

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

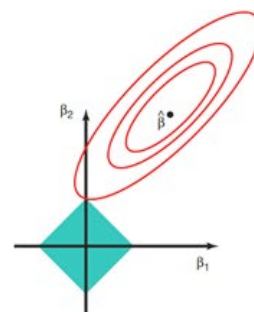
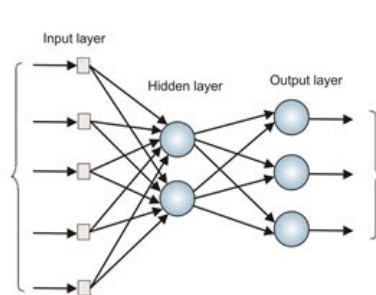
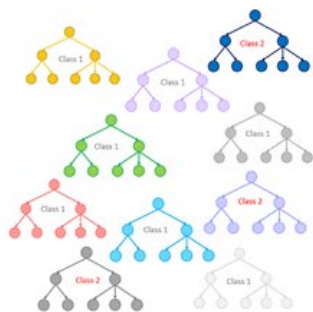
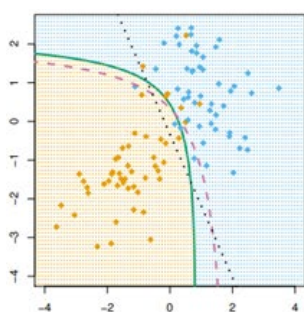
MW 4:00 PM – 5:15 PM | Online | Xiaoyue Cheng

This course will include the following topics:

1. Introduction to statistical learning: goals, model Accuracy
2. Understanding and predicting predictor performance: Cp, AIC, BIC, and  $R^2$
3. Resampling methods for error estimation: cross-validation and bootstrap
4. Shrinkage methods for multiple regression: ridge regression and Lasso
5. Linear methods for classification: logistic regression
6. Non-parametric methods for classification: nearest neighbors, support vector machines
7. Trees and ensemble methods: decision tree, C50, bagging, random forest, Adaboost, gradient boosting machine, xgboost
8. Bayesian methods for classification: Naïve Bayes, linear discriminant analysis, quadratic discriminant analysis
9. Other methods for classification: neural networks, deep learning
10. Model selection and dimension reduction: best subset and stepwise selection, principal components regression, partial least squares
11. Non-linear parametric regression methods: polynomial regression, regression splines, smoothing splines, local regression, generalized additive models

Three real-data machine learning competitions will be held in class.

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



## For More Information:

Xiaoyue Cheng | 402.554.2848 | [xycheng@unomaha.edu](mailto:xycheng@unomaha.edu)