

In this assignment, we want to calculate the probability of a model ensemble that uses simple majority voting making an incorrect prediction in a few different scenarios (differing number of models and their error rates). Essentially, we want to calculate the probability that a majority of the models in the ensemble will predict incorrectly. For this, we can use the binomial distribution probability, or the probability of getting one of two outcomes given a number of parameters: the number of independent models, the error rates of the models, and the majority number making incorrect predictions (always the lowest whole number to make a fraction above 50%, e.g., 6 if there are 11 models).

To calculate this probability, a function in Python was constructed that took the above parameters as arguments and output the probability, a number between 0 and 1. We found that the probability for the first scenario, 6 of 11 models predicting incorrectly with error rates of 0.2 each, was only 0.97% because the error rate for each model was so small. The probability for the second scenario, 6 of 11 models predicting incorrectly with error rates of 0.49 each, was much higher at 22% because each model now had an error of 0.49 or 49%. The probability for the third scenario, 11 of 21 models predicting incorrectly with error rates of 0.49 each, dropped to 16.4% because the number of models needed to get a majority predicting incorrectly increased to 11.