Machine Learning Course - CS-433

Unsupervised Learning

Nov 7, 2017

minor changes by Martin Jaggi 2016

changes by Martin Jaggi 2017

Last updated: November 7, 2017

ÉCOLE POLYTECHNIQUE
FÉDÉRALE DE LAUSANNE

Unsupervised learning

How can systems learn a meaningful internal representation for data examples? I.e., to represent them in a way that reflects the semantic and statistical structure of the overall collection of input patterns? This question is the central focus of unsupervised learning.

In unsupervised learning, our data consists only of features (or inputs) $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N$, 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.

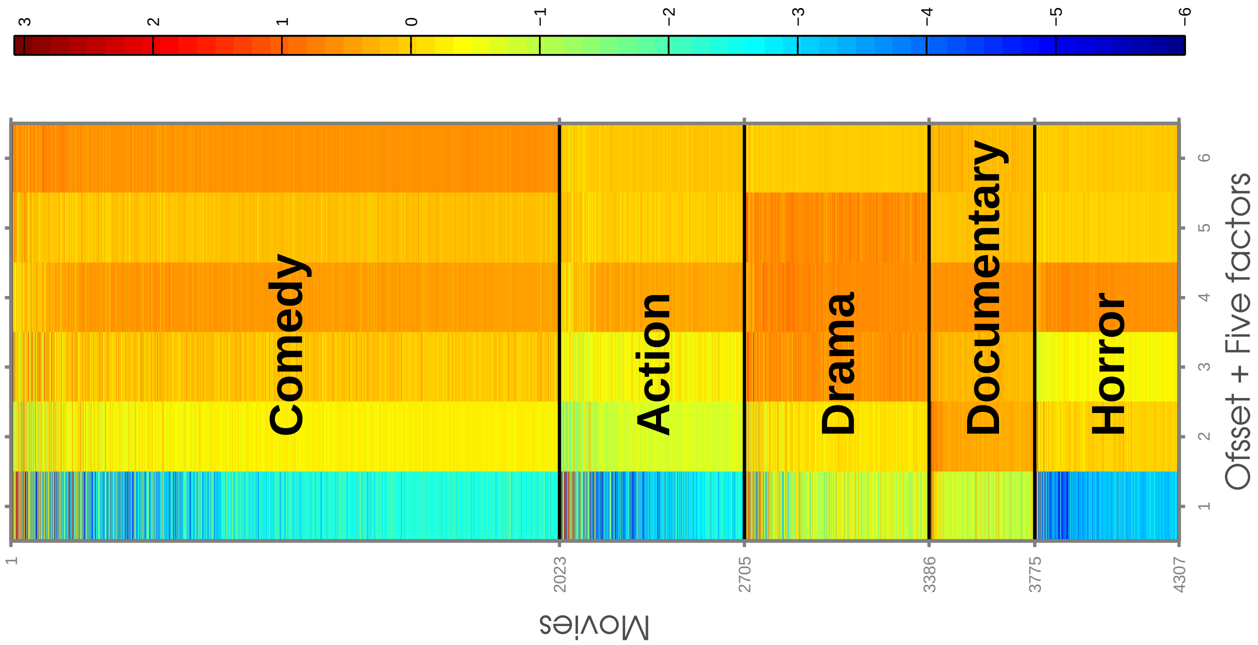
Two main directions in unsupervised learning are

- representation learning & feature learning and
- density estimation & generative models

