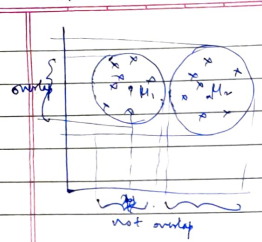
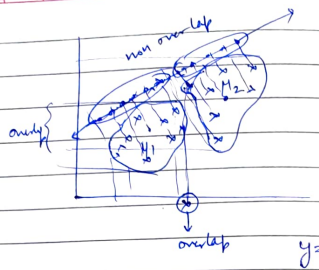
Linear discriminant Analysis





$y = w^T x$

Fischer Discriminant Ratio

$$J(w) = \frac{(\tilde{\mu}_1 - \tilde{\mu}_2)^2}{\tilde{s}_1^2 + \tilde{s}_2^2} \rightarrow \text{maximum} \rightarrow \text{minimum}$$

$S_W \rightarrow$ (within) inter class scatter matrix

$S_B \rightarrow$ (Between) class scatter matrix

$$S_W = \sum_{i=1}^{2} S_i \qquad J(w) = \frac{w^T S_B w}{w^T S_W w}$$

$= S_1 + S_2$

$$\frac{dJ(w)}{dw} = 0$$

find eigen value of this

$$\Rightarrow S_W^{-1} S_B w - \lambda w = 0$$

$$\begin{bmatrix} -6 & 3 \\ 4 & 5 \end{bmatrix}_{2\times 2} \begin{bmatrix} 1 \\ 4 \end{bmatrix}_{2\times 1} = \begin{bmatrix} -6+12 \\ 4+20 \end{bmatrix} = \begin{bmatrix} 6 \\ 24 \end{bmatrix} = 6 \begin{bmatrix} 1 \\ 4 \end{bmatrix}$$

eigen vector → eigen value

$$AV = \lambda V$$

x_1	x_2	class
4	1	1
9	10	2
2	4	1
2	3	1
6	8	2
3	6	1
9	5	2
8	7	2
10	8	2
4	4	1

(x_1, x_2)

C_1 = Class 1 = {(4,1), (2,4), (2,3), (3,6), (4,4)}

C_2 = Class 2 = {(9,10), (6,8), (9,5), (8,7), (10,8)}

LDA → Step 1 :- (mean of classes)

Compute the di-dimensional mean

$$\mu_1 = \left[\frac{4+2+2+3+4}{5}, \frac{1+4+3+6+4}{5} \right]$$

$$= [3, 3.6]$$

$$\mu_2 = \left[\frac{9+6+9+8+10}{5}, \frac{10+8+5+7+8}{5} \right]$$

$$= [8.4, 7.6]$$

Step 2 : (compute scatter matrix)

$$S_1 = \sum_{x \in w_i} (x - \mu_1)(x - \mu_1)^T$$

$$= \begin{bmatrix} 1 & -1 & -1 & 0 & 1 \\ -2.6 & 0.4 & -0.6 & 2.4 & 0.4 \end{bmatrix}_{2\times 5} \begin{bmatrix} 1 & -2.6 \\ -1 & 0.4 \\ -1 & -0.6 \\ 0 & 2.4 \\ 1 & 0.4 \end{bmatrix}_{5\times 2}$$

$$= \begin{bmatrix} 1+1+1+0+1 & 2.6 - 0.4 + 0.6 + 0 + 0.4 \\ -2.6 - 0.4 + 0.6 + 0 + 0.4 & 6.76 + 0.16 + 0.36 + 5.76 + 0.16 \end{bmatrix}$$

$$= \begin{bmatrix} 4 & -2 \\ -2 & 13.2 \end{bmatrix}$$

normalize $= \frac{S_1}{5} = \frac{1}{5} \begin{bmatrix} 4 & -2 \\ -2 & 13.2 \end{bmatrix} = \begin{bmatrix} 0.8 & -0.4 \\ -0.4 & 2.64 \end{bmatrix}$

(i) $\begin{bmatrix} 1 \\ -2.6 \end{bmatrix}\begin{bmatrix} 1 & -2.6 \end{bmatrix} = \begin{bmatrix} 1 & -2.6 \\ -2.6 & 6.76 \end{bmatrix}$

(ii) $\begin{bmatrix} -1 \\ 0.4 \end{bmatrix}\begin{bmatrix} -1 & 0.4 \end{bmatrix} = \begin{bmatrix} 1 & -0.4 \\ -0.4 & 0.16 \end{bmatrix}$

(iii) $\begin{bmatrix} -1 \\ -0.6 \end{bmatrix}\begin{bmatrix} -1 & -0.6 \end{bmatrix} = \begin{bmatrix} 1 & 0.6 \\ 0.6 & 0.36 \end{bmatrix}$

(iv) $\begin{bmatrix} 0 \\ 2.4 \end{bmatrix}\begin{bmatrix} 0 & 2.4 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & 5.76 \end{bmatrix}$

(v) $\begin{bmatrix} 1 \\ 0.4 \end{bmatrix}\begin{bmatrix} 1 & 0.4 \end{bmatrix} = \begin{bmatrix} 1 & 0.4 \\ 0.4 & 0.16 \end{bmatrix}$

$\frac{S_1}{5} = \frac{1}{5}$ (i) + (ii) + (iii) + (iv) + (v)

$= \begin{bmatrix} 0.8 & -0.4 \\ -0.4 & 2.64 \end{bmatrix}$ 8.4, 7.6

$S_2 = \begin{bmatrix} 0.6 & -2.4 & 0.6 & -0.4 & 1.6 \\ 2.4 & 0.4 & -2.6 & -0.6 & 0.4 \end{bmatrix}_{2\times5} \begin{bmatrix} 0.6 & 2.4 \\ -2.4 & 0.4 \\ 0.6 & -2.6 \\ -0.4 & 0.6 \\ 1.6 & 0.4 \end{bmatrix}_{5\times2}$

$S_2 = \begin{bmatrix} 0.36+5.76+0.36 & 0.96 \\ 0.16+2.56 & 1.44+0.96+1.56-0.24+0.64 \\ 1.44-0.96+1.56-0.24+0.64 & 5.76+0.16+6.76+0.36+0.16 \end{bmatrix}$

$S_2 = \begin{bmatrix} 9.2 & 0.2 \\ -5.48 & 13.2 \end{bmatrix}$

normalize $\frac{S_2}{5}$ $\Rightarrow S_2 = \begin{bmatrix} 1.84 & -0.04 \\ -0.04 & 2.64 \end{bmatrix}$

$S_w = S_1 + S_2 = \begin{bmatrix} 2.64 & -0.44 \\ -0.44 & 5.28 \end{bmatrix} = S_w$

$S_w^{-1} = \begin{bmatrix} 0.3841 & 0.032 \\ 0.032 & 0.192 \end{bmatrix}$

Step 3:- Compute S_B

$S_B = (\mu_1 - \mu_2)(\mu_1 - \mu_2)^T$

$= \begin{bmatrix} -5.4 \\ -4 \end{bmatrix}\begin{bmatrix} -5.4 & -4 \end{bmatrix}$

$= \begin{bmatrix} 29.16 & 21.6 \\ 21.6 & 16 \end{bmatrix}_{2\times2}$

Step 4: Find the LDA projection vector

$S_w^{-1} S_B w = \lambda w$

$\left[S_w^{-1} S_B - \lambda I\right] = 0$

$S_w^{-1} \times S_B = \begin{bmatrix} 11.892 & 8.8092 \\ 5.0819 & 3.7644 \end{bmatrix}_{2\times2}$

$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ $A^{-1} = \frac{1}{ad-bc}\begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$

$(S_w^{-1} \times S_B) - \lambda\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = 0$

$\begin{bmatrix} 11.892 - \lambda & 8.8092 \\ 5.0819 & 3.7644 - \lambda \end{bmatrix} = 0$

$\lambda = 15.65$

$\lambda = \frac{|A|}{|\lambda_1| + |\lambda_2|}$

$44.767 + 15.654\lambda + \lambda^2 = 0$

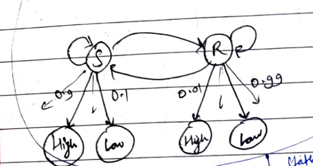
-44.767

$\lambda^2 + 15.65\lambda = 0$

$\lambda = |15.65| = 15.65$

Hidden Markov Model (Benchmark) → poor performance / this is not a good model / similarly NBC is Benchmark in classification

