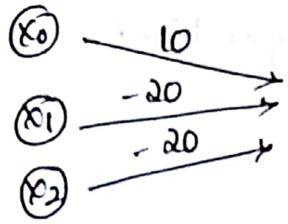
Ex 4:




$x_0=0, x_1=0, \hat{y} = \frac{1}{1+e^{-10}}$     10

$= 1$

$x_0=0, x_1=1, \hat{y} = \frac{1}{1+e^{10}} = 0$     $10-10 = -10$

$x_0=1, x_1=0, \hat{y} = \frac{1}{1+e^{10}} = 0$     $10-10 = -10$

$x_0=1, x_1=1, \hat{y} = \frac{1}{1+e^{30}} = 0$     $10-40 = -30$

NOR operation.

<u>ERROR/COST $F^N$</u>

$$L(w) = \prod_{i=1}^{M} h(x^i)^{y^i} \left(1-h(x^i)\right)^{1-y^i}$$

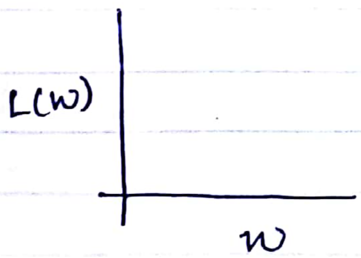
| y | h(x) |
|---|---|
| 0 | 0 |
| 0 | 1 |
| 1 | 1 |
| 1 | 0 |

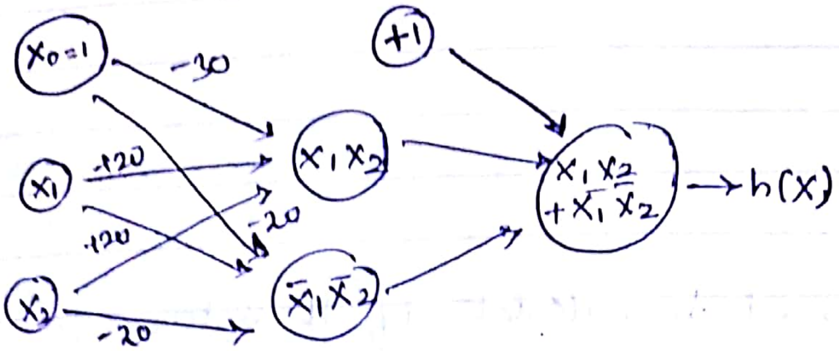
For $h(x)=0$ & $y=0$   $L(w) = 0^0 (1-0)^{1-0} = 1^0 (1)^1 = 1$

$L(w) = 1^0 (1-1)^{1-0} = 1(0) = 0$

$L(w) = 1^1 (1-1)^{1-1} = 1$

$L(w) = 0^1 (1-0)^{1-1} = 0$






$$z_1 = \frac{1}{1+e^{-(x_0 w_0^{(1)} + x_1 w_1^{(1)} + \dots \; x_n w_n^{(1)})}}$$

figure 5.1 Input for $y$ is $z_1, z_2 \dots z_m$

$$y = \frac{1}{1+e^{-(z_0 w_0^{(2)} + z_1 w_1^{(2)} + \dots \; z_n w_n^{(m)})}}$$

$w_{ji}^{(k)} \Rightarrow$ weight of neuron from (i^th neuron) first ~~second~~ layer to (j^th neuron) for ~~third~~ layer

$x_i^{(k)} \Rightarrow$ output of neuron of $k^{th}$ layer.

$a_i^{(k)} \Rightarrow$ summation o/p $k^{th}$ layer.

$$z_1^{(1)} = \frac{1}{1+e^{-a_1^{(2)}}} \qquad a_1^{(2)} = \sum_{i=0}^{d} x_i w_{1i}^{(1)}$$

$P(A \cap B) = P(A|B) \cdot P(B)$

$P(R) = P(R \cap B_1) + P(R \cap B_2)$

$P(R) = P(R|B_1) \cdot P(B_1) + P(R|B_2) \cdot P(B_2) \rightarrow$ Total probability.

$P(A \cap B) = P(A) \cdot P(B|A) + P(B) \cdot P(A|B)$

$P(A \cap B) = P(A) \cdot P(B|A) + P(B) \cdot P(A|B)$

$$P(B|A) = \frac{P(A \cap B)}{P(A)}$$

$$\boxed{P(B|A) = \frac{P(B) \cdot P(A|B)}{P(A)}} \rightarrow \text{Bayes Theorem.}$$

BAYESIAN CLASSIFIER

$X \rightarrow$ feature

$C \rightarrow$ class $\Rightarrow C_1 = 0$ , $C_2 = 0$

Given a feature, we need to calculate if it belongs to $C_1/C_2$.