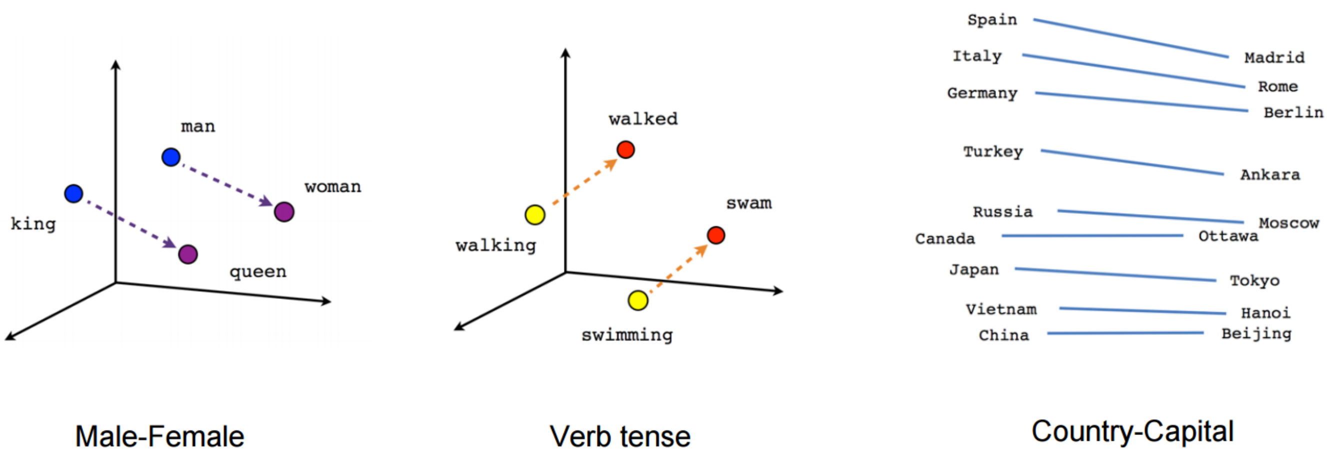
Examples

Examples for Representation Learning

Given ratings of movies and viewers, we use matrix factorization to extract useful features (see e.g. Netflix Prize).



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



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

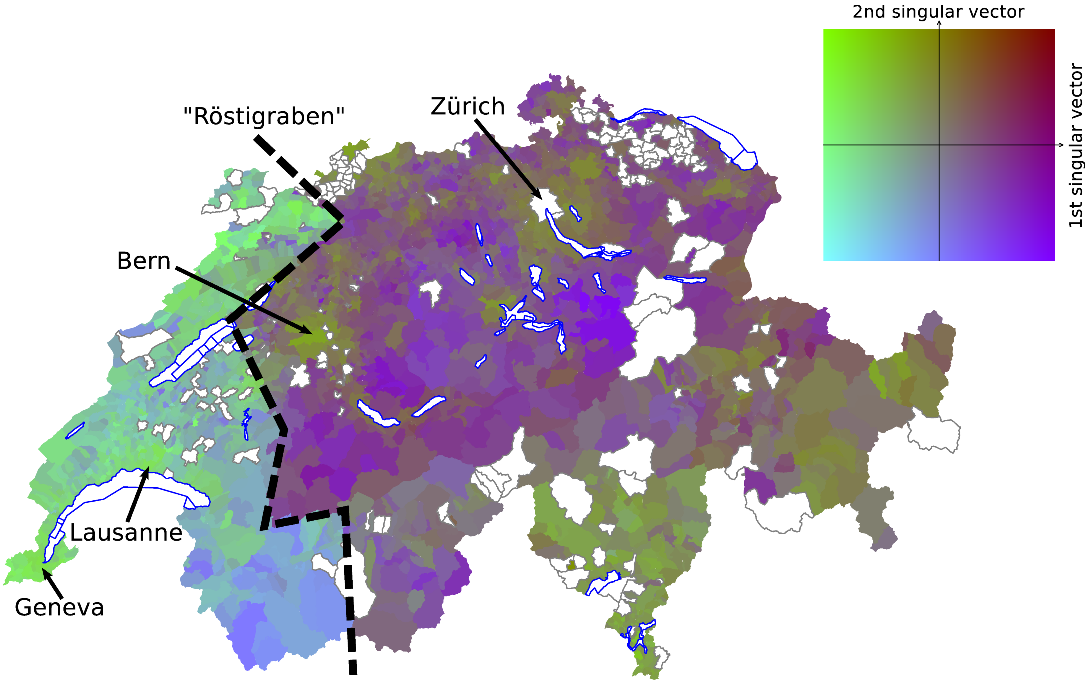


Figure 9: Voting patterns of Swiss municipalities. The color of a municipality is assigned using its location in Figure 8 and the color gradient shown in the upper right corner. Two municipalities with similar colors have similar voting patterns. The *Röstigraben*, corresponding to the cultural difference 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].