- Prior
- Transition ~~probability~~ probability table
- Observation Table

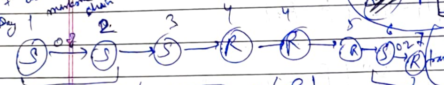


observations

States	H	L
S	0.9	0.1
R	0.01	0.99

observation probability table

1st order markov chain



transition probability table of HMM

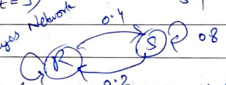
~~$P(S_{t+1}) = P(R|S_{t=S})$~~

$P(S_{t+1} = R / S_t = S)$

(t-1) \ (t)	R	S
R	0.6	0.4
S	0.2	0.8

$P(S_{t+1} = S / S_t = S)$

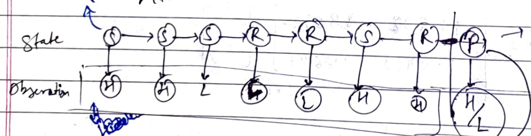
Bayes Network



markov assumption → $P(S_{t+1} | S_t, S_{t-1}, \dots S_1) = P(S_{t+1} | S_t)$

markov move

Hidden markov model



Problem 1 → Likelihood (given HMM $\lambda = (A, B)$ and obs seq O determine likelihood $P(O/\lambda)$)

$P(S | S_1 S_2 S_3 R_4 R_5 S_6 R_7 \; H_1 H_2 L_3 L_4 L_5 H_6 H_7)$

Problem 2 → Decoding (given obs seq O ~~predict~~ discover most probable state sequence (best hidden path))

Problem 3 - learning (Given obs seq O and set of states, learn the HMM parameters A and B)

---

pent Bank corpus

information extraction

POS tagging

pos → states

words → observations

word input

det — ii (adject.) nn (noun) vbd

The grand jury commented on a number of topics

determine — adj — noun — nns → noun plural

$O \to 3, 1, 3$

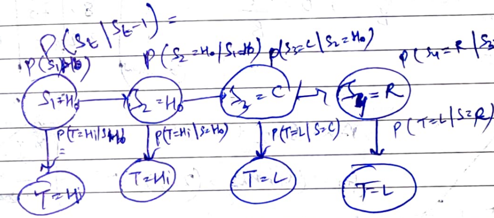
best $P(\text{states} | 3, 1, 3)$

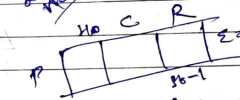
$P(\text{hot}, \text{hot}, \text{hot} | 3, 1, 3)$

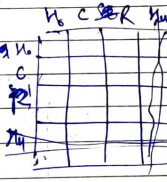
$P(\text{hot}, \text{cold}, \text{hot} | 3, 1, 3)$

$P(\text{cold}, \text{cold}, \text{cold} | 3, 1, 3)$

$P(S_t | S_{t-1}) =$

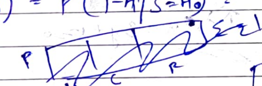


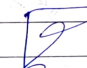




Joint probability distribution

$P(O|S) = P(T = Hi / S = Ho) =$





Likelihood of Ho Co Ho Co

$P(Ho) \cdot P(Co|Ho) \cdot P(Ho|Co) \cdot P(Co|Ho)$

Likelihood (Ho Hi Co Lo Ho Hi Co Hi)

$P(Hi|Ho) \cdot P(Ho) \cdot P(Lo|Co) \cdot P(Co|Ho) \cdot P(Hi|Ho) \cdot P(Ho|Co) \cdot P(Hi|Co) \cdot P(Co|Ho)$

or

$P(Ho) \cdot P(Co|Ho) \cdot P(Ho|Co) \cdot P(Co|Ho) \cdot P(Hi|Ho) \cdot P(Lo|Co) \cdot P(Hi|Ho) \cdot P(Hi|Co)$

Date
DELTA Pg No

Logistic Regression

$y = mx + b$

$Ax + By + C = 0$

$\frac{A}{B}x + \frac{B}{B}y + \frac{C}{B} = 0$

$y = -\frac{A}{B}x - \frac{C}{B}$

$m = -\frac{A}{B}$, $b = \frac{-C}{B}$

$Ax_1 + Bx_2 + C = 0$

$Ax_1 + Bx_2 + Cx_3 + D = 0$

for multiple inputs

x_0	x_1 (gpa)	x_2 (iq)	Level	
1	3.8	7.5	Yes	1
1	8.2	6.5	No	0
(assume)				

+ve region

-ve region

(2,5)

for $(x_1, x_2) = (2,5)$

$2A + 5B + C \geq 0$ then

$2x + 3y + 5 = 0$

$\Rightarrow 2x_2 + 3x_5 + 5 \geq 0$ (point is above the line)

○ (positive region)

× (negative region)

if $z = 0$ (point is on the line)

if $z < 0$ (point is below the line)

if A is change

if B is changed

if C is changed

$2x + 3y + 5 = 0$

$x + 2y + 4 = 0$

$\Rightarrow 7x + 5y + 6 = 0$

sudden change

so slow change

⇒ use gradient descent or gradient ascent

$w_1x_1 + w_2x_2 + w_3x_3 + w_4x_4 + w_5x_5 + w_0 = 0$

$y = w_0 + w_1x_1$

$y = w_0 + w_1x_1 + w_2x_2$

$x_0 = 1$

perceptron

x_1, x_2, x_3, x_4, x_n

w_0 bias 'b'

w_1, w_2, w_3, w_4, w_n

Σ Summation

y

desmos.com

$y = \left(\sum_{i=1}^{n} w_i x_i + b\right)$ or $\sum_{i=0}^{n} w_i x_i$ (for $b = w_0$)

$y = w_0 x_0 + w_1x_1 + w_2x_2 + \dots$

b

$w_0 x_0$

$w_0 \times 1$

$w_0 = b$

$2x^2 + 3x + 5 = 0$

$\eta = 0.01$

new value = old value + η × coordinates

$= (2, 3) + 0.01 \times (2, 5)$

$= 1.98, 2.95, 1.99$

for i in range (epoch):

{ randomly select points (x_1, x_2)

$x_i \in N$ and $\sum w_i x_i > 0$

{ $w_{new} = w_{old} + \eta(x_i)$

if $x_i \in P$ and $\sum w_i x_i \leq 0$

$w_{new} = w_{old} - \eta(x_i)$ }

x_0 (assum)	x_1	x_2	y	ŷ
1			0	
1			1	
1			0	
1			1	

for x_1, x_2 if $y = 0$ then i.e. $\sum w_i x_i < 0$

$w_{new} = w_{old} - \eta(y - \hat{y}) x_i$

y	ŷ	$y = \sum w_i x_i$	
1	1		$w_{new} = w_{old}$ no change
0	0		$w_{new} = w_{old}$ no change
1	0		$w_{new} = w_{old} - \eta x_i$
0	1		$w_{new} = w_{old} + \eta x_i$

public key → encryption
private key → decryption

digital sign → public sign

$37 \times 7 / 900$

Public key cryptography

RSA

Knapsack cryptography :–

we have K types $\Rightarrow a = [a_1, a_2 \ldots a_k]$ (publicly announced)

and $x = [x_1, x_2 \ldots x_k]$ ← plaintext

ciphered text ← $s = \text{KnapsackSum}(a, x) = x_1 a_1 + x_2 a_2 + \ldots x_k a_k$

given a and x we can easily calculate s

Given s and a_i it is difficult to find x

$a_i \geq a_1 + a_2 \ldots a_{i-1}$ (superincreasing)

KnapsackSum (x[1...k], a[1...k])

$s \leftarrow 0$

for i ← 1 to k

$s \leftarrow s + a_i x_i$

return s

Inv-knapsack (s, a[1...k])

for (i = k to 1)

{ if $s \geq a_i$

$x_i \leftarrow 1$

$s \leftarrow s - a_i$

else $x_i \leftarrow 0$ }

return x[1...k]

i	a_i	s	$s \geq a_i$	x_i	$s \leftarrow s - a_i x_i$
6	400	272	F	$x_6 = 0$	272
5	201	272	T	$x_5 = 1$	71
4	94	71	F	$x_4 = 0$	71
3	46	71	T	$x_3 = 1$	25
2	25	25	T	$x_2 = 1$	0
1	17	0	F	$x_1 = 0$	0

$\Rightarrow x_i = [0, 1, 1, 0, 1, 0]$ ⇒ knapsack has 25, 46, 201

$25 + 46 + 201 = 71 + 201 = 272$

permute (4) → 1 2 3 4 ⇒ 2 3 1 4

1 2 3 4

$$2165 \times (37^{-1} \bmod 900)$$

$$(a \times b) \bmod n = (a \bmod n \times b \bmod n) \bmod n$$

58-51 = 9.86

$$(10 \times 9) \bmod 7 = 10 \bmod 2 \times 9 \bmod 2$$

$= 0 \times 1$

$= 0$

9 mod 2 = 0

127 mod 7 = 5 = (6 × 2) mod 7 = 12 mod 7

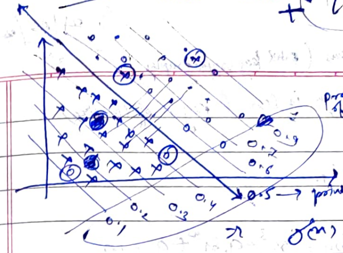
= 5

(x × 37) mod 900 = 1

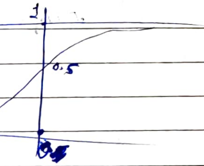
ML

$$w_{new} = w_{old} + \eta(y_i - \hat{y}_i)x_i$$

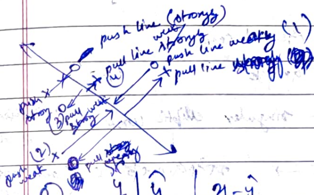
we will use sigmoid function so that when $y_i = \hat{y}_i$ it does not give zero



