

Homework №6 (deadline: 23:59 12.12.2015)

Acceptable extension: .pdf

Email to send: amidatamining@gmail.com

Filename example: Смолов_Кокорин_Дзюба_hw16.pdf.

1. **(2)** In your data set, select at least 3 features related to the same aspect and explain your choice
2. **(6)** Visualize the data with these features using standardization with the normalization both over ranges and standard deviations. At these visualizations, use a distinct shape/colour for points representing a prespecified group of objects. Also, apply the conventional PCA for the visualization. Make a comment on which of the normalizations is better and why.
3. **(4)** Compute and interpret a hidden factor behind the selected features. The factor should be expressed in a 0-100 rank scale as well as the features.

Homework №7 (deadline: 23:59 19.12.2015)

1. **(6)** K-Means Clustering at K=3,4,7 clusters. Comment and explain which is better.
2. **(6)** Nearest Neighbor clustering with MST for 7 clusters. Compare with that at K-means and explain which is better.