/A] = P$

$P[\bar{A}] = P_0$

$P[A] = Q_0 \Rightarrow (1 - P_0)$

To Find: $P[A/B]$

Solution: $$P[A|B] = \frac{P(A)\ P[B/A]}{P(A)\ P(B/A) + P(\bar{A})\ .\ P(B/\bar{A})}$$

$$= \frac{q_0\ (1-p)}{q_0(1-p) + p_0\ (p)}$$ $\because$ using (2) in Given

$$P[A|B] = \frac{q_0\ (1-p)}{q_0 - pq_0 + p_0 p}.$$

The probability that a student passes an exam is 0.9 given that he studied. Probability that he passes an exam w/o studying is 0.2. Assuming that the Prob that a student studies is 0.75. Given that the student has passed the exam, what is the probability that he studied.

<u>Solution:</u>

let A be an event denoting that he studies.

$\bar{A}$ be an event denoting that he does not study

B: Student passes an exam

$\bar{B}$: Student not passes an exam.

Scanned with CamScanner

$P[B] = 0.9$

$P[B/A] = 0.9$ $\quad P[A] = 0.75$

$P[B/\bar{A}] = 0.2$

$P[\bar{A}/B] = ?$

$$P[A/B] = \frac{P(A)\ \cancel{P(A/B)}\ P(B/A)}{P(A)\ P(B/A) + P(\bar{A}) \cdot P(B/\bar{A})}$$

$$\begin{array}{r} 75 \\ 9 \\ \hline 675 \end{array}$$

$$= \frac{0.75\,(0.9)}{0.75\,(0.9) + 0.25\,(0.2)}$$

$$= \frac{0.675}{0.675 + 0.05} = \frac{0.675}{0.725}$$

$$= 0.9310$$

For a certain binary commn. channel the Prob that transmitted 0, is received as 0 is 0.95 (The Prob that transmitted 1, is received as 1 is 0.9) If the Prob that a zero is transmitted is 0.4. Find the Prob that (i) 1 is received (ii) 1 was transmitted given that 1 was received.

$P(B/A)$

$A/B$

$\bar{A}/\bar{B}$ $\quad \bar{A}/\bar{B}$

Let $A$ be 1 transmission

$\bar{A}$ be 0 "

$B$ be 1 received

$\bar{B}$ be 0 received.

Given : $\cancel{P(B/A)}$ $P(A/B) = 0.95$ $\quad P(A) = 0.4$

$P(\bar{A}/\bar{B}) = 0.9$ $\quad P(\bar{B}) : ?$

$P(\bar{A}/\bar{B}) = ?$

$P(\bar{B}/\bar{A}) = 0.95$ $\quad P(\bar{A}) = 0.4$

$P(B/A) = 0.9$ $\quad A(A) = 0.6$

$$P(B) = ? \quad \frac{P(A)\ P(\bar{A}/B)}{P(\bar{A})\ P(\bar{A}/B) +}$$

$$P(B) = \frac{0.6(0.9)}{(0.4)(0.01)} = 0.56$$

$$P(A/B) = ? \quad \frac{P(A)\ (P(B))}{P} = \frac{0.6(0.9)}{0.56}$$

$$= \frac{27}{28}$$

# Conjugate gradient algorithm

Step 1: Start with arbitrary pt $X$

$$\nabla f(x_1) = \left[\frac{\partial F}{\partial x_1}, \frac{\partial F}{\partial x_2}\right]$$

Step 2: Set the first search direction $S_i = -\nabla f(X_i)$
$= -\nabla f_i$

Step 3: find the point $X_2$ according to relation

$$X_2 = X_i + \lambda_i S_1$$

Step 4: Set $i=2$ & go to next step

Step 5: Find $\nabla f_i = \nabla f(X_i)$ & set

$$S_i = -\nabla f_i + \frac{|\nabla f_i|^2}{|\nabla f_{i-1}|^2} S_{i-1}$$

Step 6: find optimum step length $\lambda_i^*$ in direction of $S_i$ & find new point,

$$X_{i+1} = X_i + \lambda_i^* S_i$$

Step 7: Test for optimality of $X_{i+1}$. If $X_{i+1}$ is Optimum, stop

Step 8: Otherwise set $i = i+1$ & go to step 4

# ML Algorithms

- Algorithms
- Learning Hidden Patterns
- Predict O/P
- Performance & Experience from past experience

**Supervised Learning** | **Unsuper** | **Reinforcement**

→ Reg
  * Linear
  * Polynomial

→ Decision Tree

→ Ran. Forest

→ Clustering
  * SVD
  * PCA
  * K-means

→ Continuous

→ Classification
  * KNN
  * Trees
  * Logistic Regression
  * Naive Bayes
  * SVM

→ Association Analysis
  * Aprion
  * FP Growth

→ Hidden Markov Model

Categorical

Scanned with CamScanner

## Capacity of the Model

* ability to fit variety of functions

* Model with low capacity struggles to fit training set

* Models with high capacity can overfit by memorising properties.

SVM : Inverse problem of curve fitting Algorithm.

- Optimising structural similarities
-

- appropriate capacity

models will perform well when capacity is appropriate.

models with Insufficient capacity struggles to fit

models with high capacity, can overfit

- Representational Cap :

- Effective capacity :

Terms : Total Loss = Bias + Variance (+ noise)

Signal : True underlying Pattern of Data

Noise Irrelevant data

Bias : prediction Error

Variance : Performs well with training dataset but not with test data set

Why variance?

training dataset may be strong, but so variance is less
test dataset may be minimum, so variance will be more.