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



$0 \le x \le 1$



$$z = \sum w_i x_i$$

		y	$\hat{y}$	$y-\hat{y}$
1 1	1)	1	0.8	0.2
0 1	4)	0	0.65	-0.65
1 0	3)	1	0.3	0.7
0 0	2)	0	0.15	-0.15

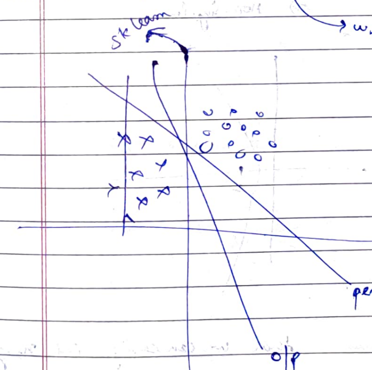
$$\sigma(z) = \frac{1}{1+e^{-z}}$$

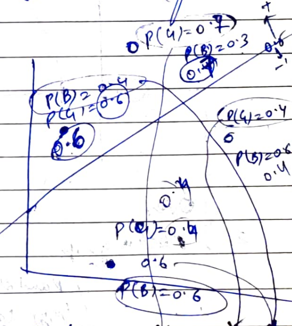
$w_{new} = w_{old} + 0.2\eta x_i$ (push) weak

$w_{new} = w_{old} - 0.65\eta x_i$ (pull) strong

$w_{new} = w_{old} + \eta \times 0.7 x_i$ (push) strong

$w_{new} = w_{old} - 0.15\eta x_i$ (pull) weak





Max likelihood = 0.7 × 0.4 × 0.4 × 0.6

= 0.28 × 0.24

= 0.0672

log (max likelihood) = $\log_{10}(a \times b) = \log a + \log b$

= log (0.0672) = -1.17263

240
240
240
240
-48
672

NoSecurity

Date 21/03/23 DELTA Pg No

Ecliptic curve cryptosystems (public key cryptography)

used when memory is less (key is of less size)

Elliptic Curve over Reals

$$y^2 + b_1xy + b_2y = x^3 + a_1x^2 + a_2x + a_3$$

↓ special case

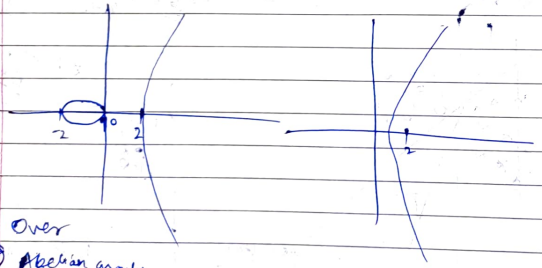
$$y^2 = x^3 + ax + b$$

if $4a^3 + 27b^2 \neq 0$ → no singular elliptic curve (3 distinct roots of $x^3+ax+b=0$)

singular → $x^3+ax+b=0$ (does not have 3 distinct roots)

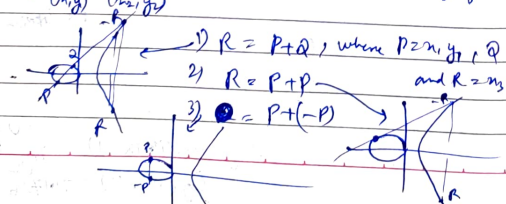
$y^2 = x^3 - 4x$ → $a=-4, b=0$ → $4x(-4)^3 + 27x0^2 \neq 0$ ⇒ non-singular

$y^2 = x^3 - 1$ → $a=0, b=-1$ → $4x0^3 + 27x(-1)^2 \neq 0$ ⇒ non-singular



Over Abelian groups

if P and Q are known then we can easily find x_3, y_3 (x_1, y_1) (x_2, y_2)



1) $R = P+Q$, where $P = x_1, y_1$, $Q = x_2, y_2$

2) $R = P+P$

3) $\infty = P + (-P)$ and $R = x_3, y_3$

1) $R = P+Q$

P and Q are distinct $x_1 \neq x_2$, $y_1 \neq y_2$

$$\lambda = (y_2 - y_1)/(x_2 - x_1)$$

$$x_3 = \lambda^2 - x_1 - x_2 \qquad y_3 = \lambda(x_1 - x_3) - y_1$$

2) $R = P+P$ (P and Q are same)

$$\lambda = (3x_1^2 + a)/2y_1$$

$$x_3 = \lambda^2 - x_2 - x_2 \qquad y_3 = \lambda(x_1 - x_3) - y_1$$

3) P and P are additive inverse

9 mod 13 = 4

9, 4 are additive inverses

(9+4) mod 13 = 0

13 mod 13 = 0

Operations → Closure

Associativity $(P+Q)+R = P+(Q+R)$

Commutativity $P+Q = Q+P$

Existence of identity $P = P+\infty = \infty+P$ (additive identity)

Existence of inverse $P+Q = \infty$

⇒ Group and Field

Elliptic Curves over GF(p)

inverse of a pt (x, y) is $(x, -y)$

if p = 13 i.e. $(4, 2) \rightarrow (4, 11)$ is the additive inverse (2+11) mod 13 = 0

elliptic curve points (p, a, b)

```
{ x ← 0
  while (x < p)
  { w ← (x^3 + ax + b) mod p
    if (w is a perfect sq in Zp) output (x, √w), (x, -√w)
    x ← x+1
  }
```

write y^2

find points and plot graph (define graph)

10.14

$y^2 = x^3 + x + 1$ calculate over modulo 13

a=1, b=1

p = 13

$w = (x^3 + x + 1) \mod 13$

x	$w = (x^3+x+1) \mod 13$	w is perfect sq in positive Integer	y
0	w = 1 mod 13 = 1	Yes	1, 12
1	w = (1+1+1) mod 13 = 3	No	3, 16
2	w = (8+2+1) mod 13 = 11	No	–
3	w = 27+3+1 = 31 mod 13 = 5	No	–
4	w = (64+4+1) mod 13 = 69 mod 13 = 4	yes	2, 11
5	w = (125+6) mod 13 = 131 mod 13 = 1	Yes	1, 12
6	2	No	–
7	343+8 = 0	Yes	0, 0
8	1	Yes	1, 12
9	11	No	–
10	10	No	10+26 = 36
11	4	Yes	2, 11
12	12	No	12+13

p−1 = 13−1 = 12

points	additive inverse of points
(0,1)	(0,12)
(4,2)	(4,11)
(5,1)	(5,12)
(7,0)	(7,0)
(8,1)	(8,12)
(10,6)	(10,7)
(11,2)	(11,11)
(12,5)	(12,8)

(1+12) = 13

2+11

points	reflection
(0,1)	(0,12)
(1,4)	(1,9)
(4,2)	(4,11)
(5,1)	(5,12)
(7,0)	(7,0)
(8,1)	(8,12)
(10,6)	(10,7)
(11,2)	(11,11)
(12,5)	(12,8)

farrozan

(y−y) mod 13

(x,y) (x,−y)

16 mod 13 = 3 → √16 = 4

36 mod 13 = 10 ⇒ √36 = 6, y^2 = 10 mod 13

25 mod 13 = 12 → √25 = 5

10.15

R = P+Q P = (4,2), Q = (10,6)

a) $\lambda = (6-2) \times (10-4)^{-1} \mod 13 = 4 \times 6^{-1} \mod 13$

$6^{-1} \mod 13 = 11$ ← 11 × 6 mod 13 = 66 mod 13 = 1

⇒ (11×4) mod 13 = 44 mod 13 = 5

b) $x_3 = (5^2 - 4 - 10) \mod 13 = 11 \mod 13 = 11 = x_3$

c) $y_3 = (\lambda((4-11)) - 2) \mod 13$

= 5×(−7) − 2 = −35 − 2 = −37 mod 13 ⇒

2+37 mod 13 = 39 mod 13 = 0

⇒ = 2 mod 13 = 2 = y_3

R = (11,2)

$(11 + 13i) \mod 13 = 11$

$\sqrt{(11+13i)}$

P = (4,2) and Q = (4,2) = P R = P+P

$y^2 = x^3 + x + 1$: a=1, b=1

a) $\lambda = ((3x_1^2 + a)/2y_1) \mod 13 = (3 \times 4^2 + 1)(2 \times 2)^{-1} \mod 13$

