Homework: Week 8 (Due Monday, November 18, 2024)

FW 599: Multivariate Analysis of Ecological Data

Instructions

Please submit all homework assignments to Canvas as an **R markdown document** (Markdown or Quarto) including **visible code** and **relevant output**. A tidy *.pdf document is preferable to an *.html or a "raw" *.qmd file. Note that homework questions are intended to directly accompany lab exercises, building up to the final class project. Consequently, it is in your best interest to answer them thoroughly and thoughtfully.

Questions

Question 1) Using your dataset, create a classification or regression tree to predict a chosen response variable from a set of predictors. Then, build a Random Forest using the same data.

- **1a.** Use cross-validation to summarize the accuracy or MSE for each model.
- **1b.** Examine the confusion matrix output. Explain how the Random Forest approach affects predictive accuracy compared to the single tree.
- **1c.** Discuss any improvements you observe with Random Forests over single trees, and why this might be the case.

Question 2) Using the Random Forest output from Question 1, evaluate the importance of individual predictors in the model.

- **2a.** Report and visualize the predictor importance scores. What are the top predictors in each model?
- **2b.** Are the most important features in the Random Forest model the same as those in the single decision tree? Explain any differences.

Question 3) How does the ensemble nature of Random Forests affect model interpretation, and why might it present challenges or advantages for understanding predictor influence?