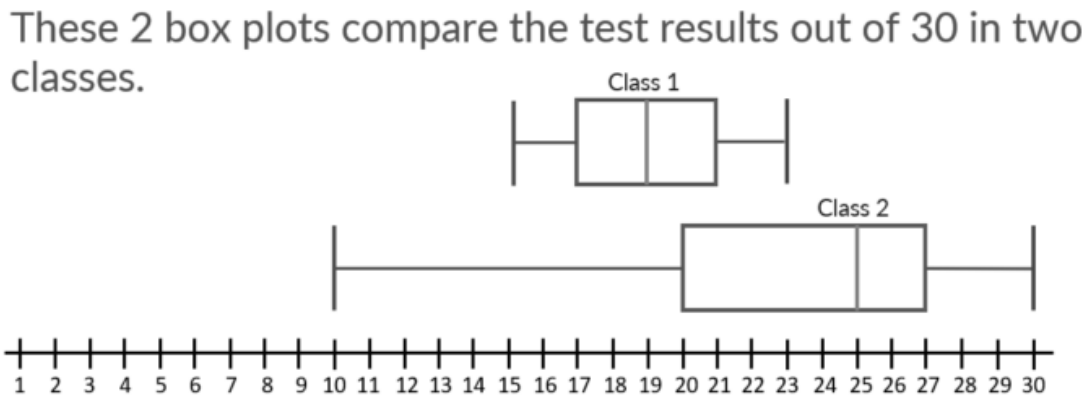
PSG College of Technology

M. Sc (TCS) - Semester 8
20XT83 - Data Mining
Tutorial

1. a.i) Analyse the box plots given and write your interpretation. (4)



b. Using the data for age 13, 15, 16, 16, 19, 20, 20, 21, 22, 22, 25, 25, 25, 25, 30, 33, 33, 35, 35, 35, 35, 36, 40, 45, 46, 52, 70.
(a) Use smoothing by bin means to smooth these data, using a bin depth of 3. Illustrate your steps. Comment on the effect of this technique for the given data.
(b) Use min-max normalization to transform the value 35 for age onto the range [0.0,1.0].

c) Briefly outline how to compute the dissimilarity between objects described by the following: (a) Nominal attributes (b) Asymmetric binary (c) Numeric attributes

2. a.i) Mention the various methods for improving the efficiency of the Apriori Algorithm. (2)
   ii) Given the Conditional Pattern base, generate the frequent itemsets. (2)

Item	Conditional Pattern Base	Conditional FP-tree
I5	{{I2, I1: 1}, {I2, I1, I3: 1}}	⟨I2: 2, I1: 2⟩
I4	{{I2, I1: 1}, {I2: 1}}	⟨I2: 2⟩
I3	{{I2, I1: 2}, {I2: 2}, {I1: 2}}	⟨I2: 4, I1: 2⟩, ⟨I1: 2⟩
I1	{{I2: 4}}	⟨I2: 4⟩

b. Given the transactional database. (6)

Customer ID	Transaction ID	Items Bought
1	1	{a,d,e}
1	2	{a,b,c,e}
2	3	{a,b,d,e}
2	4	{a,c,d,e}
3	5	{b,c,e}
3	6	{b,d,e}
4	7	{c,d}
4	8	{a,b,c}
5	9	{a,d,e}
5	10	{a,b,e}

1. Considering transaction ID as a market basket, find support for each itemset {e}, {b,d}, and {b,d,e}
2. Based on the results in the previous question, find the confidence of association rules {b,d}->{e} and {e}->{b,d}
3. Considering customer ID as market basket. An item is treated as 1 if it appears in at least one transaction done by the customer, 0 otherwise. Find the support of itemsets {e}, {b,d}, {b,d,e}.
4. Based on the previous question, find the confidence of association rules {b,d}->{e} and {e}->{b,d}

c) A database has five transactions. Let min sup D 60% and min conf D 80%.

TID	items_bought
T100	{M, O, N, K, E, Y}
T200	{D, O, N, K, E, Y }
T300	{M, A, K, E}
T400	{M, U, C, K, Y}
T500	{C, O, O, K, I, E}

Find all frequent itemsets using Apriori Algorithm.