DM

Assignment 02.

① Explain C4.5 differ from ID3 algorithm.

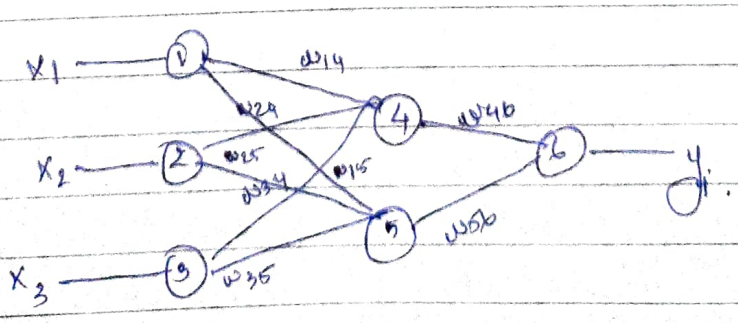
→ ID3 & C4.5 are used to generate decision tree which is used in natural language processing domain.

→ They use the entropy of an attribute & picks the attribute with highest reduction in entropy to determine which attribute should the data be split.

Ofference

	ID3	C4.5
Splitting Criteria	Information Gain	Gain Ratio.
Attribute type	Handles only categorical value.	handles both categorical & numerical value.
Missy value.	Do not handle missy value.	Handles missy value.
Prunny strategy	No prune is done	Error based prunny is used
Outlier election	Suspectible to outlier	Suspectible to outlier.

② .



x_1	x_2	x_3	w_{14}	w_{15}	w_{24}	w_{25}	w_{34}	w_{35}	w_{46}	w_{56}
1	0	1	0.2	-.3	.4	0.1	-.5	.2	-.3	-0.2

Sigmoid func $\Rightarrow \frac{1}{1+e^x}$

net o/p $\Rightarrow \sum_{i=0}^{n} w_i x_i + b_j$

unit	net i/p	o/p.
4	$(.2*1) + 0 + (0.5*1) - .4 = -0.7$	$1/(1+e^{-.7}) = -.332$
5	$(-.3*1) + 0 + (0.2*1) + 0.2 = 0.1$	$1/1+e^{-0.1} = 0.525$
6	$(-0.3 * 0.332) + (0.2 * 0.525 + 0.1 = -0.105$	$1/1+e^{-0.105} = 0.474$

$y_4 = (w_{14} * x_1) + (w_{24} x_2) + (w_{34} x_3) + \theta_4$

$y_5 = (w_{15} x_1) + (w_{25} x_2) + (w_{35} x_3) + \theta_5$

$y_6 = (w_{46} y_4) + (w_{56} y_5) + \theta_6$