## Overfitting

Fitting data too well

- features are noisy / uncorrelated to concept
- Modelling ~~data~~ process very sensitive
- Too much SEARCH

How to avoid Overfitting

* cross validation
* Train with more data.
* Remove Features
* Early stopping of training (when optimum is reached, stop)

## Underfitting:

* Learning too little of True Concepts
  → features do not capture concept
  → Too much bias in model
  → Too little search to fit model

How to avoid UF

Tuning time of training of model

Summary:

More Powerful Model is not always Perfect

Learn to Identify Overfitting & Underfitting

Tuning Parameters

## Principle of Least squares Approximate Solution for a linear system

**Pseudo Inverse:**

$AX = B$

$X = A^{-1}B$

True only if $A^{-1}$ exists.

In case of rectangular matrix, $A^{-1}$ does not exists, thus we go for Pseudo Inverse

let $A = \begin{bmatrix} 1 & 2 \\ -2 & 3 \\ 1 & 1 \end{bmatrix}_{3\times 2}$ $B = \begin{bmatrix} 5 \\ 4 \\ 4 \end{bmatrix}_{3\times 1}$ $X = \begin{bmatrix} x \\ y \end{bmatrix}_{2\times 1}$

The soln for ① is given by

$X = A^{+}B$

$A^{+} = (A^{T}A)^{-1}A^{T}$

$$A^TA = \begin{bmatrix} 1 & -2 & 1 \\ 2 & 3 & 1 \end{bmatrix}_{2\times3} \begin{bmatrix} 1 & 2 \\ -2 & 3 \\ 1 & 1 \end{bmatrix} = \begin{bmatrix} 6 & -3 \\ -3 & 14 \end{bmatrix}$$

$$(A^TA)^{-1} = \frac{1}{75}\begin{bmatrix} 14 & +3 \\ +3 & 6 \end{bmatrix}$$

$$(A^TA)^{-1} = \frac{1}{75}\begin{bmatrix} 14 & 3 \\ 3 & 6 \end{bmatrix}\begin{bmatrix} 1 & -2 & 1 \\ 2 & 3 & 1 \end{bmatrix}$$

$$= \frac{1}{75}\begin{bmatrix} 20 & -19 & 17 \\ 15 & +12 & 9 \end{bmatrix}_{2\times3}$$

$$X = \begin{bmatrix} x \\ y \end{bmatrix} = \frac{1}{75}\begin{bmatrix} 20 & -19 & 17 \\ 15 & +12 & 9 \end{bmatrix}_{2\times3}\begin{bmatrix} 5 \\ 4 \\ 4 \end{bmatrix}_{3\times1}$$

$$= \frac{1}{75}\begin{bmatrix} 92 \\ 149 \end{bmatrix}$$

$$x = \frac{92}{75} = 1.2266$$

$$y = \frac{149}{75} = 2.12$$

$$\begin{array}{r} 100 \\ -76 \\ 68 \\ \hline \end{array}$$

$$\begin{array}{r} 100 \\ -48 \\ \hline 75 \\ 48 \\ 36 \\ \hline 159 \end{array}$$

$x + 2y = 1.2266 + 2(2.12) = 6.466$

$-2x + 3y = 3.9068$

$x + y = 3.3466$

$Err_1 = 5 - 5.466 = -0.466$

$Err_2 = 4 - 3.9068 = 0.0932$

$Err_3 = 4 - 3.3466 = 0.6534$

Scanned with CamScanner

RMS > AM > GM > HM

Arithmetic mean, Geometric mean, Harmonic mean.

$|E_1| = 0.466 \quad |E_2| = 0.0932 \quad |E_3| = 0.6534$

Total Abs ERROR $= \dfrac{|E_1| + |E_2| + |E_3|}{\cancel{3}} = 1.2126.$

Mean Abs Error $= \dfrac{|E_1| + |E_2| + |E_3|}{3}$

$= \dfrac{0.466 + 0.0932 + 0.6534}{3}.$

$= \dfrac{1.2126}{3} = 0.4042$

$= 0.4042.$

Find the least square soln for the linear s/m

$x + 2y + 3z = 0$

$2x + 3y + 4z = 1$

$3x + 4y + 6z = 1$

$x - y + 2z = -1$

let A be $A = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 3 & 4 \\ 3 & 4 & 6 \\ 1 & -1 & 2 \end{bmatrix}_{4\times3}$

$X = \begin{bmatrix} x \\ y \\ z \end{bmatrix}_{3\times1} \qquad B = \begin{bmatrix} 0 \\ 1 \\ 1 \\ -1 \end{bmatrix}_{4\times1}$

$$AX = B$$

$$X = A^+ B$$

$$A^+ = (A^T A)^{-1} \cdot A^T$$

$$A^T \cdot A = \begin{bmatrix} 1 & 2 & 3 & 1 \\ 2 & 3 & 4 & -1 \\ 3 & 4 & 6 & 2 \end{bmatrix} \begin{bmatrix} 1 & 2 & 3 \\ 2 & 3 & 4 \\ 3 & 4 & 6 \\ 1 & -1 & 2 \end{bmatrix}$$

$$15 \quad 19 \quad 31$$