DSC 630: Predictive Analytics

Christopher M. Anderson

05/09/2020

Week 9

Assignment 9.4: Calculate Probability of a Model Ensemble

Calculate the probability of a model ensemble that uses simple majority voting making an incorrect prediction in the following scenarios. (Hint: Understanding how to use the binomial distribution will be useful in answering this question.)

1. The ensemble contains 11 independent models, all of which have an error rate of 0.2.

2. The ensemble contains 11 independent models, all of which have an error rate of 0.49.

3. The ensemble contains 21 independent models, all of which have an error rate of 0.49.

The Binomial Distribution Formula

The binomial distribution formula can be used to calculate the discrete probability of number of successes and failures in *n* number of independent trials or experiments. *P(x)* is the probability of *x* successes occur in the *n* number of events, *p* is the probability of success and *q* is the probability of failure often denoted by *q = (1 - p)*.

Here is the formula we use:

$$P(X = x) =^{n} C_x p^x (1-p)^{(n-x)}$$

$$x = 0, 1, 2, . . , n$$

Where:

- *n* = Number of events
- *x* = Number of successes
- *p* = Probability of success

With the formula established, we can now work through the three examples.

Ensemble Model 1:

The ensemble contains 11 independent models, all of which have an error rate of 0.2.

Now let's enter this information into the formula:

- n = 11
- x = 6 (6 is the majority of 11)
- p = 0.8 (since the error rate is 0.2, the probability of success is 1 - 0.2)

$$^{11}C_6 0.8^6 (1-0.8)^{(11-6)} = 0.0388$$

Ensemble Model 2:

The ensemble contains 11 independent models, all of which have an error rate of 0.49.

Now let's enter this information into the formula:

- n = 11
- x = 6 (6 is the majority of 11)
- p = 0.51 (since the error rate is 0.49, the probability of success is 1 - 0.49)

$$^{11}C_6 0.51^6 (1-0.51)^{(11-6)} = 0.2296$$

Ensemble Model 3:

The ensemble contains 21 independent models, all of which have an error rate of 0.49.

Now let's enter this information into the formula:

- n = 21
- x = 11 (11 is the majority of 21)
- p = 0.51 (since the error rate is 0.49, the probability of success is 1 - 0.49)

$$^{21}C_{11} 0.51^{11} (1-0.51)^{(21-11)} = 0.1709$$