Examples for Clustering

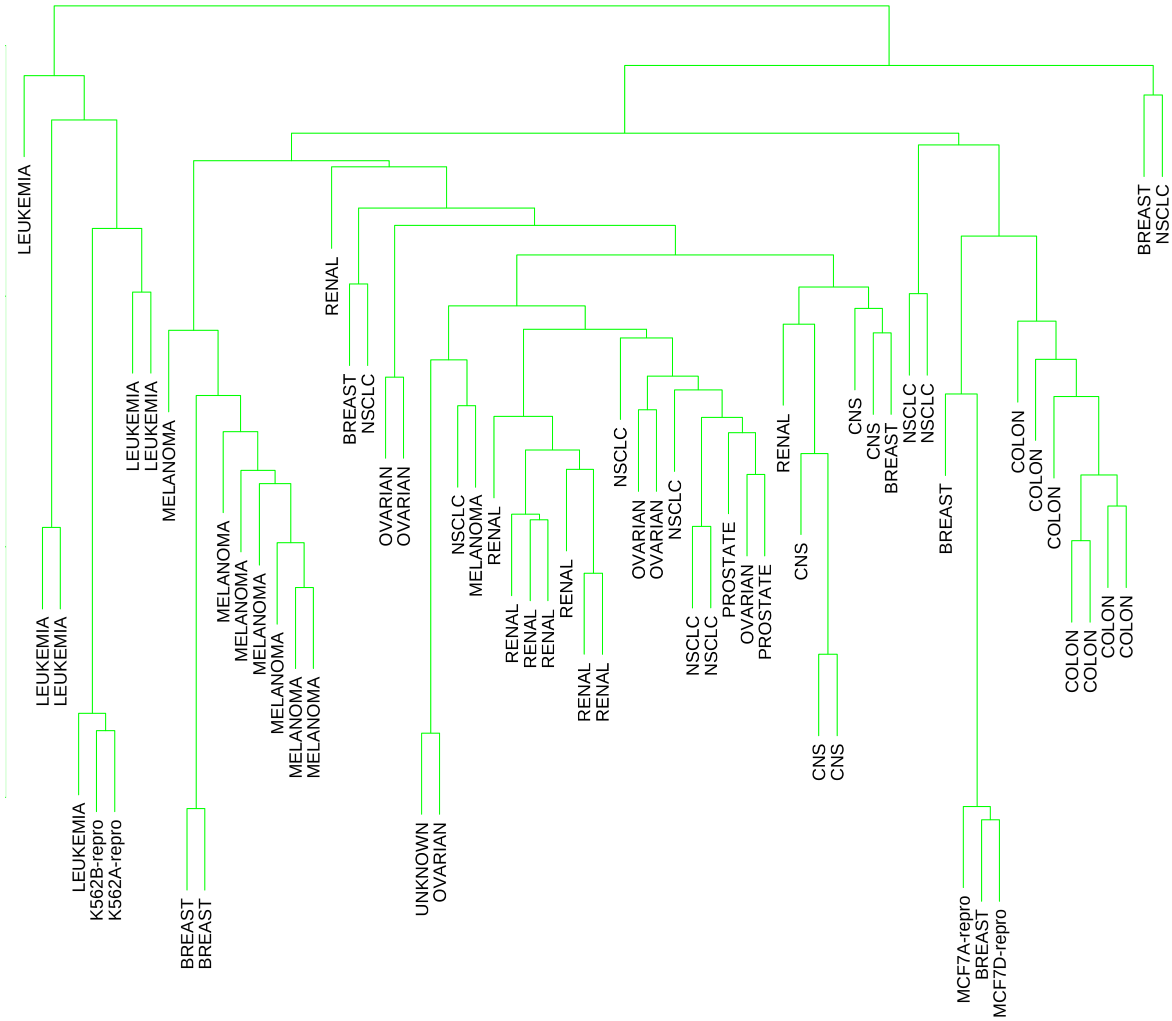


FIGURE 14.12. *Dendrogram from agglomerative hierarchical clustering with average linkage to the human tumor microarray data.*

