

Machine Learning Course - CS-433

Unsupervised Learning

Nov 8, 2016

©Mohammad Emtiyaz Khan 2015

minor changes by Martin Jaggi 2016



ÉCOLE POLYTECHNIQUE
FÉDÉRALE DE LAUSANNE

Unsupervised learning

How can systems learn to represent particular input patterns in a way that reflects the statistical structure of the overall collections of input patterns? This question is the central focus of [unsupervised learning](#).

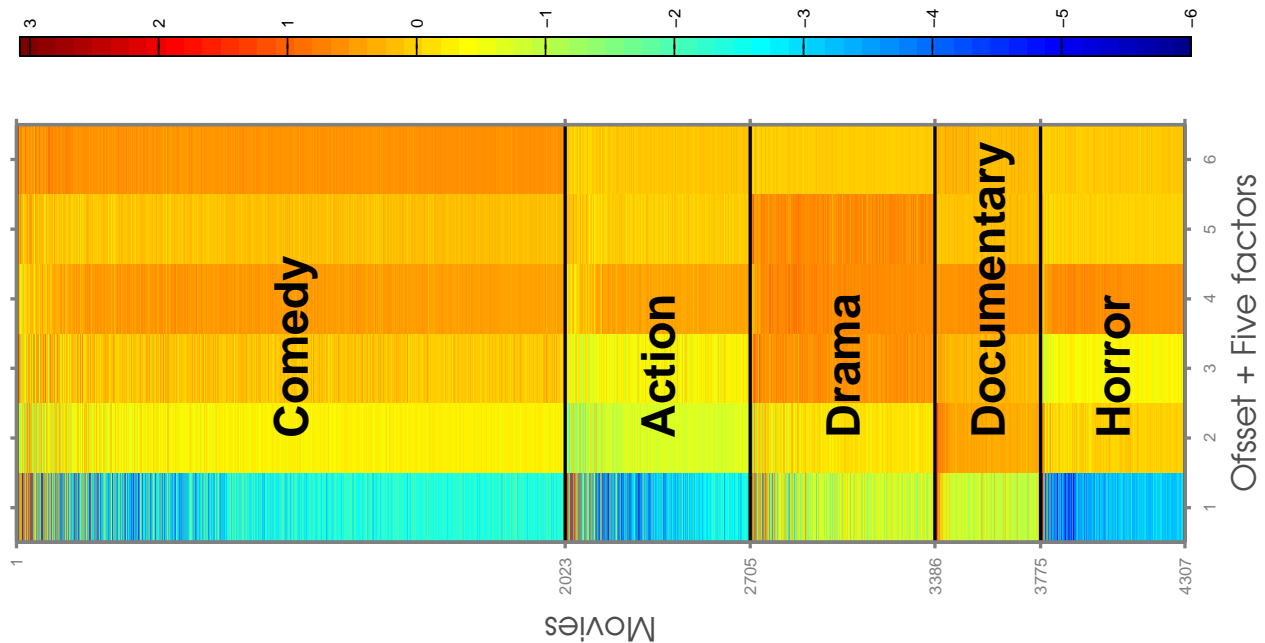
In unsupervised learning, our data consists only of features (or inputs) $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N$, vectors in \mathbb{R}^D , and there are **no outputs** y_n available.

Unsupervised learning seems to play an important role in how living beings learn. Variants of it seem to be more common in the brain than supervised learning.

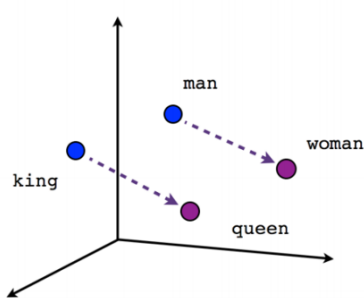
The two most common types of unsupervised learning are [density estimation](#) and [feature learning / representation learning](#).

Examples

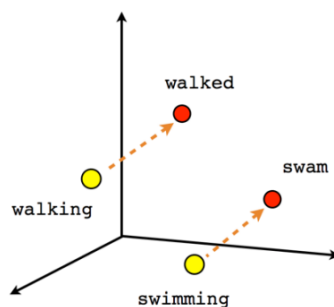
Given ratings of movies and viewers, we use matrix factorization to extract useful features (Khan et al. 2014).