$P(Ci/X)$

$\hookrightarrow$ Posterior probability.

If $P(C_1|X) < P(C_2|X)$ then $x$ is assigned to $C_2$

$$P(Ci|X) = \frac{P(Ci).P(X|Ci)}{P(X)}$$

$\hookrightarrow$ Total probability (coz feature may occur in both classes)

$$= \frac{P(Ci).P(X|Ci)}{P(C_1).P(X|C_1) + P(C_2).P(X|C_2)}$$

$\rightarrow$ Prior probability

$$= \frac{P(Ci).P(X|Ci)}{\sum_{i=1}^{k} P(Ci)P(X|Ci)}$$

$P(X|Ci) \rightarrow$ class conditional likelihood

$\downarrow$

Evidence

P1

- A doctor knows that meningitis causes stiff neck 50% of the time.
- Prior probability – meningitis – $1/50,000$
- " " – stiff neck – $1/20$

If a patient has stiff neck, P(Menangitis) = ?

$$P(M|S) = \frac{P(M).P(S|M)}{\cancel{P(M).P(S|M) + P(S) + P(\overset{5}{M}|S)}} \rightarrow P(S)$$

$$= \frac{\frac{1}{50,000}\left(\frac{1}{2}\right)}{\cancel{\frac{1}{50,000}\left(\frac{1}{2}\right) \pm} \frac{1}{20} \cancel{\left(\frac{1}{2}\right)}}$$

$$\boxed{P(M|S) = \frac{1}{5000}}$$

Consider the "Cancer test kit" problem, which has the following features. Given that the subject has cancer "c", the probability of the test kit producing a positive decision "+" is =0.98
Probability of the kit producing a negative decision "−" given that the subject is healthy "H" is 0.97. The prior probability of cancer in the population = 0.01
We would like to know the probability that the subject has cancer given that the test kit generated a +'ve decision?

$P(C|+)$

<u>During Training</u>.

$(X, Y)$

Compute: $P(X|C_i)$ for all $C_i$ & for all $X$. & $P(C_i)$

X is multinational

$X \to x_1, x_2 \dots x_p$

$P(X|C_i) = P(x_1, x_2 \dots x_p | C_i)$

$= P(x_1|C_i)\, P(x_2|C_i) \dots P(x_p|C_i)$

| Day | | | | | | |
|---|---|---|---|---|---|---|
| D1 | Sunny | Hot | High | | N. | 2/5 |
| D2 | Sunny | Hot | High | | N | |
| D3 | Overcast | Hot | High | | Y. | |
| D4 | Rain | Mild | High Normal | weak | Y | |
| D5 | Rain | Cool | N | W | Y. | |
| D6 | Rain Overcast | Cool | N | S | N, | |
| D7 | Sunny | Cool | N | S | Y.. | |
| D8 | Sunny | Mild | H | W | N | |
| D9 | | Cool | N | W | Y | |
| D10 | Rain | Mild | N | W | Y. | |
| D11 | Sunny | Mild | N | S | Y. | |
| D12 | Overcast | Mild | H | S | Y. | |
| D13 | Overcast | Hot | N | W | Y. | |
| D14 | Rain | Mild | H | S | N. | |

$C_1 = Yes \quad C_2 = No$

$P(x'|C_1)$ & $P(x'|C_2)$

$x'$ is assigned to $C_1$ if $P(x'|C_1) > P(x'|C_2)$

$$P(C_1|x_1) = \frac{P(C_1)|P(x'/C_1)}{P(x)} = 9/14$$

$$P(C_2|x_2') = \frac{P(C_2)|P(x'|C_2)}{P(x)} = 5/14$$

| | Yes | No |
|---|---|---|
| Sunny | 2/9 | 3/5 |
| Rain | 3/9 | 2/5 |
| Overcast | 4/9 | 0/5 = 0 |

Yes

P(x|Yes)

P(Outlook|Yes).

P(Sunny|Yes) = 2/9.

