MATH-5399 Independent Study for Bryan Barrios

The University of Texas at Tyler, Summer 2025

Instructor: Dr. Ivan Ramirez-Zuniga
Email: iramirezzuniga@uttyler.edu
Office: RBN 4005

Textbook:

- **Applied Predictive Modeling**, M. Kuhn & K. Johnson (Springer, 2013)
- **Machine Learning with R**, B. Lantz (Packt, 2015)

Course Description

This 3-credit independent study provides an intensive, hands-on exploration of predictive modeling and machine learning techniques. In this independent study, we will cover the following broad areas (subject to changes):

1. **Modeling Workflow**
    - Data pre-processing (cleaning, handling missing values, feature scaling/encoding)
    - Model validation & performance assessment (train/test split, cross-validation, metrics)
    - Model tuning & overfitting prevention
2. **Supervised Learning**
    - *Regression Models*
        - Linear and generalized linear models
        - Non-linear regression models
        - Regression trees and rule based models
    - *Classification Models*
        - Probabilistic classifiers (e.g. logistic regression, Naive Bayes)
        - Distance-based methods (k-NN)
        - Margin-based methods (SVMs)
        - Tree-based classifiers & ensemble methods (random forest, boosting)
        - Neural networks for classification
3. **Unsupervised Learning**
    - Clustering (K-means, hierarchical)
    - Dimensionality reduction (PCA)
    - Association rules

Course Format & Expectations

- **Weekly Meetings**: One-on-one with faculty mentor.
- **Preparation & Presentation**:
  - Study assigned chapter(s) and complete exercises before each meeting.
  - Begin each meeting with a student presentation covering key concepts learned during the previous week, exercise results, and questions.
- **Exercise Submission & Code Sharing**:
  - Written solutions to textbook exercises due by meeting time.
  - R code and scripts pushed to the course GitHub repository by meeting time.
  - Code should be well–commented, modular, and reproducible.
- **Final Project**:
  - Design and implement a complete predictive modeling pipeline on a real dataset.
  - Deliverables: reproducible report in an electronic notebook (eg. Jupyter Notebook) and 25 min presentation.

Grading

Grading is based on student performance and engagement:

- Weekly presentations & participation: 30%
- Exercise solutions & code quality: 30%
- Final project report & presentation: 40%

Grading scale

A:90-100%, B: 80-89%, C: 70-79%, D: 60−69%, F:<60%.