Mathematics of Machine Learning and Signal Recognition
COMS E4995

Instructor:

Prof. Homayoon Beigi <beigi@recotechnologies.com> (hb87@columbia.edu)

Textbooks:

Required:
H. Beigi, "Fundamentals of Speaker Recognition, Springer, New York, 2011.

Reference Books:
K.P. Murphy, "Machine Learning, A Probabilistic Perspective," The MIT Press, Cambridge, MA, 2012.
H. Beigi, "Fundamentals of Speaker Recognition," Springer, New York, 2nd Edition, 2022.
M. Loève, "*Probability Theory*," Springer, New York, 4th Edition, 1977.
P.R. Halmos, "Measure Theory," Springer, New York, 1974.
I.T. Jolliffe, "Principal Component Analysis," Springer, New York, 2nd Edition, 2002.
R. Courant and D. Hilbert, "Methods of Mathematical Physics," John Wiley & Sons, New York, 1989.
C. F. Gerald and P. O. Wheatley, "Applied Numerical Analysis," Pearson College Div., 7th Edition, New York, 2003.
G.J. McLachlen and T. Krishnan, "The EM Algorithm and Extensions," John Wiley & Sons, 2nd Edition, New York, 2008.
W.E. Boyce and R.C. DiPrima, "Elementary Differential Equations and Boundary Value Problems," John Wiley & Sons, 11th Edition, New York, 2017.
P.W. Berg and J.L. McGregor, Holden Day, San Francisco, 1966.
R. Fletcher, "Practical Methods of Optimization," John Wiley & Sons, 2nd Edition, New York, 2000.

Grading:

Homework (20%):
- Problems and coding assignments.

Midterm (20%):
- Coding assignment and Problems.

Project Proposal (10%):
- 2-page proposal, including state of the art and proposed methodology.

Final Project (50%):
35% - Test/report of the methodology and results.

15% - Code and Results.

Course Description:

Mathematics of Machine Learning and Signal Recognition provides the background mathematical background for addressing in-depth problems in machine learning, as well as the treatment of signals, especially time-dependent signals, specifically non-stationary time-dependent signals – although spatial signals such as images are also considered. The course will provides the essentials of several mathematical disciplines which are used in the formulation and solution of the problems in the above fields. These disciplines include Linear Algebra and Numerical Methods, Complex Variable Theory, Measure and Probability Theory (as well as statistics), Information Theory, Metrics and Divergences, Linear Ordinary and Separable Partial Differential Equations of Interest, Integral Transforms, Decision Theory, Transformations, Nonlinear Optimization Theory, and Neural Network Learning Theory. The requirements are Advanced Calculus and Linear Algebra. Knowledge of Differential Equations would be helpful.

Lectures:

Week 1

- Linear Algebra and Numerical Methods

  Basic Definitions
  Norms
  Gram-Schmidt Orthogonalization
  Ordinary Gram-Schmidt Orthogonalization
  Modified Gram-Schmidt Orthogonalization
  Sherman-Morrison Inversion Formula
  Vector Representation under a Set of Normal Conjugate Direction
  Stochastic Matrix
  Linear Equations

Week 2

- Complex Variable Theory

  Complex Variables
  Limits
  Continuity and Forms of Discontinuity
  Convexity and Concavity of Functions
  Odd, Even and Periodic Functions
  Differentiation
  Analyticity
  Integration
  Power Series Expansion of Functions
  Residues
  Relations Between Functions
  - Convolution
  - Correlation
  - Orthogonality of Functions

Weeks 3 & 4

- Measure and Probability Theory and Statistics

  Set Theory
  - Equivalence and Partitions
  - R-Rough Sets (Rough Sets)
  - Fuzzy Sets

  Measure Theory
  - Measure
  - Multiple Dimensional Spaces
  - Metric Space
  - Banach Space (Normed Vector Space)
  - Inner Product Space (Dot Product Space)

  Infinite Dimensional Spaces (Pre-Hilbert and Hilbert)
  Probability Measure

Integration
Functions
Probability Density Function
Densities in the Cartesian Product Space
Cumulative Distribution Function
Function Spaces
Transformations

Statistical Moments
Discrete Random Variables
Combinations of Random Variables
Convergence of a Sequence
Sufficient Statistics
Moment Estimation
Estimating the Mean
Law of Large Numbers (LLN)
Different Types of Mean
Estimating the Variance
Multi-Variate Normal Distribution