$x' = $ (Outlook = Sunny, Temperature = Cool, Humidity = High, Wind = Strong)

$P(x'|Yes) = P(Yes)\, P(Sunny|Yes)\, P(Cool|Yes)\, P(High|Yes)\, P(Strong|Yes)$

$P(No|x') = P(No)$

| Tid | Refund | Marital Status | Taxable Income |
|---|---|---|---|
| 1 | Y | S | 125 |
| 2 | N | M | 100 |
| 3 | N | S | |
| 4 | Y | M | |
| 5 | N | D | |
| 6 | N | M | |
| 7 | Y | D | |
| 8 | N | S | |
| 9 | N | M | |
| 10 | N | S | |

Normal Distribution,

$$P(x_i | c = c_i) = \frac{1}{\sqrt{2\pi}\sigma_p} \exp\left(- \frac{(x_j - \mu_p)^2}{2\sigma_\beta^2}\right)$$

$$\mu_{Yes} = \frac{95+85+90}{3} = 90$$

$$\mu_{No} = \frac{125+100 \ldots + 75}{7} =$$

$$\sigma_{Yes}^2 = \frac{(90-95)^2 + (90-85)^2 + (90-90)^2}{2} = \frac{25+25}{2} = 25$$

$P(X | class = NO) =$

## Entropy

$$(D) = \sum_{i=1}^{k} - p_m^i \log_2 p_m^i$$

$$Gain(D, A) = Entropy(D) - \sum_{j=1}^{v} \frac{|D_j|}{|D|} Entropy(D_j)$$

↳ for an attribute.

A → attribute considered

v → total outcome for an attribute.

$D_j$ → Subset satisfying the attribute.

Dataset H, if a student buys computer/not.

$Y_1 = yes \quad Y_2 = no$

$$Entropy(D) = \sum_{i=1}^{k} - p^i \log_2 p^i$$

$$= -9/14 \log_2 9/14 - 5/14 \log_2 5/14.$$

$$= -9/14(-0.6375) - 5/14(-1.4856)$$

$$= 0.4098 + 0.5306$$

$$= \underline{0.9404}$$

$$Gain(Age) = 0.9404 - \left[ +\frac{5}{14} Entropy(Young) + \frac{4}{14} Entropy(middle) + \frac{5}{14} Entropy(Senior) \right]$$

$$Entropy(Y) = -\frac{2}{5} \log_2 2/5 - \frac{3}{5} \log_2 3/5 = +0.5288 + 0.4422$$

$$= 0.971$$

$$Entropy(M) = -\frac{4}{4} \log_2 4/4 = \underline{0}$$

$$Entropy(S) = -3/5 \log_2 3/5 - 2/5 \log 2/5 = \underline{0.971}$$

Scanned by CamScanner

$$Gain = 0.9404 - \left[\frac{5}{14}(0.971) + \frac{4}{14}(0) + \frac{5}{14}(0.971)\right]$$

$$= 0.9404 - [0.6936]$$

$$= \cancel{0.0038}$$

$$= \underline{0.2468}$$

Gain (Income) = 0.029

Gain (student) = 0.151

Gain (credit) = 0.048

High value is for age ∴ Select age as the root.

## K-MEANS.

$$d_{ij} = \sqrt{\sum_{i=1}^{p} (x_{il} - x_{jl})^2}$$

Distance b/w $x_i$ & $x_j$ can be elaborated as $x_i = (x_{i1}, x_{i2} \dots x_{ip})$

$x_j = (x_{j1}, x_{j2} \dots x_{jp})$

Ex:

| Ex | x | y |
|---|---|---|
| 1 | 0.5 | − |
| 2 | 3.0 | − |
| 3 | 4.5 | + |
| 4 | 4.6 | + |
| 5 | 4.9 | + |
| 6 | 5.2 | − |
| 7 | 5.3 | − |
| 8 | 5.5 | + |
| 9 | 7.0 | − |
| 10 | 9.5 | − |

Test point is 5.

$V1 = \sqrt{(5-0.5)^2} = 4.5$

$V_2 = \sqrt{(5-3)^2} = 2$

$V_3 = \sqrt{(5-4.5)^2} = 0.5$

$V_4 = \sqrt{(5-4.6)^2} = 0.4$

$V5 = \sqrt{(5-4.9)^2} = 0.1$

$V_6 = \sqrt{(5-5.2)^2} = 0.2$

$V7 = \sqrt{(5-5.3)^2} = 0.3$

$V8 = \sqrt{(5-5.5)^2} = 0.5$

$V9 = \sqrt{(5-7)^2} = 2$

$V_{10} = \sqrt{(5-9.5)^2} = 4.5$

Nearest is 0.1 i.e from 5 to 4.9 (I nearest neighbour)
class label for 5 is '+'.

3 Nearest neighbour points are 6th & 7th, majority is -.
∴ class label of 5 is -

Problem

| Example | x1 (Acidity) | x2 (Strength) | Classification |
|---|---|---|---|
| 1 | 7 | 7 | Bad |
| 2 | 7 | 4 | Bad |
| 3 | 3 | 4 | good |
| 4 | 1 | 4 | Good. |

Test Data = (3, 7)  k = 3

$\sqrt{(3-7)^2+(7-7)^2} = 4$

$\sqrt{(3-7)^2+(7-4)^2} = \sqrt{16+9} = 5$

$\sqrt{(3-3)^2+(7-4)^2} = 3$

$\sqrt{(3-1)^2+(7-4)^2} = \sqrt{4+9} = \sqrt{13} = 3.6055$

The value of K = 3

3, 3.6055, 4.