$= (49 \times 4^{-1} \mod 13)$

b) $x_3 = (\lambda^2 - x_1 - x_2) \mod 13$

$= 9^2 - 4 - 4 = 81 - 8$

= (73) mod 13

= 7

x × 4 mod 13 = 1

10×4 mod 13 = 40 mod 13 = 1

⇒ $4^{-1} \mod 13 = 10$

→ 49×10 mod 13 = 9

c) $y_3 = (9(4-7) - 2) \mod 13$

= −29 mod 13

= (x+29) mod 13 = 0

→ x = 10 → 10 mod 13 = 10

⇒ x_3, y_3 = (7,10)

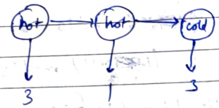
NLP

Transition probability matrix, observation / emission probability matrix

Date 21/02/23

Likelihood computation :- given an HMM $\lambda = (A, B)$ and an observation sequence O, determine the likelihood of $P(O|\lambda)$ =

eg:-



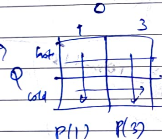
$P(hot, 3, hot, 1, cold, 3) = P(hot) \cdot P(3|hot) \cdot P(hot|hot) \cdot P(1|hot) \cdot P(cold|hot) \cdot P(3|cold)$

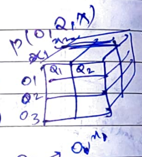
H	C	P
0.4	0.5	0.1

$= P_1 \prod_{i=2}^{T} P(S_t | S_{t-1}) \cdot \prod_{i=1}^{T} (O_t | S_t)$

$P(3,1,3) \rightarrow$ given

$P(O) = \sum_Q P(O, Q) \rightarrow$





	H	C
1		
2	0	0
3		

$Q \rightarrow O_0$

$P(O_1, Q_2) = P(O_1) \cdot P(O_2)$ independent

$= \sum_Q P(O|Q) \cdot P(Q)$

$P(1) = P(1, cold) + P(1, hot)$

likelihood

$P(3,1,3) = P(3,1,3, cold, cold, cold) + P(3,1,3\ cold\ cold\ hot)$ ——

$P(3) \cdot P(3|3) \cdot P(3|1) \cdot P(3|cold) \cdot P(1|cold) \cdot P(3, cold)$ + —— $P(3,1,3,$ hot hot hot$)$

ML

log(0.4)

Date

max likelihood = $-\log(0.7) + \log(0.6)$

(ml)

$ml_i = -y_i \log(\hat{y}_i) - (1-y_i)\log(1-\hat{y}_i)$

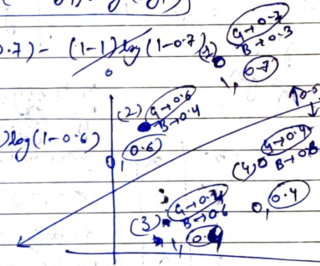
$ml = \prod ml_i$

right classified — point 1 green :- $ml_1 = -1\log(0.7) - (1-1)\log(1-0.7)$

$= -\log 0.7$

wrong classified — point 2 Black :- $ml_2 = -0\log(0.6) - (1-0)\log(1-0.6)$

$= -\log 0.4$



point 3 :- Black — right classified — $ml_3 = -1\log(0.4) - (1-1)\log(1-0.4)$

$= -\log 0.4$

wrong — point 4 :- $ml_4 = -0\log(0.4) - (1-0)\log(1-0.4)$

$= -\log 0.6$

no log: $ml = \prod m_i$ — log $\Rightarrow ml = \sum ml_i = -\log 0.7 + \log 0.4 + \log 0.4 + \log 0.6$ $= -1.17$

$ml = 0.7 \times 0.4 \times 0.4 \times 0.6 = 0.0672$

log loss function $= L = \frac{1}{n}\sum_{i=1}^{n}\left[-y_i\log(\hat{y}_i) - (1-y_i)\log(1-\hat{y}_i)\right]$

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

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

$\left(\frac{1}{1+e^{-z}}\right)\left(\frac{e^{-z}}{1+e^{-z}}\right) \Leftarrow \frac{1}{(1+e^{-x})^2} \Leftarrow \frac{+e^{-x}}{(1+e^{-x})^2}$

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

Over $GF(2^m)$ ⟶ Inverses → P = (x, y)

10-16

$GF(2^3)$

0	000	$g^3 = g+1$
1	001	$g^4 = g^2+g$
g	010	$g^5 = g^3+g^2 = g^2+g+1$
g^2	100	$g^6 = g^3 \cdot g^3 = g^2+1$

011
110
111
101

$IP = f(m) = x^3+x+1$

$g^3+g+1=0$

$\Rightarrow g^3 = g+1$

$y^2+xy = x^3+g^3x^2+1$ ⟶ $a = g^3,\ b = 1$

$y^2+xy = x^3+ax^2+b$

x	x	f(x) mod IP
0	0	$1 \bmod (g^3+g+1) = 1$
1	1	$g \bmod (g^3+g+1) = g$
2	g	$(g^2+1) \bmod (g^3+g+1) = g^2+1$
3	g^2	$(g^4+g^2+1) \bmod (g^3+g+1) = g^2+g+g^2+1 = g$
4	g^3	$g^3+g^3 \bmod (g^3+g+1) = g^3+g^3+g+1 = g+1$
5	g^4	$g^{12}+g^4 \bmod (g^3+g+1) = g^2+g$
6	g^5	$g^{15}+g^5 \bmod (g^3+g+1) = g^2+g+1$
7	g^6	$g^{18}+g^6 \bmod (g^3+g+1) = g^2+1$
8		
9		
10		
11		
12		

0, 1	0, 1	⇒ (0, 1)
g^2, 1	$g^2, g^2+1 = g^6$	⇒ (g^2, g^6)
g^3, g^2	$g^3, g^3+g^2 = g^2+g+1 = g^5$	⇒ (g^3, g^5)
g^5, g^6	$g^5, g^2+g = g^4$	⇒ (g^5, g^4)
g^6, g	$g^6, g^2+g+1 = g^5$	⇒ g^5

g^2, 1	g^2, g^6
g^3, g^2	g^3, g^5
$g^5, 1$	g^5, g^4
g^6, g	g^6, g^5

$y^2+xy = x^3+g^3x^2+1$

a b

$IP = f(m) = x^3+x+1 = 0 \Rightarrow g^3+g+1 = 0$

$g^3 = -g-1$

$g^3 = g+1$

$F_3 = F_{p^m}$ $GF(2^3)$

$x, y \in \{0, 1, g, g^2, g^3, g^4, g^5, g^6\}$

$(x,y) = \{(0,0)\ (0,1)\ (0,g)\ (0,g^2)\ (0,g^3)\ (0,g^4)\ (0,g^5)\ (0,g^6)$

0	(1,0)	(1,1)	(1,g)	(1,g²)	(1,g³)	(1,g⁴)	(1,g⁵)	(1,g⁶)
1	(g,0)	(g,1)	(g,g)	(g,g²)	(g,g³)	(g,g⁴)	(g,g⁵)	(g,g⁶)
g	(g²,0)	(g²,1)	(g²,g)	(g²,g²)	(g²,g³)	(g²,g⁴)	(g²,g⁵)	(g²,g⁶)
g³	(g³,0)	(g³,1)	(g³,g)	(g³,g²)	(g³,g³)	(g³,g⁴)	(g³,g⁵)	(g³,g⁶)
g⁴	(g⁴,0)	(g⁴,1)	(g⁴,g)	(g⁴,g²)	(g⁴,g³)	(g⁴,g⁴)	(g⁴,g⁵)	(g⁴,g⁶)
g⁵	(g⁵,0)	(g⁵,1)	(g⁵,g)	(g⁵,g²)	(g⁵,g³)	(g⁵,g⁴)	(g⁵,g⁵)	(g⁵,g⁶)
g⁶	(g⁶,0)	(g⁶,1)	(g⁶,g)	(g⁶,g²)	(g⁶,g³)	(g⁶,g⁴)	(g⁶,g⁵)	(g⁶,g⁶)

