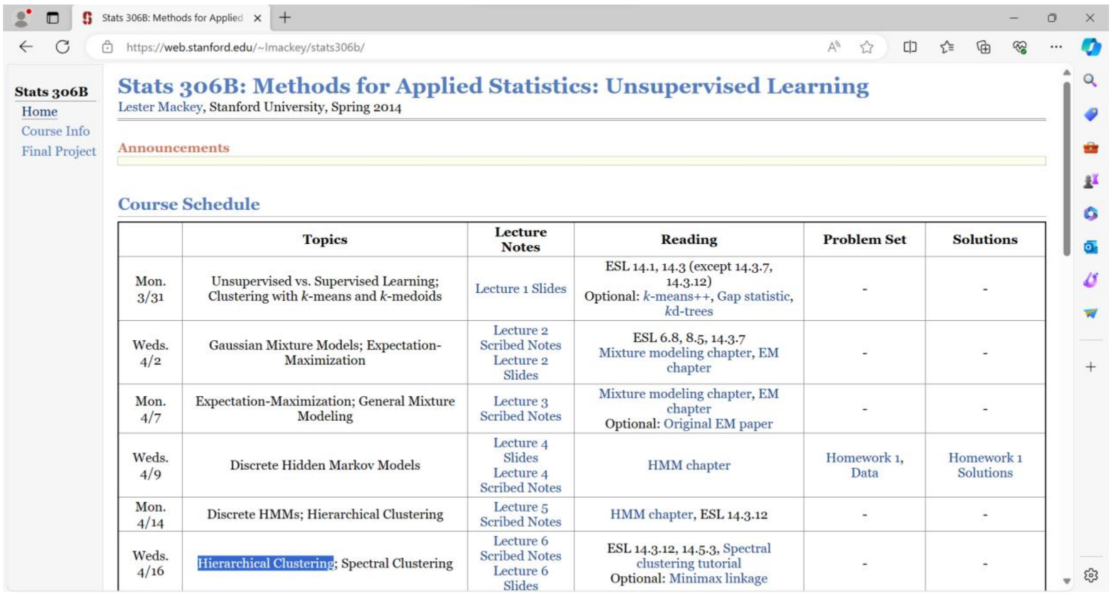
Stanford University

University Undergraduate Course: Unsupervised Learning

https://web.stanford.edu/~lmackey/stats306b/doc/stats306b-spring14-lecture6_scribed.pdf
https://web.stanford.edu/~lmackey/stats306b/
(retrieved 11 March 2024)



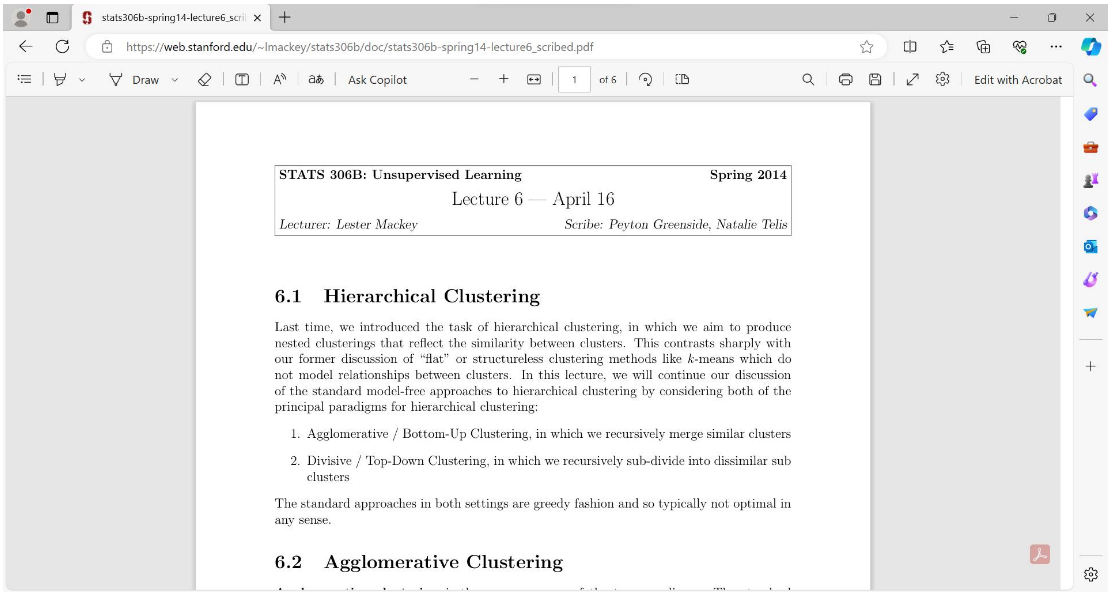
Stats 306B: Methods for Applied Statistics: Unsupervised Learning
Lester Mackey, Stanford University, Spring 2014

