

MATH/STAT 4450/8456: Introduction to Machine Learning and Data Mining

TR 4:00 PM – 5:15 PM | Xiaoyue Cheng

This course focuses on statistical machine learning techniques in classification problems, as well as aspects of model development. It will include the following topics:

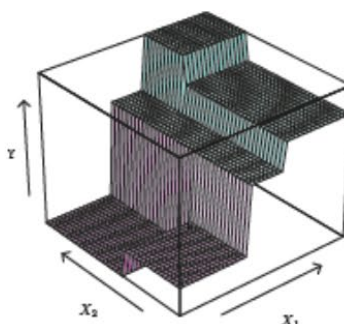
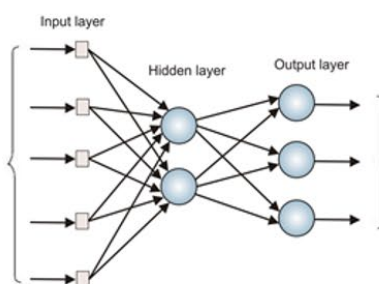
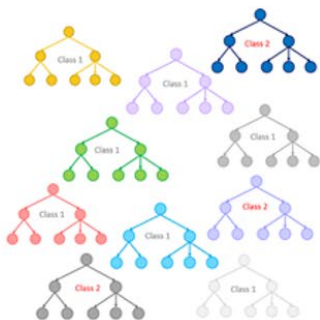
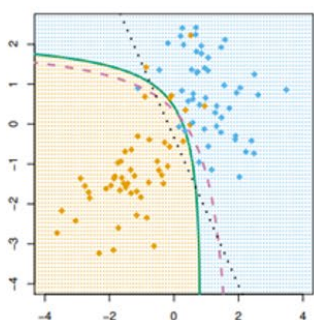
1. Introduction to statistical learning: goals, model accuracy, workflow
2. Understanding and predicting predictor performance: C_p , AIC, BIC, R^2 , ROC
3. Resampling methods for error estimation: cross-validation and bootstrap
4. Linear methods for classification: logistic regression
5. Trees and ensemble methods: decision tree, c5.0, bagging, random forest, adaboost, gradient boosting machine, xgboost
6. Non-parametric methods for classification: nearest neighbors, support vector machines
7. Bayesian methods for classification: naive Bayes, linear discriminant analysis, quadratic discriminant analysis
8. Other methods for classification: neural networks, deep learning, convolutional neural networks

Pre-requisites: MATH 4740/8746 or STAT 3800/8805, or permission of instructor.

Textbooks: *An Introduction to Statistical Learning*, by James, Witten, Hastie, and Tibshirani.

Software: R or Python.

Projects: Three real-data machine learning competitions will be held via Kaggle InClass.



For More Information:

Xiaoyue Cheng | 402.554.2848 | xycheng@unomaha.edu