(x, y)	LHS = $(y^2+xy) \bmod IP$	RHS = $(x^3+g^3x^2+1) \bmod IP$	LHS = RHS
(0, 1)	$(1^2+0) \bmod (g^3+g+1) = 1$	$(0+0g^3+0) \bmod g^3+g+1 = 1$	Yes
$(g^2, 1)$	$(1^2+g^2) \bmod (g^3+g+1) = 1+g^2$	$g^6+g^3+1 = (g^3)^2(1+g)+1$ $= (g+1)^2(1+g)+1$ $= (g^2+1)(1+g)+1$ $= g^3+g^2+g^2+g+1$ $= g^3+g+1+g$ $= g^2+1$	Yes
g^3, g^2	$g^4+g^5 = g(g+1)[1+g] = g(g^2+1)$ $= g^3+g$ $= g+1+g = 1$	$g^9+g^9+1 = 1$	Yes
$g^5, 1$	$1+g^5 = 1+g^2(g+1) = g^3+g^2$ $= g+1+g^2$ $= g^2+g$	$g^{15}+g^{13}+1 = g^{13}(1+g^2)+1$ $g(g^3)^4(1+g^2)+1 = g(g^4)^3(1+g^2)+1$ $= g(g^4+g+1)(1+g^2)+1$ $= (g^4+g+1)(1+g^2)+1$ $= g^4+g^3+g^2+g^2+g+1$ $= g^2+g+1+1$ $= g^2+g$	Yes
g^6, g	$g^2+g^7 = g^2+g(g^3)^2 = g^2+g(g+1)^2$ $= g^2+g(g^2+1)$ $= g^2+g^3+g = g^2+g+1+g$ $= g^2+1$	$g^{18}+g^{15}+1 = g^{15}(g^3+1)+1$ $= (g+1)^5(g+1)+1 = (g+1)(g^2+1)g+1$ $= (g^5+g+g^2+1)g+1 = (g^2+g^3+g^2+g)g+1$ $= g^2+1$	Yes

order of E $y^2 + xy = x^3 + g^3x^2 + 1$.

$\#E = 2Tr(g^3) \bmod 4 = (2 \times 7) \bmod 4 = 14 \bmod 4 = 2$

$Tr(g^3) = \sum_{0}^{2} (g^3)^{2^i} = g^3 + (g^3)^2 + (g^3)^4$

$= g^3 + g^6 + g^{12 \bmod 7} = g^3 + g^6 + g^5$

011, 101, 111 → 111 $= g^5$

LHS $= y^2 + y$

for $y = g^0, g^1, g^2, g^3, g^4, g^5, g^6$

and $x = 0, g^0, g^1, g^2, g^3, g^4, g^5, g^6$

	$y^2 + xy = x^3 + g^3x^2 + 1$	Yes/no	inverse
$x=0, y=0$			
0, 0	$0 + 0 \cdot 0 = 0 + 0 + 1$ ✗	NO	
(0, 1) ①	$1 + 0 = 0 + 0 + 1$ ⇒ ✓	Yes →	$(0, 1+0) = (0,1)$ ✓ ①
1, 0	$0 + 0 = 1 + g^3 + 1 = g^3$ ✗	No	
1, 1	$1 + 1 \neq 1 + g^3 + 1$ ✗	No	
1, g	$g^2 + g \neq 1 + g^3 + 1 = g^2 + g = g+1$ ✗	No	
1, g^2	$g^4 + g^2 \neq g^3$ ✗		
1, g^3	$g^6 + g^3 \neq g^3$ ✗		
1, g^4	$g^8 + g^4 \neq g^3$ ✗		
1, g^5	$g^{10} + g^5 = g^5(g^5+1) = (g^2+g)(g^2+1) + g) \neq g^3$ ✗		
1, g^6	$g^{12} + g^6 = g^6(g^6+1) = (g^2+1)(g^2+g+1) \neq g^3$ ✗		
g, 0 ✗	$0 = g^3 + g^5 + 1$ ✗		
g^2, 0	$0 = g^6 + g^7 + 1 = (g^2+1)(g+1) = g^3 + g^2 + g + 1 + 1$ ✗		
(g^2, 1) ②	$1 + g^2 = g^6 + g^7 + 1 = g^2 + 1$ ✓	Yes	$(g^2, 1+g^2) = (g^2, g^6)$ ③
g^3, 0	$0 = g^9 + g^9 + 1 = 1$ ✗		
g^3, 1	RHS $= 1$, LHS $= y^2 + xy = 1 + g^3 = g^4 + g$		
g^3, g	LHS $= g^2 + g^4 = g^2 + g^2 + g = g$		
(g^3, g^2) ④	LHS $= g^4 + g^5 = g^2 + g + g^3 + g^2 = g + g + 1 = 1$ ✓		$(g^3, g^2 + g^3) = (g^3, g^5)$ ⑤

RHS $= x^3 + g^3x^2 + 1$, LHS $= y^2 + xy$ — inverse

$g^4, 0$: for $x = g^4$ ⇒ RHS $= g^{12} + g^{11} + 1 = (g^2+1)^2 + (g^2+1)(g^2+g+1)$

$= g^4 + 1 + g^4 + g^3 + g^2 + g^2 + g + 1$

$= g^3 + g = g + g + 1 = 1$

for $x = g^4, y = 0$: LHS $= 0 + 0 = 0$ ✗

$g^4, 1$: LHS $= 1 + g^4 = 1 + g^2 + g \neq 1$ ✗

g^4, g: LHS $= g^2 + g^5 = g^2 + g^2(g+1) = g^2 + g^3 + g^2 = g + 1 \neq 1$ ✗

g^4, g^2: LHS $= g^4 + g^6 = g^2 + g + (g+1)^2 = g^2 + g + g^2 + 1 = g + 1 \neq 1$ ✗

g^4, g^3: LHS $= g^6 + g^7 = (g+1)^2(g+1) = (g^2+1)(g+1) = g^3 + g^2 + g + 1 = g^2 \neq 1$ ✗

g^4, g^4 ✗

g^4, g^5 ✗

g^4, g^6 ✗

$g^5, 0$: RHS $= (g^5)^3 + g^3(g^5)^2 + 1$

$= (g^2+g+1)^3 + (g+1)(g^2+g+1)^2 + 1$

$= (g^2+g+1)^2 (g^2+g+1+g+1) + 1$

$= g^{12} + 1 = g$

$g^{15} + g^{13} + 1$

$= g^{13}(g^2+1) + 1$

$= g(g^{12})(g^2+1) + 1$

$= (g+1)^4(g^3+g) + 1$

$= (g^4+1)(g+g+1) + 1$

$= g^4 + 1 + 1 = g^2 + g$

LHS $= y^2 + xy = 0 + g^5 \times 0 = 0$

⑥ (g^5, 1): LHS $= 1 + g^5 = 1 + g^2(g+1) = 1 + g^3 + g^2 = g^2 + g = g^2 + g$ ✓

$(g^5, 1 + g^5) = (g^5, g^2 + g) = (g^5, g^4)$ ⑦

$g^6, 0$: RHS $= g^{18} + g^{15} + 1$

$= g^{15}(g^3+1) + 1$

$= (g+1)^5 (g+1+1) + 1$

$= (g+1)(g^4+1)g + 1$

$= (g^5 + g + g^4 + 1)g + 1$

$= (g^2 + g + 1 + g + g^2 + g + 1)g + 1 = g^2 + 1$

$g^6, 1$: LHS $= 1 + g^6 = 1 + g^2 + 1 = g^2$ ✗

⑧ (g^6, g): LHS $= g^2 + g^7 = g^2(1 + g^5) = g^2(g^2+g) = g^4 + g^3$

$= g^2 + g + g + 1$

$= g^2 + 1 = g^2 + 1$ ✓

$(g^6, g^6 + g) = (g^6, g^2 + 1 + g) = (g^6, g^5)$ ⑨

ML

Binary cross

log loss function $(L) = -\frac{1}{n}\sum_{i=1}^{n} y_i \log \hat{y}_i + (1-y_i)\log(1-\hat{y}_i)$ → cost fun

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

$\sigma'(z) = \sigma(z)[1-\sigma(z)]$

Date 27/03 DELTA Pg No.

loss func → one row

cost func → all rows (dataset point)

loss function = $y_i \log \hat{y}_i - (1-y_i)\log(1-\hat{y}_i)$

minimise cost function

$z = \sum_{i=1}^{n} w_i x_i$

$\hat{y} = \sigma(z)$

inputs →

$$y = \begin{bmatrix} x_{11} & x_{12} & x_{13} & \cdots & x_{1n} \\ x_{21} & x_{22} & \cdots & & x_{2n} \\ \vdots & & & & \\ x_{m1} & x_{m2} & \cdots & & x_{mn} \end{bmatrix}_{m \times n} \quad \sum_{i=1}^{n} w_i x_i \quad \hat{y} = \sigma(z)$$

$$\begin{bmatrix} w_1 & w_2 & \cdots & w_n \end{bmatrix}_{1 \times n} \times \begin{bmatrix} w_1 \\ w_2 \\ w_3 \\ \vdots \\ w_n \end{bmatrix}_{n \times 1}$$

$$y \cdot (xw)_{m \times 1} = (z)_{m \times 1}$$

$\hat{y}_1 = \sigma(x_{11}w_1 + x_{12}w_2 + \cdots + x_{1n}w_n)$

$\hat{y}_2 = \sigma(x_{21}w_1 + x_{22}w_2 + \cdots + x_{2n}w_n)$

$$\hat{y} = \begin{bmatrix} \hat{y}_1 \\ \hat{y}_2 \\ \hat{y}_3 \\ \vdots \\ \hat{y}_m \end{bmatrix} = \sigma \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1n} \\ x_{21} & x_{22} & \cdots & x_{2n} \\ \vdots & & & \\ x_{m1} & x_{m2} & \cdots & x_{mn} \end{bmatrix} \cdot \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_n \end{bmatrix}$$

$$\hat{y} = \sigma(xw) = \sigma(z)$$

$L = -\frac{1}{n}\sum_{i=1}^{n}\left(y_i \log \hat{y}_i + (1-y_i)\log(1-\hat{y}_i) \right)$