Weeks 5

- Information Theory

Sources
The Relation between Uncertainty and Choice
Discrete Sources
Entropy or Uncertainty
Generalized Entropy
Information
The Relation between Information and Entropy
Discrete Channels
Continuous Sources
Differential Entropy (Continuous Entropy)
Relative Entropy
Mutual Information
Fisher Information

Weeks 6

- Metrics and Divergences

Distance (Metric)
Distance Between Sequences
Distance Between Vectors and Sets of Vectors
Hellinger Distance

Divergences and Directed Divergences
Kullback-Leibler's Directed Divergence
Jeffreys' Divergence
Bhattacharyya Divergence
Matsushita Divergence
F-Divergence
δ -Divergence
χ α Directed Divergence

Weeks 7 and 8

- Review of Linear Differential Equations (Ordniary and Separable Partial)

- Integral Transforms

Integral Equations
Kernel Functions
Hilbert's Expansion Theorem
Eigenvalues and Eigenfunctions of the Kernel
Fourier Series Expansion
Convergence of the Fourier Series

Parseval's Theorem
Wavelet Series Expansion
The Laplace Transform
Inversion
Some Useful Transforms
Complex Fourier Transform (Fourier Integral Transform)
Translation
Scaling
Symmetry Table
Time and Complex Scaling and Shifting
Convolution
Correlation
Parseval's Theorem
Power Spectral Density
One-Sided Power Spectral Density
PSD-per-unit-time
Wiener-Khintchine Theorem
Discrete Fourier Transform (DFT)
Inverse Discrete Fourier Transform (IDFT)
Periodicity
Plancherel and Parseval's Theorem
Power Spectral Density (PSD) Estimation
Fast Fourier Transform (FFT)
Discrete-Time Fourier Transform (DTFT)
Power Spectral Density (PSD) Estimation
Complex Short-Time Fourier Transform (STFT)
Discrete-Time Short-Time Fourier Transform DTSTFT
Discrete Short-Time Fourier Transform DSTFT
Discrete Cosine Transform (DCT)
Efficient DCT Computation

Week 9

- Difference Equations and The z-Transform
  Difference Equations
  z-Transform – Definition
  Translation
  Scaling
  Shifting – Time Lag
  Shifting – Time Lead
  Complex Translation
  Initial Value Theorem
  Final Value Theorem
  Real Convolution Theorem
  Inversion
- Cepstrum

Week 10
- Decision Theory
  Hypothesis Testing
  Bayesian Decision Theory
  Bayesian Classifier
  Decision Trees

- Unsupervised Clustering and Learning

Vector Quantization (VQ)
Basic Clustering Techniques
Estimation using Incomplete Data

- Parameter Estimation
  Maximum Likelihood Estimation (MLE, MLLR, fMLLR)
  Maximum A-Posteriori (MAP) Estimation
  Maximum Entropy Estimation
  Minimum Relative Entropy Estimation
  Maximum Mutual Information Estimation (MMIE)
  Model Selection (AIC and BIC)

Week 11

- Transformation
  Principal Component Analysis (PCA)
  Linear Discriminant Analysis (LDA)
  Factor Analysis (FA)
  Probabilistic Linear Discriminant Analysis (PLDA)

- Hidden Markov Modeling (HMM)
  Memoryless Models
  Discrete Markov Chains
  Markov Models
  Hidden Markov Models
  Model Design and States
  Training and Decoding
  Gaussian Mixture Models (GMM)
  Practical Issues

Week 12

- Nonlinear Optimization Theory
  Gradient-Based Optimization
    The Steepest Descent Technique
    Newton's Minimization Technique
    Quasi-Newton or Large Step Gradient Techniques
    Conjugate Gradient Methods
  Gradient-Free Optimization
  Search Methods
  Gradient-Free Conjugate Direction Methods
  The Line Search Sub-Problem
  Practical Considerations
    Large-Scale Optimization
    Numerical Stability
    Nonsmooth Optimization
    Constrained Optimization
      The Lagrangian and Lagrange Multipliers
      Duality
  Global Convergence

Week 13
- Neural Network Learning
  Perceptron
  Feedforward Networks
  Time-Delay Neural Networks (TDNN)
  Convolutional Neural Networks (CNN)

Recurrent Neural Networks (RNN)
Long-Short Term Memory Networks (LSTM)
End-to-End Sequence (Encoder/Decoder) Neural Networks
Embeddings and Transfer Learning