**R Short Course: Tree-Based Methods In-Class Activity**

**March 9, 2021**

Problems from James, Gareth, et al. *An introduction to statistical learning*. Vol. 112. New York: Springer, 2013. Chapter 8.

1. In the class example, a classification tree was applied to the Carseats data after converting Sales into a qualitative response variable. Now we will seek to predict Sales using regression trees and related approaches, treating the response as a quantitative variable.

	1. Split the data set into a training set and a test set.
	2. Fit a regression tree to the training set. Plot the tree, and interpret the results. What test MSE do you obtain?
	3. Use cross-validation in order to determine the optimal level of tree complexity. Does pruning the tree improve the test MSE?
2. Consider the Gini index, classification error, and cross-entropy in a simple classification setting with two classes. Create a single plot that displays each of these quantities as a function of $\hat{p\_{m1}}$. The $x$-axis should display $\hat{p\_{m1}}$, ranging from 0 to 1, and the $y$-axis should display the value of the Gini index, classification error, and entropy.

Hint: In a setting with two classes, $\hat{p\_{m1}}=1-\hat{p\_{m2}}$. You could make this plot by hand, but it should be much easier to make in R.