EE 445 Introduction to Machine Learning

Credits: 3
Categorization of credits: engineering topic

Instructor(s): Professors Narayana Santhanam

Catalog Description: EE 445 Introduction to Machine Learning (3) This course covers the basics of machine learning, including the statistical underpinnings of linear regression, classification techniques, resampling methods including cross validation, jackknife and bootstrap, shrinkage methods, singular value decomposition of matrices and applications to principal components analysis, decision trees, boosting and random forests. Pre: 342; or consent. Special note: This course has been offered as EE 491B for the last two years, and will be incorporated as a permanent course.
Credits: 3
Prerequisites: EE 342 "Probability and Statistics".

Class/Lab Schedule: 3 lecture hours per week in Spring Semester

Topics Covered:

The following plan is for Spring semesters (15 weeks). The total planned duration is 14 weeks to account for 3 holidays (MLK day, Presidents day and Good Friday).

1. Review of EE342: Random variables, independence (1 week)
2. Hypothesis testing: null vs linear dependence (1 week)
3. Linear regression: single and multiple features, Intro to t-statistic, the t-test, quadratic forms and Cochran's theorem, F-tests (incremental and direct) (2.5 weeks)
4. Classification (overview, perceptron, logistic*, LDA, QDA) (2.5 weeks)
5. Resampling techniques (Cross validation, jackknife and bootstrap) (2 weeks)
6. Shrinkage (LASSO, Ridge penalties as well as their interpretations in sparsity and computational stability respectively) (1 week)
7. Singular value decomposition (detailed) and Principal components analysis (2 weeks)
8. Decision trees, Pruning and random forests, brief introduction to boosting (2 weeks)

Textbook and Other Required Materials: "An introduction to statistical learning with applications in R" by James, Witten, Hastie and Tibshirani (Springer).

Course Objectives and Relationship to Program Objectives:

This course has multiple objectives. The first is for the students to get a grasp on a fundamentals in regression and classification, and basic primitives such as resampling techniques, singular value decomposition, thus providing a strong base for future graduate study and/or a career that

involves statistical analysis. Second, students are taught how to think about data in various domains (including but not limited to biology, engineering problems, statistical inference) using the fundamentals and primitives learnt above, including how to approach cases where the abstractions taught in class may not exactly apply. The final objective is to appreciate how to combine various methods learnt in class with domain knowledge for effective solutions. As a result EE442 addresses Program objectives [1,2,3 and 4].

Course Outcomes and Their Relationship to Program Outcomes
- Understanding the statistical and computer science primitives of machine learning [1]
- Using theory learnt in class to model a real life challenge (particular challenge varies from year to year---in Spring 2015, the project was to predict cancer survival rates using clinical data) [1, 2, 3, 6]
- Understanding generally how prior fundamentals from earlier courses are used, and how fundamentals taught here translate to problems beyond the course [4, 7]

Contribution of Course to Meeting the Professional Component
"Engineering topics: 100%"

Computer Usage:
Students have a concurrent project that spans more than half the semester. This project will involve extensive use of design tools in R and MATLAB as well as substantial coding in R or any another language the student may choose (python/C).

Design Credits and Features:
EE445 will have 0.5 design credit. As mentioned above, in addition to homeworks, students do an extended project involving real data. In Spring 2015, students predicted cancer survival rates using clinical data made available through the Dream Challenge (sponsored by Sage Bionetworks).

Person(s) Preparing Syllabus and Date: Narayana Santhanam, May 15, 2015. Y. Dong, June 14, 2021.