Nitte Meenakshi Institute of Technology
Department of Computer Science & Engineering

K-Nearest Neighbour

Data Pre-processing and Classification

1. (a) Apply Min -Max Normalization to **transform** *Age, Gross Salary and Deduction* Columns of the given dataset to [0,1] range.

$$v' = \frac{v - min_A}{max_A - min_A}(new\_max_A - new\_min_A) + new\_min_A$$

Payroll Dataset			
Name	Age	Gross Salary	Class
Mike	38	5500	Low
Mary	36	4000	Low
Bill	41	9900	Med
Jim	42	12200	Med
Dave	50	15600	High
Anne	58	16500	High

Solution:

AgeN	GrossN
0.09	0.12
0	0
0.23	0.47
0.27	0.66
0.64	0.93
1	1

(b) Use K Nearest Neighbors Algorithm to train the normalized observations from the above dataset and test the below observations. Let K = 3.

Note: Use Euclidean Distance

Name	Age	Gross Salary	Class
Minnie	40	10350	?
Jack	57	16300	?

Solution:

Name	Age	Gross Salary	AgeN	GrossN	Distance1	Distance2		Rank1	Rank2
Mike	38	5500	0.09	0.12	0.40024992	1.21622366	L	3	5
Mary	36	4000	0	0	0.54083269	1.36488095	L	4	6
Bill	41	9900	0.23	0.47	0.06403124	0.88232647	M	1	4
Jim	42	12200	0.27	0.66	0.17492856	0.75153177	M	2	3
Dave	50	15600	0.64	0.93	0.62289646	0.31400637	H	5	2
Anne	58	16500	1	1	0.95524866	0.05385165	H	6	1

Name	Age	Gross Salary	Class	AgeN	GrossN				
Minnie	40	10350	?	0.18	0.51	M	M	L	M
Jack	57	16300	?	0.95	0.98	H	H	M	H