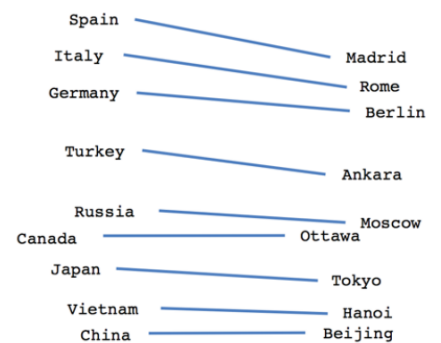
Learning word-representations using matrix-factorizations, **word2vec** (Mikolov et al. 2013).



Male-Female



Verb tense



Country-Capital

Given voting patterns of regions across Switzerland, we use PCA to extract useful features (Etter et al. 2014).

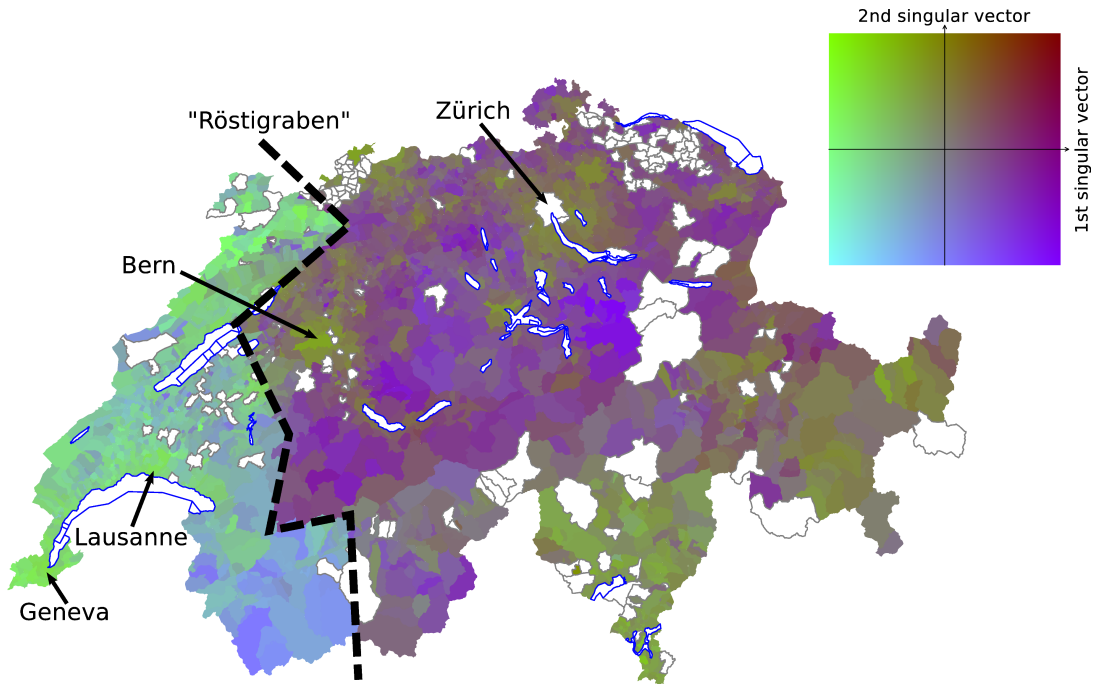


Figure 9: Voting patterns of Swiss municipalities. The color of a municipality is assigned using its location in Figure 8 and the color gradient shown in the upper right corner. Two municipalities with similar colors have similar voting patterns. The *Röstigraben*, corresponding to the cultural difference between French-speaking municipalities and German-speaking ones, is clearly visible from the difference in voting patterns. Regions shown in white are lakes or municipalities for which some vote results are missing (due to a merging of municipalities, for example). A more detailed map can be found online [2].

