Machine Learning Fall 2020

1 Machine Learning Description

Machine learning is a branch of artificial intelligence, concerned with the construction and study of systems that can learn from data. Data may be numeric or symbolic and typically has the form of an N-tuple. The anthropomorphic term learning in the machine learning context means being able to predict some unobserved components of an N-tuple given some observed components of the N-tuple. This course provides a detailed explanation of many of the techniques used in machine learning and statistical pattern recognition.

The prerequisites for the course are

- A recent Linear Algebra course or knowledge of Linear Algebra at the graduate level
- A recent Probability and Statistics course or knowledge of Probability and Statistics
- A recent Algorithms course or knowledge of algorithms
- A fluency in programming in at least one of C++, Python, or Julia

The text book for the class is *The Elements of Statistical Learning, Hastie, Tibshirani, and Friedman, Springer* Although this semester I will not be teaching from the text book. But the text book will have in it material related to much of what is going to be taught.

The website for the course can be found at

http://haralick.org/ML/lecture_slides.shtml

The website will have some reference text pdfs and some reference paper pdfs.

I will try to have the lecture slides posted over the weekend before each class. The latest the lecture slides will be posted for the class is on Monday 11AM. I would recommend that you print out the pdf file before the class so that you have it in front of you with some blank paper on which you can write notes or questions. Of course the slides will be shared in the zoom session.

The class zoom session will be opened ten minutes before the class. I have not set up a waiting room so you can join when you arrive. I have never taught by zoom before and do not know how the class interaction will work. For now I suggest that if you have a question, unmute yourself and pose the question if nobody else is speaking. I promise you that I will do the best I can to answer every question.

After we finish at 4PM, I will leave the zoom session going so that anyone who would like to talk with me and ask questions can stay in the zoom session and I forward to being available for one on one interaction.

2 Topic List

All on the list of topics will not be covered and there may be somethings we do cover, but are not on the list. We will not cover topics exactly in the order of the list.

- Bayesian Classification
  - Class conditional probabilities
  - Prior Probabilities
  - Gain Matrix
  - Maximizing Expected Gain
  - Minimax Classification
- Principal Component Analysis
- Subspace Classifiers
- Probability Models
  - Parametric Probability Models
  - Non-parametric Probability Models
  - Making Decisions in Context
  - Conditional Independence
  - Hidden Markov Model
  - Forward Backward Algorithm

- Graphical Models
  - Semi-graphoids
  - Graphoids
  - Bayesian Nets
- Decision Trees
- Nearest Neighbor Classifiers
- Linear Regression
- Logistic Regression
- Neural Networks
  - The Perceptron Algorithm
  - The Back Propagation Algorithm
  - Deep Learning Networks
- Linear Decision Rules
  - Fisher Linear Decision Rule
  - Support Vector Machines
  - Kernel Methods
- Clustering
  - K-Means Clustering
  - Expectation Maximization
  - Linear Manifold Clustering
  - Gaussian Mixture Models
  - Clustering Evaluation Measures
- Experimental Protocols
  - Training Sets
  - Test Sets
  - Cross-Validation
  - Performance Characterization

3 Learning Goals

For this class, the student must be able to demonstrate a working knowledge of the theoretical foundations and software of machine learning represented by the topics of

- Bayesian Classification
- Principal Components
- Subspace Classifiers
- Support Vector Machines
- The Kernel Trick
- Regression
- Clustering
- Performance Characterization

4 Assessment

The software and theoretical aspects will be assessed by a midterm project and report . Project reports will be due every two weeks beginning September 21, 2020. Midterm project due October 19, 2020. Final Project and report due December 21,2020. Midterm project 35%); Final project (55%); Attendance (10%).

The last two slides on the Discrete Bayes topic slides will explain all the contents that the Mid Term Project will have to have. It is due on October 12, 2020.

The format and sections in the Final Project Report can be found at

http://haralick.org/ML/machine_learning_syllabus_fall_2020.pdf