

CSE 251U/291: Unsupervised Learning Introduction

A: What can be done without labels?

Where machine learning shines

Supervised learning with:

- highly specific tasks
- large labeled data sets



Problems with this setting:

- ① Large labeled data sets are usually hard to come by.
- ② What about loosely related tasks?

Humans seem to learn well in primarily unsupervised mode.

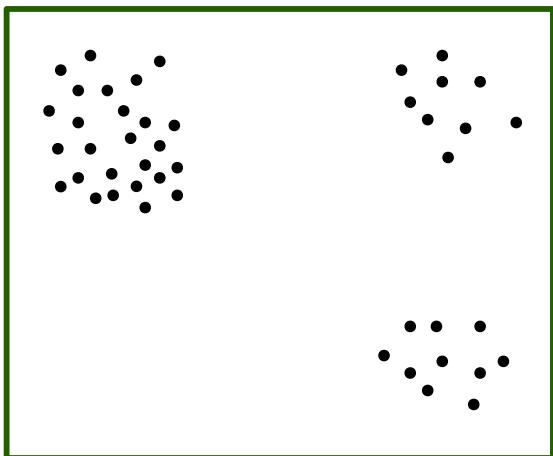
What can be done without labels?

1. REMEMBER



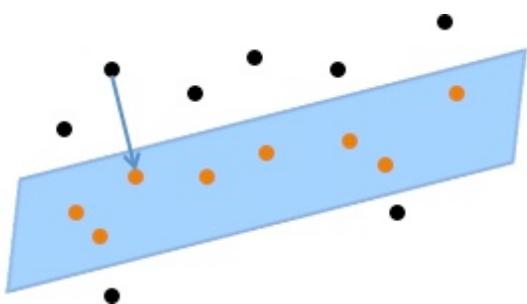
What can be done without labels?

2. SUMMARIZE



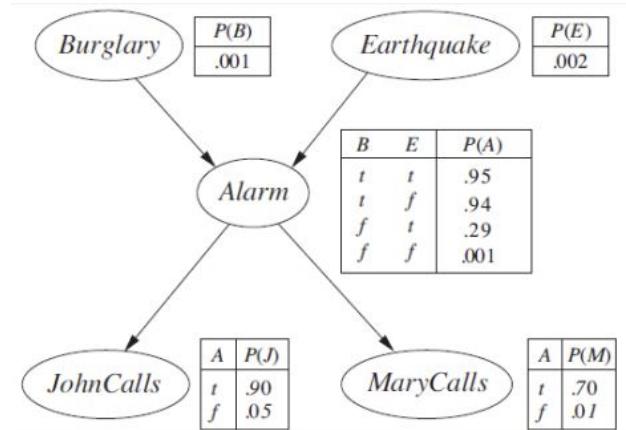
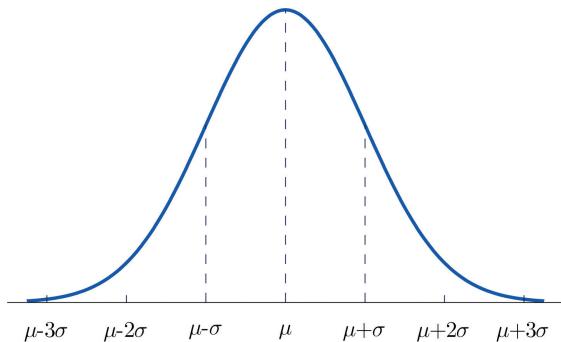
What can be done without labels?

3. COMPRESS



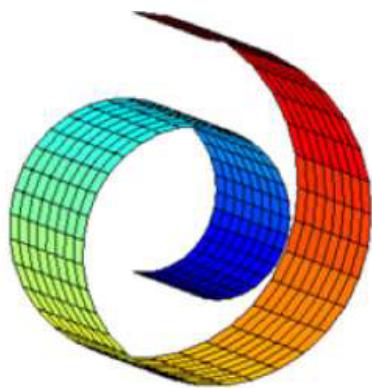
What can be done without labels?

4. GENERATE



What can be done without labels?

5. EMBED



Some approaches to unsupervised learning

- ① Remember
- ② Summarize
- ③ Compress
- ④ Generate
- ⑤ Embed

B: Course goals and structure

Course goals

An overview of the field of unsupervised learning.

- ① What, precisely, are the problems we're trying to solve?
- ② What are general principles for coming up with solutions?
- ③ What algorithms are commonly used?
- ④ What are the statistical properties of these procedures, and what are some of their failure modes?
- ⑤ What are typical use cases?

Part 1: Basic primitives

Fast, easy, and widely used.

- Nearest neighbor
- Mean, median, variance, covariance
- Histograms and other simple data sketches
- K-means and hierarchical clustering
- Principal component analysis
- Singular value decomposition
- Random projection

Part 2: Probabilistic models

More powerful and expressive, with a well-developed family of estimators.

- The basic distributions: exponential families
- Bayesian estimation
- Multivariate Gaussian
- Graphical models
- Sampling and variational inference
- Latent variable models
- Nonparametric Bayes and topic models

Part 3: More advanced structure

- Spectral embedding
- Sparse coding
- Projection pursuit
- Autoencoders
- Self-supervised learning

Helpful background

- Linear algebra
- Probability and statistics
- Algorithmic analysis
- Python

Logistics

Things to do:

- Download slides
- Attend lectures and fill in any gaps in the slides
- Work on assignments
- Office hours

Assessments:

- Weekly homeworks submitted via Gradescope