Stats 306B
Home
Course Info
Final Project

Announcements

Course Schedule

	Topics	Lecture Notes	Reading	Problem Set	Solutions
Mon. 3/31	Unsupervised vs. Supervised Learning; Clustering with k-means and k-medoids	Lecture 1 Slides	ESL 14.1, 14.3 (except 14.3.7, 14.3.12) Optional: k-means++, Gap statistic, kd-trees	-	-
Weds. 4/2	Gaussian Mixture Models; Expectation-Maximization	Lecture 2 Scribed Notes Lecture 2 Slides	ESL 6.8, 8.5, 14.3.7 Mixture modeling chapter, EM chapter	-	-
Mon. 4/7	Expectation-Maximization; General Mixture Modeling	Lecture 3 Scribed Notes	Mixture modeling chapter, EM chapter Optional: Original EM paper	-	-
Weds. 4/9	Discrete Hidden Markov Models	Lecture 4 Slides Lecture 4 Scribed Notes	HMM chapter	Homework 1, Data	Homework 1 Solutions
Mon. 4/14	Discrete HMMs; Hierarchical Clustering	Lecture 5 Scribed Notes	HMM chapter, ESL 14.3.12	-	-
Weds. 4/16	Hierarchical Clustering; Spectral Clustering	Lecture 6 Scribed Notes Lecture 6 Slides	ESL 14.3.12, 14.5.3, Spectral clustering tutorial Optional: Minimax linkage	-	-



STATS 306B: Unsupervised Learning — Spring 2014
Lecture 6 — April 16
Lecturer: Lester Mackey — *Scribe: Peyton Greenside, Natalie Telis*

6.1 Hierarchical Clustering

Last time, we introduced the task of hierarchical clustering, in which we aim to produce nested clusterings that reflect the similarity between clusters. This contrasts sharply with our former discussion of "flat" or structureless clustering methods like k-means which do not model relationships between clusters. In this lecture, we will continue our discussion of the standard model-free approaches to hierarchical clustering by considering both of the principal paradigms for hierarchical clustering:

1. Agglomerative / Bottom-Up Clustering, in which we recursively merge similar clusters
2. Divisive / Top-Down Clustering, in which we recursively sub-divide into dissimilar sub clusters

The standard approaches in both settings are greedy fashion and so typically not optimal in any sense.

6.2 Agglomerative Clustering

An Unusual Example: Clustering Great Paintings

After clustering 57 paintings by ratings for composition, color, drawing type, and expression, the authors discussed in slide 9 generated a dendrogram of great artwork labeled by painter. The authors note that the value of this analysis is dependent on the interesting hypotheses generated by the clustering, as in the example above.

6.4.2 Practicalities and Challenges

Model selection presents challenges similar to k-means cluster selection.

Model selection (choosing truncation level, referred to in 6.4.1, in **A Brief Note...**) is a challenge in creating meaningful interpretations of hierarchical clusterings. Although there is no single solution to interpretation, many of the methods we have discussed for k selection in k-means may apply equally well here.

Interpreting dendrograms is challenging for large datasets.

It is difficult to interpret dendrograms even past the selection of a model because large datasets may make visualizing clusters impossible, as well as choosing a meaningful k difficult (if there are several thousand data points, there may be tens or hundreds of meaningful clusters).

One solution is labeling each interior node with a prototype data point. One could simply choose the average or median of those in the cluster. A much more informative choice would be labeling each cluster with a prototypical datapoint which is minimally dissimilar from every point in the cluster. This is called **minimax linkage**: when aiming to label a cluster C, choose the prototype $x^* = \text{argmin}_{x \in C} \max_{x' \in C} d(x, x')$.