# Syllabus for MA598-550 Machine Learning and Uncertainty Quantification for Data Science

MA59800-550, Fall 2016

Office: MATH 410, 150 N. University Street, West Lafayette, IN, 47907 Office hours: TR 04:20-05:00pm (tentative) or by appointment Email: guanglin [at] purdue.edu

\_\_\_\_\_

# **Lectures Time and Location**

TR 03:00-04:15pm Classroom: UNIV 103

## **Graduate Course Description**

This introductory course will cover many concepts, models, algorithms and Python codes in machine learning and uncertainty quantification in data science. Topics include classical supervised learning (e.g., regression and classification), unsupervised learning (e.g., principle component analysis and K-means), uncertainty quantification algorithms (e.g., importance sampling, Markov Chain Monte Carlo) and recent development in the machine learning and uncertainty quantification field such as deep machine learning, variational Bayes, and Gaussian processes. While this course will give students the basic ideas, intuition and hands on practice behind modern machine learning and uncertainty quantification methods, the underlying theme in the course is probabilistic inference for data science.

## **Tentative Topics**

- Review on basic concepts in information theory and probability distributions
- Unsupervised machine learning algorithm
  - Principal component analysis
- Supervised machine learning algorithm

- Active subspace algorithm
- Sliced inverse regression
- o Localized sliced inverse regression
- Dimension reduction algorithms & data compression
- Compressive sensing algorithm
- Linear regression and classification
- Bayesian inference
- Clustering analysis (K-means Clustering, mixture models and Expectation Maximization)
- Stochastic gradient descent algorithm
- Random forest
- Hidden Markov models
- Deterministic approximate inference: Variational Bayes, and expectation propagation
- Support Vector Machines regression and classification
- Gaussian process regression, Gaussian process classification
- Uncertainty quantification algorithms
  - Monte Carlo,
  - Latin Hyper Cube sampling,
  - Importance sampling
  - Polynomial chaos method
  - Gaussian process regression
  - Compressive sensing
- Data assimilation (Kalman filter, particle filter)

#### ~ • •

#### Prerequisites

Basic linear algebra, calculus, and probability, or permission of instructor.

\_\_\_\_\_

### Textbooks

Pattern Recognition and Machine Learning, Christopher M. Bishop, 2007

Gaussian processes for machine learning, Carl Edward Rasmussen and Christopher K. I. Williams, 2005

# Lecture Notes in Python Interactive Jupyter Notebook: https://github.com/PredictiveModelingMachineLearningLab /MA598

### Assignments

- Homework (links will be activated as homework is assigned). Copying will not be tolerated.
- Homework will be submitted through blackboard.
- Review of recent research Students will choose a subtopic of machine learning research, select three recent conference papers on the topic, and write a 2 page report outlining the main ideas of papers and relate them to the context of the course. And give a presentation on the main ideas during the class.
- Final project: You are required to complete a class project. The choice of the topic is up to you so long as it clearly pertains to the course material. To ensure that you are on the right track, you will have to submit a one paragraph description of your project a month before the project is due. You are encouraged to collaborate on the project, but we expect a four-page write-up about the project, which should clearly describe the project goal, methods, and your results. Each group should submit only one copy of the write-up and include all the names of the group members (a two-person group will have 6 pages, a three-person group will have 8 pages, and so on).

\_\_\_\_\_

### Grading

- 10% Participation
- 15% Review of recent research
- 40% Homework
- 35% Final Project