

Homework #5: Image Analytics

Gather a collection of 50 to 100 images (the more, the better), possibly related to your research or interest. If you can not find those, find any image set of your liking on the web or even in your photo album. Make sure the images are in jpg or png format, and not too big to avoid overburdening the embedding server. Place the image files in a folder (and sub-folders, to indicate classes), and load them using the Import Images widget. Check them out in Image Viewer to make sure they have loaded correctly.

Now apply the skills you have learned in this class to get insight into your image set.

Cluster images using either hierarchical clustering or k-means and comment on the quality and meaningfulness of clusters.

Arrange your images into groups (classes) by placing them in appropriate sub-folders. Can image classes be predicted from image embeddings, that is, from their vector-based descriptions you get from Image Embedding widget? Report on cross-validated accuracy (if your image set is small, use 3-fold or 5-fold cross-validation instead of default 10-fold cross-validation). You can also comment on the types of mistakes that your selected learner makes (e.g., use Confusion Matrix widget).

Project images into two-dimensional space: use either PCA, MDS or t-SNE. Report if the projection makes sense. For illustration, you can include the "image map": a graph with the projection of images and points marked or labeled with class. Comment on the groups you can spot from this visualization.

Include anything else that you can think of, and that makes sense.

Submit your homework as a short report in PDF where you tell us about the data and the analysis results. The report should include a title of the homework, your name, and email. It should be at most three pages long (this limit is strict!), use 11 pt Times. Submit your homework as a PDF (not as a Word!) document to bzupan@gmail.com with subject "DM-HW5". The deadline is 9:00 a.m., Wednesday, February 17.