

coursework12

Fortunat Mutunda

May 15, 2016

There is an image with shuffled rows - <http://biit.cs.ut.ee/imgshuffle/index.cgi?fname=DM2016&dname=DM2016> You can

get access to RGB values in here - <http://biit.cs.ut.ee/imgshuffle/data/DM2016/DM2016.txt> (uploaded file here) You can re-order row id's in any order and upload them to the same webpage to recover image in that new order of rows. New image for those who know what is in the first image and want to have fun finding out what's in the picture (txt file).

- Apply any clustering techniques (hierarchical, SOM, K-Means) that you wish and try to recover what is pictured on the image.
- Use the same data matrix from task 1 and run a PCA analysis on it. Plot first three principal components as 2-dimensional plots PC1-PC2, PC1-PC3, PC2-PC3 of these data or as a 3D plot. Check out PCA example
- Grab US census data (e.g. medium size) in here - <http://biit.cs.ut.ee/~vilo/edu/Data/census2000/> Make Pivot table summary about people's earnings based on various variables. E.g. the gender and education level. Make sure to apply heatmaps on top of pivot table.
- On the same data - try to visualize other relationships in data - based on ancestry, industry, marital status and education, for example.
- Read the Jim Gray - Data Cube abstraction. Describe the key operators from this article using examples based on above census data (tasks 3-4). (Alternative list of operations - https://en.wikipedia.org/wiki/OLAP_cube#Operations)
- (Bonus 2p) Attempt running a TSP or other techniques to recover as well as possible the original image of tasks 1.-2.
- (Bonus 2p) Load the same census data sets (you can attempt larger ones, too) into a DB and run SQL queries to achieve summarization as in pivot tables.