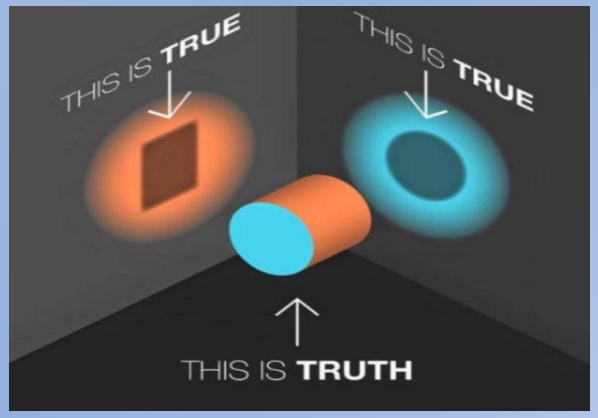
Information Theory & Machine Learning

Two Sides of the Same Coin

Information Theory
Theoretical understanding
Performance benchmarks
Guides intuition



Machine Learning
Algorithmic understanding
Computational aspects
Training & Learning

Information Theory & Coding (ECSE 6530) -- Spring 2021: The course will have three main components:

- Foundations of information theory
- Applications of information theory to machine learning
- Applications of information theory to high-dimensional statistics

Instructor: Ali Tajer (ECSE)
Prerequisite: Engineering Probability (ECSE-2500) or its equivalent.
Day/time: TF 12:50– 2:10pm

ECSE 6530 - Information Theory

Part I - Information Measures & Applications in ML

1- Information & Entropy

1. Entropy: axiomatic characterization
2. Properties of entropy
3. Entropy & submodularity
4. Relative entropy
5. Differential entropy
6. Properties of differential entropy

2 - Information Geometry

1. Divergence
2. Conditional divergence
3. Properties of conditional divergence
4. Mutual information
5. Properties of mutual information
6. Conditional mutual information
7. Properties of conditional mutual information
8. Geometric interpretation of mutual information
9. Variational characterization of mutual information
10. Sufficient statistics

3 - Applications of Information Measures in Machine Learning

1. Sensor placement
2. Clustering
3. Estimating entropy
4. Estimation mutual information
5. Feature selection via information gain scores
6. Independence testing
7. Machine learning on distributions
8. Structure learning in graphical models
9. Maximum entropy density estimation
10. Information projection

Part II - Data Compression & Applications in ML

4- Source Coding & Applications in ML

1. Basics of course coding
2. Source coding theorem
3. Kraft and McMillan theorems
4. Huffman codes
5. Prefix codes
6. Minimum description length principle
7. Rate distortion theory

5 - Applications of Source Coding in Machine Learning

1. Empirical risk minimization
2. Histogram classifiers
3. Decision tree classifiers
4. Unbounded loss functions and infinite classes
5. Histogram regression
6. Universal prediction
7. Information bottleneck

Part III - Information Theory & High-dimensional Statistics

6 - Statistical Decision Theory

1. Parametric data models
2. Non-parametric data models
3. Best estimator
4. Risk functions (Bayes and minimax)
5. Relationship between Bayes and minimax risks
6. Tensoring experiments
7. Sample complexity information

7 - Hypothesis testing

1. Binary hypothesis testing
2. Neyman-Pearson formulation
3. Likelihood ratio tests
4. Achievable and converse bounds
5. Asymptotes
6. Information projection and large deviation

8 - Estimation

1. f-divergences
2. Total variations
3. Inequality between f-divergences
4. Variational representation
5. Fisher information
6. Alternative forms to Fisher information
7. Hammersley-Chapman-Robbins (HCR) lower bound
8. Cramér-Rao (CR) lower bound
9. Bayesian CR lower bound

9 - Unstructured Estimation

1. Exact minimax risk for Gaussian location model
2. Le Cam's method
3. Two-point method
4. Mutual information method
5. Fano's method
6. Density estimation