

Data Mining

Homework Assignment #4

Dmytro Fishman, Anna Leontjeva and Jaak Vilo

March 12, 2014

You are free to use any programming language you are comfortable with. Hints in R are optional. Check appendix for more hints at the end of the document.

Task 1

Listen to the presentation by Tamara Munzner: Keynote on Visualization Principles - <http://vizbi.org/Videos/26205288> (use the PDF slide-deck from there as well <http://bit.ly/nCJM5U>)

Task 2

Summarise the key take-home messages from her presentation on one nice page.

Task 3

Read data from file, data.txt, calculate mean, median, variance and standard deviation for the each feature (column). Compare them. Do you see the difference? Plot each feature separately and observe the trick :)

Task 4

Perform a "Single Link" clustering of 2-D data from slide 28. Use Euclidean distance as a distance measure. Draw a dendrogram/tree with node height at the distance at where the clusters were merged. Hint: Draw the points first on 2D and then perform manual simulation. (Solutions on paper are ok :-)

	X	Y
A	2	4
B	7	3
C	3	5
D	5	3

E	7	4
F	6	8
G	6	5
H	8	4
I	2	5
J	3	7

Task 5

Perform the same clustering on the following data set ([link to the USArrests](#)).

Task 6 (2pt)