ONLINE MASTERS IN **DATA SCIENCE**

DSC 257R - UNSUPERVISED LEARNING

COURSE GOALS & STRUCTURE

SANJOY DASGUPTA, PROFESSOR

UC San Diego

COMPUTER SCIENCE & ENGINEERING

HALICIOĞLU DATA SCIENCE INSTITUTE

Course Goals

An overview of the field of unsupervised learning.

1. What, precisely, are the problems we're trying to solve?
2. What are the 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?

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

Assessments:

- Weekly homeworks submitted via Gradescope