Academic Calendar Entry: Regression, classification, resampling, model selection and validation, fundamental properties of matrices, dimension reduction, tree-based methods, unsupervised learning. Prerequisite: Either (a) STAT 230 or (b) a score more than 75% in one of APSC 254, BIOL 202, PSYO 373; and one of COSC 111, APSC 177. Course Overview: This course will introduce students to some popular machine learning techniques for making sense of complex datasets. Students will be applying these methods to datasets using R. Learning Outcomes: At the end of this course, you should be able to: build a model and validate it; understand fundamental proofs for techniques that rely on matrix algebra; compute linear regression and apply hypothesis testing; perform logistic regression and discriminant analysis; apply the K-fold cross-validation methods; apply the LASSO and ridge regression methods; apply bagging and boosting on tree-based methods; apply some methods of unsupervised learning (e.g., principal components, or k-means clustering); manipulate data sets in R including applying the above methods. Course Objectives: The course is designed to introduce students to classical machine learning methods for regression and classification with an emphasis on model validation (i.e., it is not enough to fit a model, students should be able to estimate how good the resulting model is). By taking this course, students will gain experience in applying machine learning algorithms in R and develop skills for effectively communicating a proper interpretation of the results. Topics include Introduction, notation, terminology, Simple regression, model assessment, nonlinearity, Multiple linear regression, variable selection, categorical predictors, interactions, Classification via logistic regression, Discriminant analysis, distance measures, Unsupervised learning, Cross validation, bootstrap, Tree based methods, tuning parameters, Neural nets, shrinkage methods (LASSO and ridge regression), Dimensionality reduction via PCA, FA, NMF, Unsupervised learning with mixture models