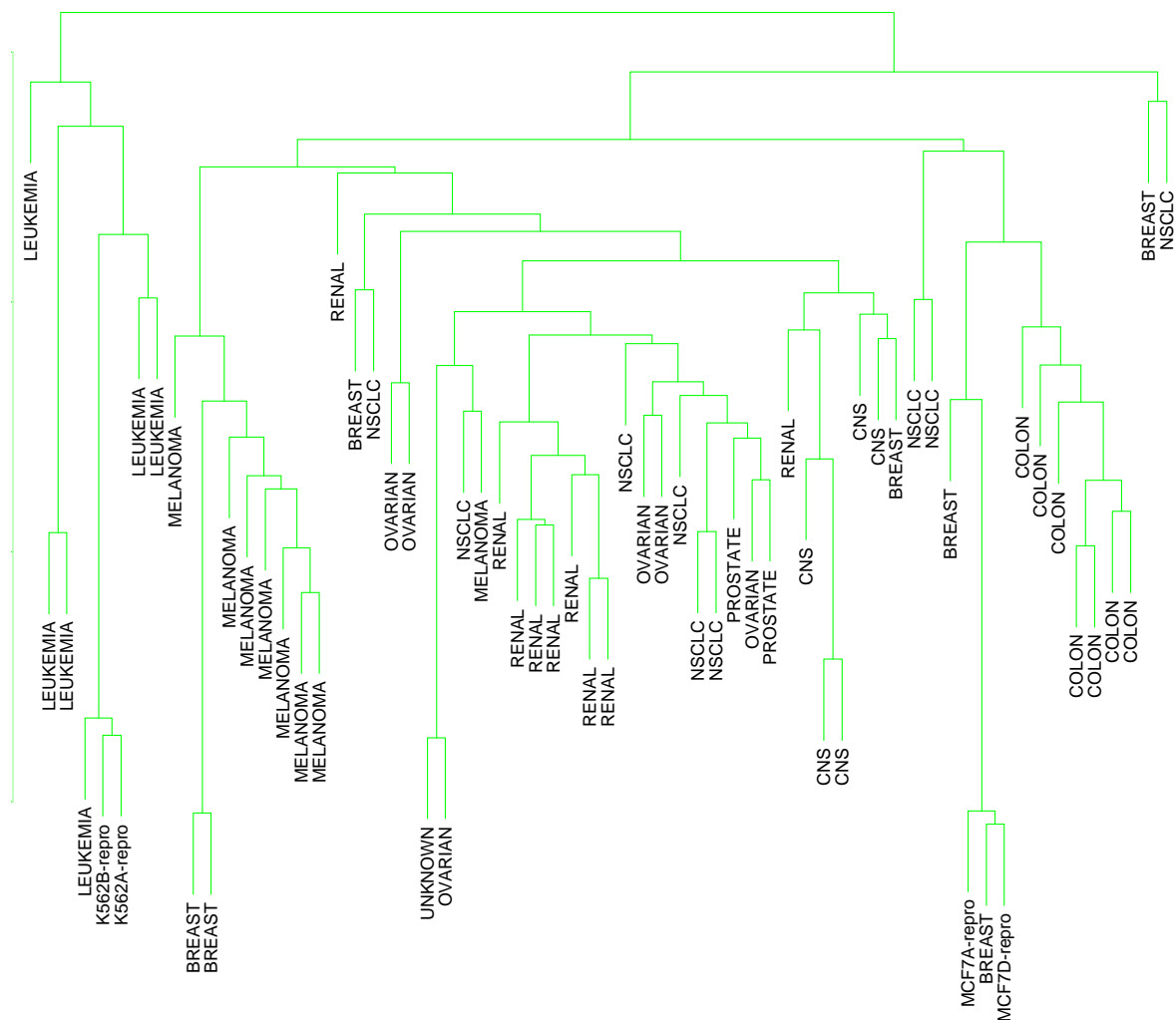


FIGURE 14.12. *Dendrogram from agglomerative hierarchical clustering with average linkage to the human tumor microarray data.*

