DSC 630
Laura Hoffmann
Assignment 7.3

Calculating Probability of Model Ensembles Report

Setup

The solution to this assignment seemed relatively simple. All I needed to do was create a function that used the number of models and error rate as arguments in calculating the probability of an incorrect prediction. In the function, an empty list was created, and then p was calculated by subtracting the argument of error rate from 1. I created a list of r-values by using the number of independent models argument and then used the binom function from scipy.stats to help calculate mass function values for each of the values in the R-values list. For each of the scenarios, I noticed the values for number of independent models were odd and we needed a majority rule, so I made a for loop that added one to each of the independent model variables and then divided by two in order to calculate for the probability of making an incorrect prediction. I was able to print results through the function for each of the scenarios.

Results

For the model ensemble with 11 independent models and an error rate of .2, there was a probability of 1.17% of the majority voting producing an incorrect prediction. For the ensemble with 11 independent models and an error rate of .49 the probability of an incorrect prediction jumped all the way up to 47.29%. For the model ensemble with 21 independent models and a .49 error rate the probability was still up to 46.3%. So we can see that the error rate effects this probability more than the number of models, although since it was majority rule the model numbers were basically cut in half, so 6 in the first two ensembles vs. 11 in the third ensemble. I printed a visual at the end to see the probabilities with each model.