

PROBLEM SET 1

BCB4003/503 CS4083/583 Biological and Biomedical Database Mining – A term / Fall 2013

Prof. Carolina Ruiz

Student: _____

Data Formats:	/10
• Converting dataset to Weka's arff input format	
• Converting dataset to Matlab's dat input format	/10
Data Exploration:	
• Observations and visualizations	/10
• Co-variance matrix and visualization	/07
• Co-relation matrix and visualization	/08
• Selection of 3 attributes for removal	/05
Data Preprocessing:	
• Sampling: ___ supervised; ___ unsupervised; ___ distribution	/15
• Discretization: ___ equalFreq.; ___ equalWidth; ___ supervised	/15
• Weka's discretization code	/05
• Missing Values	/05
• Feature Selection	/10
Optional Part	/20
TOTAL	/100

Class Presentation & participation during presentations:

_____/10

Total Problem Set 1:

_____/100