Taking $\sum_{i=1}^{n} y_i \log \hat{y}_i = y_1 \log \hat{y}_1 + y_2 \log \hat{y}_2 + \dots + y_n \log \hat{y}_n$

$$= \underset{1\times n}{\begin{bmatrix} y_1 & y_2 & y_3 & \cdots & y_n \end{bmatrix}} \log \underset{n\times 1}{\begin{bmatrix} \hat{y}_1 \\ \hat{y}_2 \\ \hat{y}_3 \\ \vdots \\ \hat{y}_n \end{bmatrix}}$$

$$= Y \log \hat{Y}$$

Taking $\sum_{i=1}^{n} (1-y_i)(1-\hat{y}_i) = (1-Y)\log(1-\hat{Y})$

$$= (1-Y)\log(1-\sigma(xw))$$

$$L = -\frac{1}{m}\left[Y \log \hat{Y} + (1-Y)\log(1-\sigma(xw)) \right]$$

$$\frac{\Delta L}{\Delta w} = \frac{\partial L}{\partial w} = \frac{\partial L}{\partial w_0} + \frac{\partial L}{\partial w_1} + \frac{\partial L}{\partial w_2} + \dots \frac{\partial L}{\partial w_n}$$

$$w_{new} = w_{old} + \eta \frac{\partial L}{\partial w_{old}}$$

x_1	x_2	$\longrightarrow$ $w_0x_0 + w_1x_1 + w_2x_2$
		1 2 3

3 values

$1-\sigma(z) = \sigma(1-z)$

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

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

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

$\Rightarrow (1 = e^{-1})$ X

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

$\sigma'(xw) = \frac{-1}{(1+e^{-xw})^2} \cdot -xe^{-xw}$

$= \frac{xe^{-xw}}{1+e^{-xw}} =$

$$\frac{\partial L}{\partial w} = \frac{\partial (Y \log \hat{Y})}{\partial w} + \frac{\partial (1-Y)\log(1-\sigma(xw))}{\partial w}$$

$$= Y \frac{\partial \log \hat{Y}}{\partial w} + (1-Y)\frac{\partial \log(1-\sigma(xw))}{\partial w}$$

$$= \frac{Y}{\hat{Y}} \frac{\partial \sigma(xw)}{\partial w} + \frac{(1-Y)}{(1-\hat{Y})} \frac{\partial (1-\sigma(xw))}{\partial w}$$

$$= \frac{Y}{\hat{Y}} \sigma(xw)[1-\sigma(xw)] - \frac{1-Y}{1-\hat{Y}} \frac{\partial \sigma(xw)}{\partial w}$$

$$\frac{\partial L}{\partial w} = Y\,x\hat{Y}(1-\hat{Y}) - \frac{1-Y}{1-\hat{Y}} x\hat{Y}(1-\hat{Y})$$

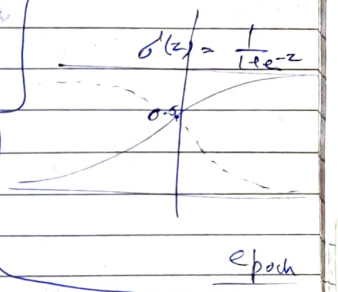
$$= xY(1-\hat{Y}) - x\hat{Y}(1-Y)$$

$$= xY - xY\hat{Y} - x\hat{Y} + xY\hat{Y}$$

$$\frac{\partial L}{\partial w} = -\frac{1}{m} x (Y - \hat{Y}) = -\frac{(Y-\hat{Y})x}{m}$$

gradient descent

$$w_{new} = w_{old} - \eta \frac{(Y-\hat{Y})x}{m}$$



$$L = -\frac{1}{m}\sum_{i=1}^{n}\left[y_i \log \hat{y}_i + (1-y_i)(1-\hat{y}_i) \right]$$

Taking $\sum_{i=1}^{n} y_i \log \hat{y}_i = y_1 \log \hat{y}_1 + y_2 \log \hat{y}_2 + \dots + y_n \log \hat{y}_n$

$$= \underset{1\times n}{\begin{bmatrix} y_1 & y_2 & y_3 & \cdots & y_n \end{bmatrix}} \log \underset{n\times 1}{\begin{bmatrix} \hat{y}_1 \\ \hat{y}_2 \\ \hat{y}_3 \\ \vdots \\ \hat{y}_n \end{bmatrix}}$$

$$= y \log \hat{y}$$

Taking $\sum_{i=1}^{n} (1-y_i)(1-\hat{y}_i) = (1-y)\log(1-\hat{y})$

$$= (1-y)\log(1-\sigma(xw))$$

$$L = -\frac{1}{m}\left[y \log \hat{y} + (1-y)\log(1-\sigma(xw)) \right]$$

$$\frac{\partial L}{\partial w} = \frac{\partial L}{\partial w_0} + \frac{\partial L}{\partial w_1} + \frac{\partial L}{\partial w_2} + \dots + \frac{\partial L}{\partial w_n}$$

$$w_{new} = w_{old} + \eta \frac{\partial L}{\partial w_{old}}$$

$x_1 \; x_2 \longrightarrow w_0 x_0 + w_1 x_1 + w_2 x_2$ (1, 2, 3)

3 values

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

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

$$1-\sigma'(z) = \sigma(1-z)$$

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

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

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

$\Rightarrow (1 = e^{-1})$ X

$$\frac{\partial L}{\partial w} = \frac{\partial (y \log \hat{y})}{\partial w} + \frac{\partial (1-y)\log(1-\sigma(xw))}{\partial w}$$

$$= y \frac{\partial \log \hat{y}}{\partial w} + (1-y)\frac{\partial \log(1-\sigma(xw))}{\partial w}$$

$$= \frac{y}{\hat{y}} \frac{\partial \sigma(xw)}{\partial w} + \frac{(1-y)}{(1-\hat{y})} \frac{\partial (1-\sigma(xw))}{\partial w}$$

$$= \frac{y}{\hat{y}} \sigma(xw)[1-\sigma(xw)] - \frac{1-y}{1-\hat{y}} \frac{\partial \sigma(xw)}{\partial w}$$

$$\frac{\partial L}{\partial w} = y x \hat{y}(1-\hat{y}) - \frac{1-y}{1-\hat{y}} x \hat{y}(1-\hat{y})$$

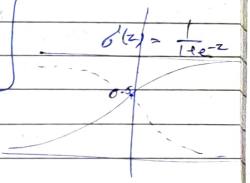
$$= xy(1-\hat{y}) - x\hat{y}(1-y)$$

$$= xy - xy\hat{y} - x\hat{y} + xy\hat{y}$$

$$\frac{\partial L}{\partial w} = -\frac{1}{m} x (y - \hat{y}) = -\frac{(y-\hat{y})x}{m}$$

gradient descent

$$w_{new} = w_{old} - \eta \frac{(y-\hat{y})x}{m}$$



NLP

Forward Algorithm for likelihood computation

we have 3 steps

Date 27/03/23 DELTA Pg No.