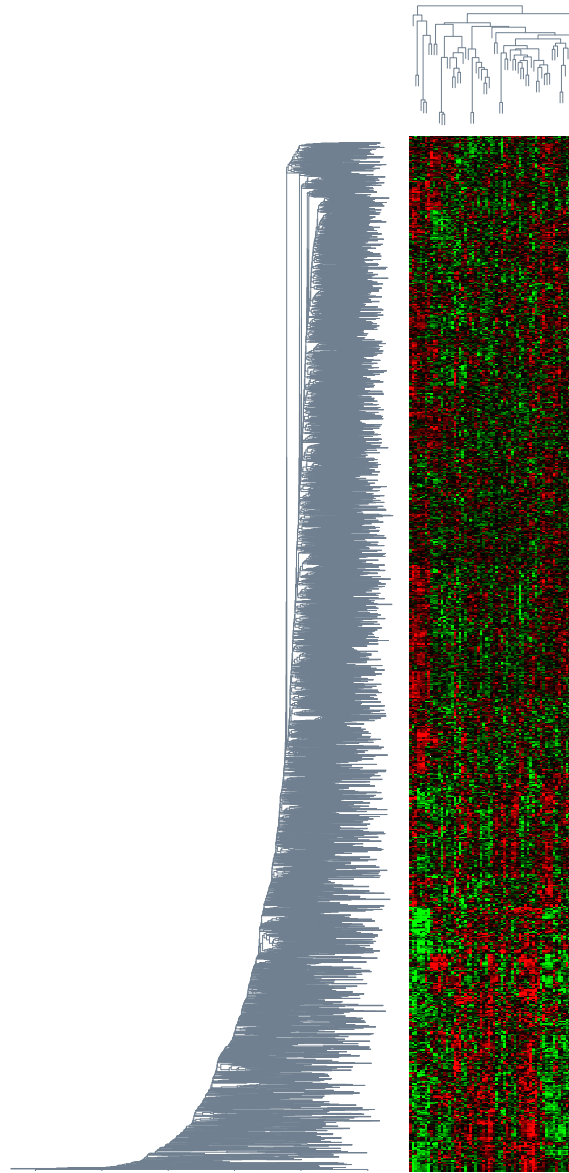
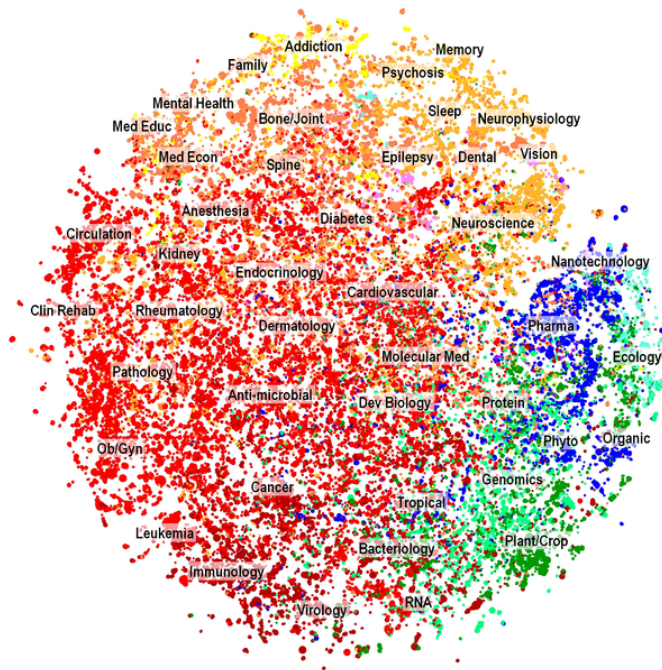


FIGURE 14.14. *DNA microarray data: average linkage hierarchical clustering has been applied independently to the rows (genes) and columns (samples), determining the ordering of the rows and columns (see text). The colors range from bright green (negative, un-*

Clustering more than two million biomedical publications (Kevin Boyack et.al. 2011)



Clustering articles published in Science (Blei & Lafferty 2007)

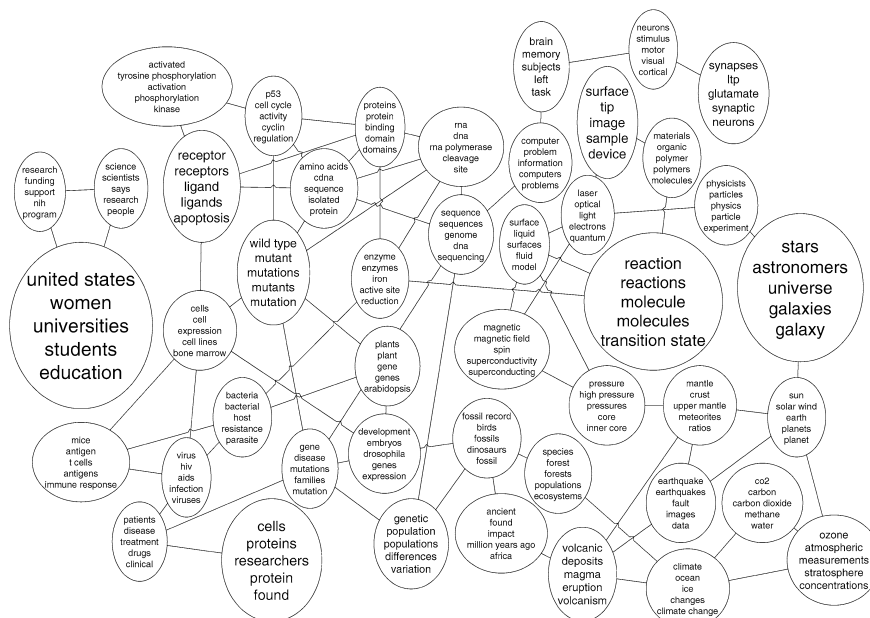


FIG. 2. A portion of the topic graph learned from 16,351 OCR articles from Science (1990–1999). Each topic node is labeled with its five most probable phrases and has font proportional to its popularity in the corpus. (Phrases are found by permutation test.) The full model can be found in <http://www.cs.cmu.edu/~lemur/science/> and on STATLIB.