# Lahore University of Management Sciences

# EE 578/CS594 – Information Theory and Machine Learning

Spring 2022

| | |
|---|---|
| Instructor(s) | Momin Uppal and Muhammad Tahir |
| Room No. | 9-246A |
| Office Hours | TBA |
| Email | momin.uppal@lums.edu.pk; tahir@lums.edu.pk |
| Telephone | 8112, 8423 |
| Secretary/TA | TBA |
| TA Office Hours | TBA |
| Course URL (if any) | https://lms.lums.edu.pk/ |

| Course Basics | | | | |
|---|---|---|---|---|
| Credit Hours | 3 | | | |
| Lecture(s) | Nbr of Lec(s) Per Week | 2 | Duration | 75 minutes |
| Recitation/Lab (per week) | Nbr of Lec(s) Per Week | - | Duration | - |

| Course Distribution | |
|---|---|
| Core | -- |
| Elective | Graduate-level Elective for EE and CS Majors |
| Open for Student Category | SSE Seniors / Graduate students |
| Close for Student Category | Anyone not fulfilling the required pre-requisite |

## COURSE DESCRIPTION

Given a noisy observation of a data sequence, what is the maximum amount of information one can extract from it? If one can withstand some level of distortion in the extracted information, what is the maximum amount of extractable information then? These are some of the questions that the field of information theory tries to address. Interestingly, these questions are intimately linked to the fields of inference and (machine) learning. Indeed, inference and learning methods attempt to extract patterns /decisions/information from observed data. **The objective of this course is to explore these links by providing an information theoretic perspective to inference and learning.**

After providing a crash course in basic concepts in probability and linear algebra, the course contents will delve into fundamental concepts in statistical inference and learning, with special emphasis on their connections with information theory. In the process, students will also be exposed to information theoretical concepts related to data representation and communication. Special emphasis will be laid on providing a Bayesian perspective to machine learning algorithms.

## COURSE PREREQUISITE(S)

- Enforced: Math-230 Probability
- Recommended: Background in Basics of Linear Algebra

## Grading Breakup and Policy (Tentative)

Homework/Computer Assignments : 20% (5)
Quizzes: 15%
Midterm Examination:30%
Final Examination:35%

# Lahore University of Management Sciences

| COURSE OVERVIEW (TENTATIVE) | | |
|---|---|---|
| **Week** | **Topics** | **Remarks** |
| 1 | • Course Introduction: Connections between Machine Learning and Information Theory<br>• Probability and Random Variables Primer \ Introduction to Bayesian Inference<br>• Introduction to Machine Learning Models: Discriminative vs Generative Models | |
| 2 | • Vector Spaces, Theory of Matrices, Projection Operators, Generalized Inverse Theory<br>• Solving Ax = y<br>• Sparse Solution of Ax = y, Fundamental bounds on sparsest solutions | |
| 3 | • Matrix Decompositions and their role in Machine Learning | |
| 4-5 | • Introduction to Information Theory<br>➢ Shannon's Information content: Entropy<br>➢ Asymptotic equi-partition property<br>➢ Lossless data compression / Source coding theorem<br>➢ Conditional entropy and mutual information, Cross-Entropy as An Objective Function of Multi-class Classification<br>➢ Relative Entropy / KL Divergence<br>➢ Data over noisy channels<br>Mutual Information: Correlation of the 21st century | |
| 6 | • Learning Theory: Estimation problem and fundamental bounds, Empirical Risk Minimization, Bias versus variance trade-off | |
| 7 | • Vector Spaces Approach to First Linear Model in Supervised Learning:<br>Least-Square Regression, Regularization, Hessian and its importance to Model Overfitting | |
| 8 | • Vector Spaces Approach to First Linear Model in Unsupervised Learning:<br>PCA, RPCA and Matrix Decompositions | |
| 9 | • A probabilistic approach to First Linear Model in Supervised Learning<br>➢ Naïve Bayes<br>➢ Maximum Likelihood Model fitting<br>➢ The Bayesian approach to model fitting and its relationship to regularization<br>➢ Feature selection using information theoretic measures | |
| 10-11 | ➢ Bayesian Inference and Non-conjugate Models<br>➢ Point Estimates – The MAP Solution<br>➢ The Laplace Approximation<br>➢ Sampling Techniques: Metropolis-Hasting<br>• Gaussian Models<br>➢ Multi-variate Gaussian distribution<br>➢ Maximum likelihood parameter estimation<br>➢ Gaussian discriminant analysis (GDA)<br>➢ Linear discriminant analysis (LDA)<br>➢ Quadratic discriminant analysis (QDA)<br>➢ LDA and Logistic regression | |
| 11-12 | • A probabilistic approach to First Linear Model in Unsupervised Learning<br>➢ A probabilistic approach to PCA<br>➢ Clustering: The K-means clustering algorithm<br>➢ Mixture Models<br>○ The expectation maximization algorithm<br>○ Model selection for latent variable models<br>➢ Fitting models with missing data | |
| 13-14 | • Deep Neural Networks: an Information Theoretical perspective<br>➢ Information Characteristics of DNN layers<br>➢ Opening the black box of DNNs via Information and the benefit of multiple Hidden Layers and more training samples | |

# Lahore University of Management Sciences

| Examination Detail (Tentative) | |
|---|---|
| Midterm Exam | Yes/No: Yes<br>Combine Separate: Combined<br>Duration: 2 hours<br>Preferred Date: During the Midweek<br>Exam Specifications: |
| Final Exam | Yes/No: Yes<br>Combine Separate: Combined<br>Duration: 3 hours<br>Exam Specifications: |

| Textbook(s)/Supplementary Readings |
|---|
| We may not follow a single textbook. Most of the material will be derived from the following.<br><br>1. Information Theory, Inference, and Learning Algorithms by David MacKay<br>2. Machine Learning: A probabilistic perspective by Kevin Murphy<br>3. Neural Networks and Learning Machines by Simon Haykin<br>4. Elements of Statistical Learning by Trevor Hastie<br><br>The instructors will also provide their own course notes for ready reference. |