1) Initialization :-

$$\alpha_1(j) = a_{0j} \cdot b_j(O_1), \quad \text{for } j = 1 \text{ to } N$$

HMM $\lambda = (A, B)$

$O = \{O_1, O_2, O_3\}$

$P(O|\lambda)$

$P(O_1, O_2, O_3 | \lambda)$

$P(O_1, S_1, O_2, S_2, O_3, S_3 | \lambda)$

$P(S_1) \cdot P(O_1|S_1) \cdot P(S_2|S_1) \cdot P(O_2|S_2) \cdot P(S_3|S_2) \cdot P(O_3|S_3)$

2) Recursion :-

$$\alpha_t(j) = \sum_{i=1}^{N} \alpha_{t-1}(i) \cdot a_{ij} b_j(O_t) \quad \text{for } j = 1 \text{ to } N, \; t = 1 \text{ to } T$$

3) Termination :-

$$P(O|\lambda) = \alpha_T(q_F) = \sum_{i=1}^{N} \alpha_T(i) \cdot a_{iF}$$

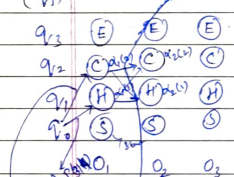
a_{ij} = transition probability from one state to another

$b_j(O_t)$ = observation probability of observation O_t given state j

$$\alpha_t(j) = P(O_1, O_2 \cdots O_{t-1}, O_t, q_t = j | \lambda)$$

$$\alpha_t(j) = \alpha_{t-1}(i) \cdot a_{ij} b_j(O_t)$$

(q_i)



$P(H|S) = 0.6$

$P(1|C) = 0.4$

Initialization

$\alpha_1(1) = a_{01} \; b_1(O_1) \quad P(O_1 = 3 \mid \text{state} = H)$

$= 0.8 \times 0.4 = 0.32$

$\alpha_1(2) = a_{02} \; b_2(O_1) = 0.2 \times 0.1 = 0.02$

$P(C|S) \cdot P(3|C)$

max

$\alpha_2(2) = \sum \alpha_{t-1}(i) \cdot a_i$

$\alpha_2(2) = \alpha_1(2) \cdot P(C|C) \cdot P(1|C)$

ML

Classification

Accuracy, Precision, Recall, F1 score, Macro F1 score, weighted F1 score

Date 29/3/23 DELTA Pg No.

y	$\hat{y}$	
0	0	✓
0	0	✓
1	0	✗
0	1	✗
1	1	✓
0	0	✓
1	1	✓
0	0	✓
1	0	✗
0	1	✗

$$\text{Accuracy} = \frac{T_P + T_N}{T_P + F_P + T_N + F_N}$$

$$= \frac{\text{No of correctly classified}}{\text{Total no of Prediction}}$$

$$= \frac{6}{10}$$

Confusion Matrix Predicted

Actual	0	1
0	T_N	F_P
1	F_N	T_P

$$\text{Precision} = \frac{T_P}{T_P + F_P} = \frac{\text{what proportion of Predicted Positive is True Positive}}{}$$

$$\text{Recall} = \frac{TP}{TP + FN} = \text{what proportion of Actual positive is True positive}$$

Spam \ Not spam	0	1
0	100	170
1	30	700

Model A

	0	1
0	100	190
1	10	700

Model B

Accuracy $= \frac{100 + 700}{100 + 170 + 30 + 700} = \frac{800}{1000} = 80\%$

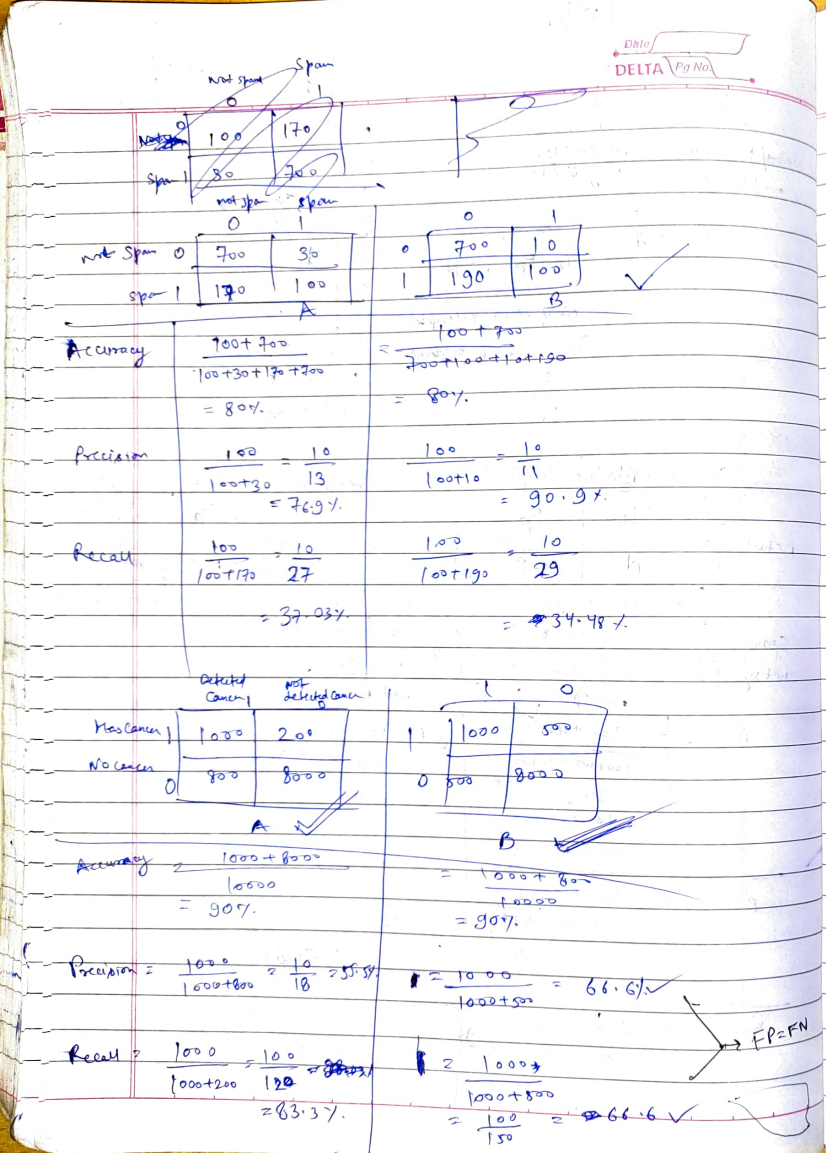
Model B: $= \frac{100 + 700}{100 + 190 + 10 + 700} = \frac{800}{1000} = 80\%$

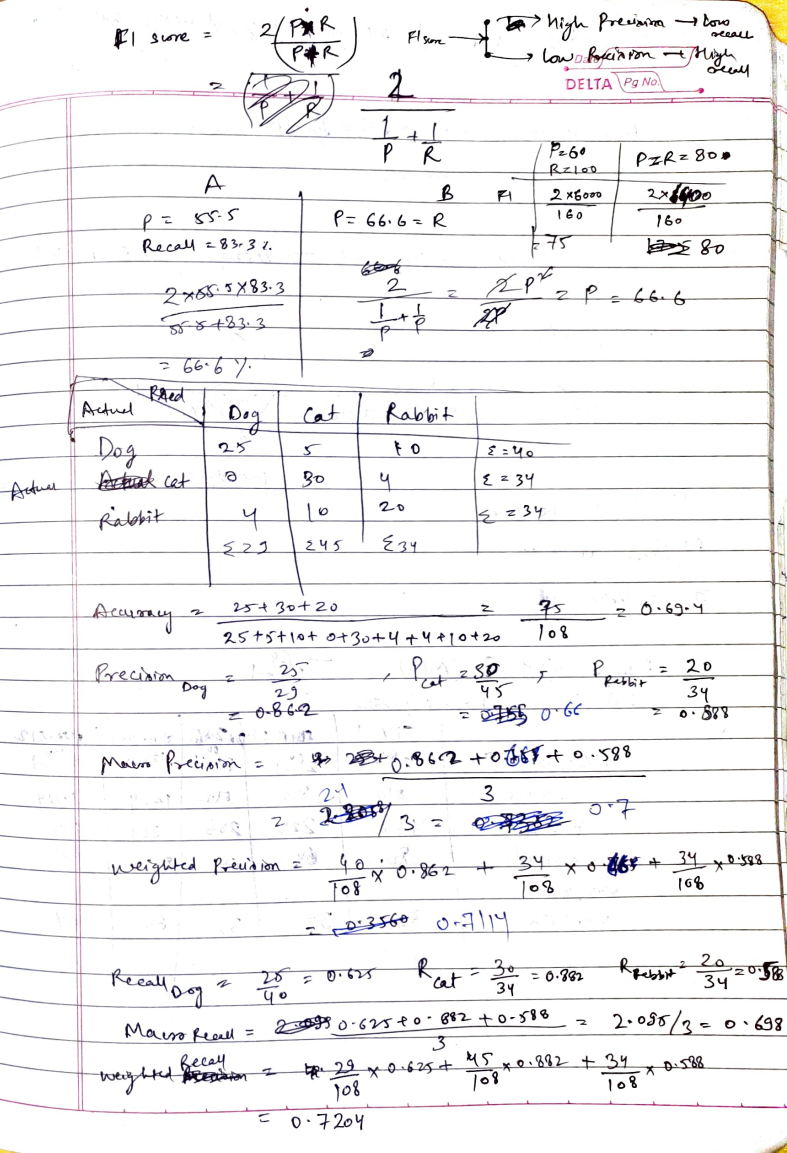
Precision $= \frac{700}{700 + 170} = \frac{700}{870} = 80.4\%$

