CS 7300 : UNSUPERVISED FEATURE LEARNING

Semester Hours:	3.0	*Contact Hours:* 3
Coordinator:	TBD	
Text:	Various	
Author(s):	VARIOUS	
Year:	Various	

SPECIFIC COURSE INFORMATION

Catalog Description:

The course covers the principles and advances in unsupervised feature learning algorithms. It focuses on development of machine learning features, considering the feature hierarchies from unlabeled data. The learning algorithms are exploited in many applications. Topics include clustering, sparse coding, Boltzmann machine, autoencoders, and deep belief networks, and ethics. The course also requires an open-ended research project. Prerequisites: CS 5200 and STAT 5020, or permission of instructor. Credit cannot be received for both DATA 7300 and CS 7300.

Course type: **ELECTIVE**

SPECIFIC COURSE GOALS

- I am able to explain the principles of the unsupervised feature learning algorithms.
- I am able to implement popular clustering algorithms.
- I am able to design unsupervised learning networks from unlabeled data.
- I am able to employ unsupervised learning networks for real applications.
- I am able to analyze ethical issues associated with unsupervised learning.

LIST OF TOPICS COVERED

- Introduction (~5%)
    - Overview
    - Basic mathematics review: probability, loss function, likelihood, regressions
- Supervised Learning & Algorithms (~10%)
    - Classification: linear models

    - Kernel method: support vector machines
    - Neural networks: back-propagation
- Clustering (~15%)
    - Hierarchical clustering
    - K-means, Bradley-Fayyad-Reina (BFR) algorithm
    - Expectation-maximization algorithm
    - Density-based spatial clustering of applications with noise
- Principle Component Analysis (PCA) (~15%)
    - Maximum variance formulation, minimum-error formulation
    - Probabilistic PCA
    - Kernel PCA
    - Nonlinear latent variable models*
- Independent Component Analysis (ICA) (~15%)
    - Models: linear noiseless ICA, linear noisy ICA, nonlinear ICA
    - Binary ICA
- Sparse Coding (~10%)
    - Over-complete set of basis vector
    - Efficiency
- Deep Network & Learning (~15%)
    - Neural network
    - Deep belief networks, deep architecture
    - Convolutional deep belief networks
    - Boltzmann Machine
    - Autoencoders
- Ethics of Machine Learning (~15%)
    - Introduction – What and Why?
    - Data bias and fairness
    - Privacy and security
    - Choose from: model interpretability, model accountability, or adversarial uses
    - Case study

(* Optional if time allows.)

RECOMMENDED REFERENCES

- Unsupervised Learning Algorithms, by Celebi and Aydin, Springer, 2016.
- Unsupervised Classification: Similarity Measures, Classical and Metaheuristic Approaches, and Applications, by Bandyopadhyay and Saha, Springer, 2013.