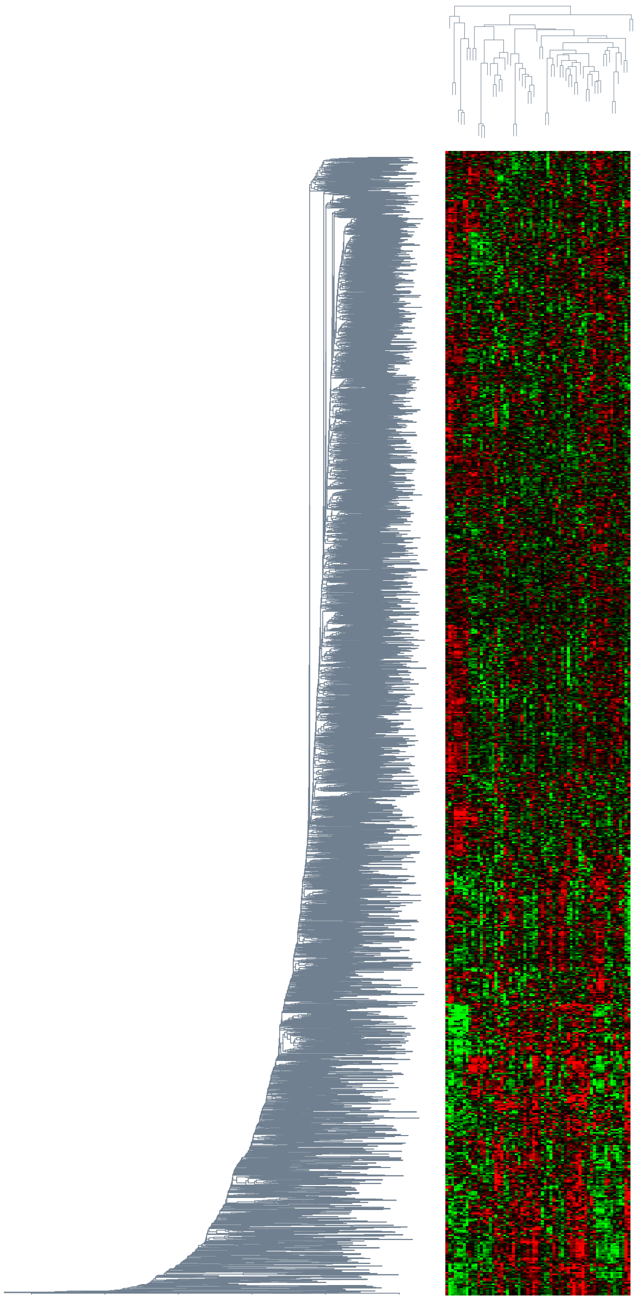
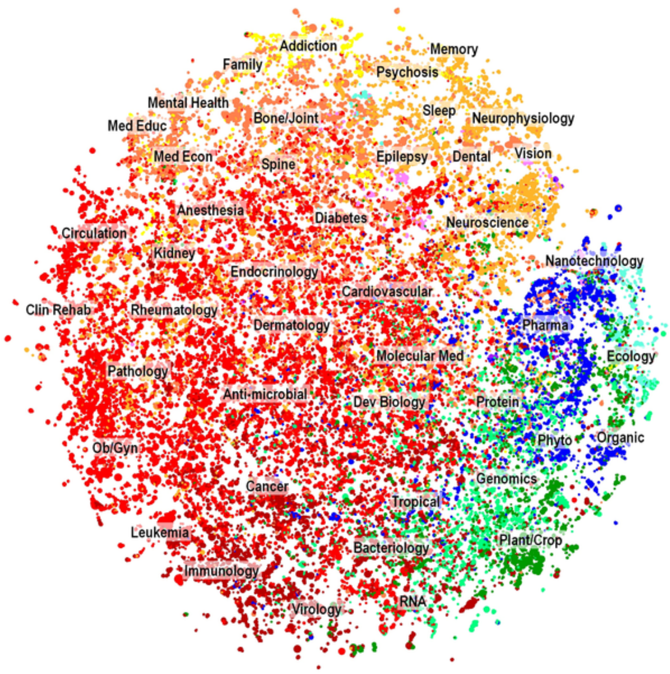


FIGURE 14.14. *DNA microarray data: average linkage hierarchical clustering has been applied independently to the rows (genes) and columns (samples), de-*

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



Clustering articles published in Science (Blei & Lafferty 2007)

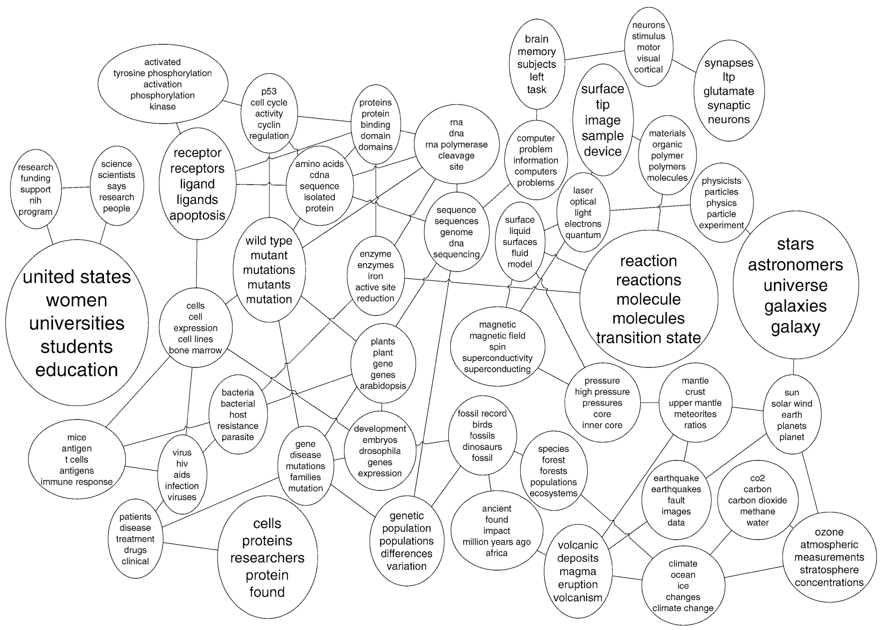


FIG. 2. *A portion of the topic graph learned from* 16,351 *OCR articles from Science* (1990–1999). *Each topic node is labeled with its five most probable phrases and has font proportional to its popularity in the corpus.* (*Phrases are found by permutation test.*) *The full model can be found in* *http://www.cs.cmu.edu/~lemur/science/*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

Examples for Generative Models

Generative Adversarial Networks (GANs)

see also *noise-contrastive estimation* and *predictability minimization*

"Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Network", ICLR 2016, https://arxiv.org/abs/1511.06434