Model B: $= \frac{700}{700 + 190} = \frac{700}{890} = 78.6\%$

Recall $= \frac{700}{700 + 30} = \frac{700}{730} = 95.9\%$

Model B: $= \frac{700}{700 + 10} = \frac{700}{710} = 98.6\%$


Not Spam
Spam
Not Spam
Spam
100
170
80
700
A
700
30
170
100
700
10
190
100
B
Accuracy
700+700 / 100+30+170+700
= 80%
100+700 / 700+100+10+190
= 80%
Precision
100 / 100+30 = 10/13
= 76.9%
100 / 100+10 = 10/11
= 90.9%
Recall
100 / 100+170 = 10/27
= 37.03%
100 / 100+190 = 10/29
= 34.48%
Actual Cancer
not Actual Cancer
Has Cancer
No Cancer
1000
200
800
8000
A
1000
500
500
8000
B
Accuracy = 1000+8000 / 10000
= 90%
90%
Precision = 1000/1000+800 = 10/18 = 55.5%
= 1000/1000+500 = 66.6%
Recall = 1000/1000+200 = 100/120 = 83.3%
= 1000/1000+800 = 100/150 = 66.6%
FP=FN


F1 score = 2(P*R / P+R)
= 2 / (1/P + 1/R)
F1 score
High Precision → low recall
Low Precision → High recall
A
P = 55.5
Recall = 83.3%
2×55.5×83.3 / 55.5+83.3
= 66.6%
B
P = 66.6 = R
2 / (1/P + 1/P) = 2P / 2 = P = 66.6
F1
2×6000 / 160
= 75
P=60
R=100
P*R = 800
2×6400 / 160
80
Actual
Predicted
Dog
Cat
Rabbit
25
5
10
Σ=40
Cat
8
30
4
Σ=34
Rabbit
4
10
20
Σ=34
Σ=29
Σ45
Σ34
Accuracy = 25+30+20 / 25+5+10+8+30+4+4+10+20 = 75/108 = 0.694
Precision Dog = 25/29 = 0.862
P Cat = 30/45 = 0.66
P Rabbit = 20/34 = 0.588
Macro Precision = 0.862+0.66+0.588 / 3
= 2.1 / 3 = 0.7
weighted Precision = 40/108 × 0.862 + 34/108 × 0.66 + 34/108 × 0.588
= 0.7114
Recall Dog = 25/40 = 0.625
R Cat = 30/34 = 0.882
R Rabbit = 20/34 = 0.588
Macro Recall = 0.625+0.882+0.588 / 3 = 2.095/3 = 0.698
weighted Recall = 29/108 × 0.625 + 45/108 × 0.882 + 34/108 × 0.588
= 0.7204

NS chapter-11 Message integrity and message authentication

Date 29/3/23

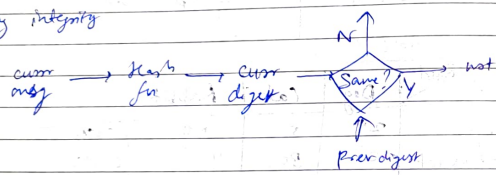
DELTA Pg No.

Message digest → Hash function (covers message)

maintains message integrity

Message (document) → Hash function → Message digest 5 (MD5)

checks integrity



1) Preimage Resistance

2) Second " "

3) Collision "

Message digest 5 (MD5) → $2^{64}-1$

SHA (secure Hash Algorithm

	SHA 1	SHA 224	SHA 256	SHA-384	SHA-512
Max message size	$2^{64}-1$	$2^{64}-1$	$2^{64}-1$	$2^{128}-1$	$2^{128}-1$
Block size	512	512	512	1024	1024
MD size	160	224	256	384	512
No of rounds	80	64	64	80	80
~~No of rounds~~					
Word size	32	32	32	64	64

Date

DELTA Pg No.

MD5 → step 1 — padding

input: msg of any length

output: 128 bit (32 hex)

no of rounds = 64

the aim of this step is to make the length of the original msg equal to a value which is 64 bit less than an exact multiple of 512

$(M + P + 64) = 0 \bmod 512$

M = 1000

1000 + P + 64 = 1536

P = 472

1 one 471 zeros

M = 800

$(800 + P + 64) = 0 \bmod 512$

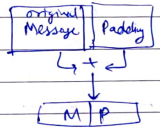
$1024 - (800 + 64)$

1024

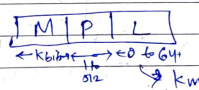
−864

160

⇒ P = 160



step 2: Append length



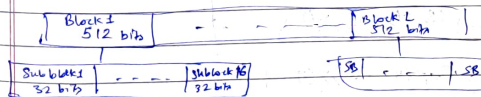
$L \times 512 = N \times 32$

32×16

$2^5 \times 2^4$

$= 2^9$

step 3: divide the input into 512 bit blocks.



step 4: initialising chain variables

A	01	23	45	67	32 bits
B	89	AB	CD	EF	
C	FE	DC	BA	98	
D	76	54	32	10	

MD5 buffer

Here

MD5 has 4 registers

$\Rightarrow 32 \times 5 = 160$

A 32 bits

B 32 bits

C 32 bits

D 32 bits

E 32 bits

1 block

Steps round 1:- 16 sub blocks | other constant

one Round → 16 iteration

a | b | c | d

A (32) B (32) C (32) D (32) → chaining (128 bits)

4 rounds and 16 iteration
⇒ 4 × 16 = 64 round

Processing

for i from 0 to 63 do

K[i] = floor (2^{32} × abs(sin(i+1)))

end for

32×4=128 bit ← K[0...3] := { 0xd76aa478 (32 bit), 0x..., 0x..., 0x... }

K[4...7] := { —, —, —, — }

K[60...63] := { —, —, —, 0xeb86d391 }

CV_q (128)

A B C D

F, T[1...16], X[i]

G, T[17...32], X[i]

H, T[33...48], X[i]

I, T[49...64], X[i]

(+) (+) (+) (+)

128

CV_{q+1}

addition (+) is mod 2^{32}

MD5 compression Function

A | B | C | D

+ is addition mod 2^{32}

X[k]

T[i]

CLS_s

A B C D

Rounds

1) F(B,C,D) = (B∧C) ∨ (B̄∧D)

2) G(B,C,D) = (B∧D) ∨ (C∧D̄)

3) H(B,C,D) = B⊕C⊕D

4) I(B,C,D) = C⊕(B∨D̄)

a ← b + ((a + g(b,c,d) + X[k] + T[i]) <<< s)

(CLS) circular left shift of 32 bit word by s bit

CV_0 = IV

CV_{q+1} = SUM_{32}(CV_q, RF_I(CV_q, RF_H(CV_q, RF_G(CV_q, RF_F)))))

Lecture 18 → 28/3/23
HMM extension
Vector semantics
distributional semantics

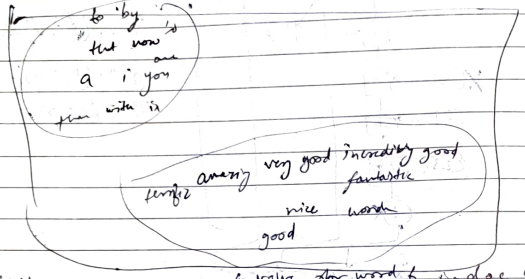
NLP

Date 29/3/23
DELTA Pg No.

each word = a vector

Not just 'word' or word 45

similar words are "nearby in space"



tf-idf → sparse vector (value for word & in doc d)
word2vec → dense vector
$w_{td} = tf_{t,d} \times idf_t$

glove
computer | computed | term-document matrix
they occur for a given corpus
Visualization of doc

word-word matrix ("term-context matrix")

context → neighbourhood to target

target	aardvark	computer	data	pinch	result	sugar
apricot	0	0	0	1	0	1
pineapple	0	0	0	1	0	1
digital	0	2	1	0	1	0
	0	1	6	0	4	0

sparse vector

feature ← x	y → class label
t_1, c	+ → presence in the cor
t_2, c	+
t_3, c	−
t_4, c	−

take a window of words for context
of left the target word t and a few
context window
target

The classifier is good

compute the prob

Compute similarity = t·c (dot product)

$sim(t,c) = t \cdot c$ to turn into a prob
we will use sigmoid func / logistic regression

$$P(+|t,c) = \frac{1}{1+e^{-sim(t,c)\theta_0 + \theta_1 x_1 + \theta_2 x_2 + \dots}}$$

maximize likelihood

$$\sum_{(t,c)\in +} \log P(+|t,c) + \sum_{(t,c)\in -} \log P(-|t,c)$$

maximize :- + label for the pairs from the positive samples
− label for the pairs samples from the negative sample

if vector is of d size

$-sim(t,c) = \theta_0 + \theta_1 x_1 + \dots + \theta_d x_d$

$t_0 + t_1 c_1 + \dots + t_d c_d$

train using stochastic gradient descent