3, 4, 1

Good Good Bad

∴ Max Good.

<u>UNSUPERVISED LEARNING</u>

2 Algorithms → Clustering → K-means

<u>K means</u>

k → no of clusters.

centroid → means of the clusters

Scanned by CamScanner

Problems

1) 

| Data Point | X1 | X2 |
|---|---|---|
| A | 1 | 1 |
| B | 1 | 0 |
| C | 0 | 2 |
| D | 2 | 4 |
| E | 3 | 5 |

Create two clusters. Let A & C be the initial clusters.

$C_1 = (1,1)$ $\quad C_2 = (0,2)$

$C_1A = 0$

$C_1B = 1$

$C_1C = \sqrt{2} = 1.414$

$C_1D = \sqrt{10} = 3.1623$

$C_1E = \sqrt{20} = 4.4721$

$C_2A = \sqrt{2} = 1.414$

$C_2B = \sqrt{5} = 2.2361$

$C_2C = 0 = 0$

$C_2D = \sqrt{8} = 2.8284$

$C_2E = \sqrt{18} = 4.2426$

| Data point | $x_1$ | $x_2$ | $C_1$ | $C_2$ | Assign (Based on the least value b/w $C_1, C_2$) |
|---|---|---|---|---|---|
| A | 1 | 1 | 0 | 1.4 | $C_1$ |
| B | 1 | 0 | 1 | 2.2 | $C_1$ |
| C | 0 | 2 | 1.4 | 0 | $C_2$ |
| D | 2 | 4 | 3.2 | 2.8 | $C_2$ |
| E | 3 | 5 | 4.5 | 4.2 | $C_2$ |

New centroid

Avg $= \frac{A+B}{2} = C_1$ $\quad \frac{C+D+E}{3} = C_2$

$\therefore C_1 = \left(\frac{1+1}{2}, \frac{1+0}{2}\right) = (1, 0.5)$

$C_2 = \left(\frac{0+2+3}{3}, \frac{2+4+5}{3}\right)$

$= (1.6667, 3.6667)$

$\therefore C_1 = (1, 0.5)$ $\quad C_2 = (1.6667, 3.6667)$

$C_1A = 0.5$

$C_1B = 0.5$

$C_1C = \sqrt{1+2.25} = 1.803$

$C_1D = \sqrt{1+(3.5)^2} = 3.640$

$C_1E = \sqrt{4+(4.5)^2} = 4.924$

Continues....

# HIERARCHICAL CLUSTERING.

- Distance b/w the adjacent points & not the mean as in k-means.

$$d(G_i, G_j) = \min_{x^r \in G_i, x^s \in G_j} \{ d(x^r, x^s) \}$$ // distance b/w each group & consider the minimum

| Data Pt | X1 | X2 |
|---|---|---|
| A | 1 | 1 |
| B | 1 | 0 |
| C | 0 | 2 |
| D | 2 | 4 |
| E | 3 | 5 |

| | A | B | C | D | E |
|---|---|---|---|---|---|
| A | 0 | 1 | 1.41 | 3.16 | 4.47 |
| B | 1 | 0 | 2.232 | 4.123 | 5.385 |
| C | 1.41 | 2.232 | 0 | 2.82 | ~~4.243~~ 3.46 ✓ |
| D | 3.16 | 4.123 | 2.82 | 0 | 1.41 |
| E | 4.47 | 5.385 | ~~4.243~~ 3.46 ✓ | 1.41 | 0 |

Consider the smallest, group them. Here it is 1 ∴ Merge A,B

| | AB | C | D | E |
|---|---|---|---|---|
| AB | 0 | 1.41 | 3.16 | 4.47 |
| C | 1.41 | 0 | 2.82 | 4.24 |
| D | 3.16 | 2.82 | 0 | 1.41 |
| E | 4.47 | 4.24 | 1.41 | 0 |

1.41 & 2.232
Min [(AB, C)] ∴ 1.41

3.16 & 4.123
Min [(AB, D)] ∴ 3.16

4.47 & 5.385
Min [(AB, E)] ∴ 4.47

Merge CD

| | ABC | ED | ~~E~~ |
|---|---|---|---|
| ABC | 0 | 2.82 | |
| ED | 2.82 | 0 | |
| ~~E~~ | | | |

4.24 & 2.82
Min [(ABC, E) & Min (ABC, D)]
∴ 2.82

Scanned by CamScanner