



END SEMESTER ASSESSMENT (ESA) M.TECH. I SEMESTER-DEC 2015/Jan 2016

UE15CS521 - DATA ANALYTICS

Time: 3 Hrs

Answer All Questions

Max Marks: 100

1	a)	What are point estimates? Why cannot we rely on point estimates to estimate parameters of the population? What are quantitative and qualitative data?	12
	c)	What is binning? Mention the various methods employed in binning and write a brief about it.	8
2	a)	Suppose that the data for analysis includes the attribute <i>salary</i> . The <i>salary</i> values for the data tuples are: 13000, 15000, 16000, 16000, 19000, 20000, 21000, 22000, 22000, 25000, 25000, 25000, 30000, 33000, 33000, 35000, 35000, 35000, 36000, 40000, 45000, 46000, 52000, 70000 a. What is the <i>mean</i> of the data? b. What is the <i>median</i> ? c. What is the <i>mode</i> of the data? d. Which data will you use to find out the <i>average</i> income of the tuple? I Assuming this data to be income of individuals belonging to a community, how do I find the <i>mid-income</i> group?	8
	b)	What is <i>Euclidean distance</i> ? What are its <i>advantages</i> ? Discuss under what circumstances does it fail to give accurate measurement, with <i>an example</i>	12
3	a)	What is time series? Explain the areas of application of time series?	8
	c)	What are the basics types of time series – name them and write about them. Explain data reduction strategies.	12
4	a)	A brief note on histograms, box plot, steam-leaf plots	8
	b)	What is smoothing? How is it different from running/moving averages?	12
5	a)	What are the goals of clustering? What are the uses of clustering?	8
	b)	What are hierarchical clustering? What is a dendrogram? Is top-down or bottom-up more accurate?	12