Assignment34

John Bute

2024-11-17

Suppose we produce ten bootstrapped samples from a data set containing red and green classes. We then apply a classification tree to each bootstrapped sample and, for a specific value of X, produce 10 estimates of P(Class is Red|X):

0.1, 0.15, 0.2, 0.2, 0.55, 0.6, 0.6, 0.65, 0.7, and 0.75.

There are two common ways to combine these results together into a single class prediction. One is the majority vote; the second approach is to classify based on the average probability. In this example, what is the final classification under each of these two approaches?

For a majority vote, we need to take each bootstrapped classifier and see if the estimate probability of each one is greater than 0.5:

P(Class is Red | X) > 0.5 -> Red P(Class is Red | X) <= 0.5 -> Red

So out of the 10 estimates:

0.1, 0.15, 0.2, 0.2 predicted Green (4 votes) 0.55, 0.6, 0.6, 0.65, 0.7, 0.75 predicted Red (6 votes) Since red got the majority vote, then the final classification is Red

For average probability: P(Class is Red | X) -> Red IF Average(predicted probabilities of bootstrapped classifier) > 0.5 otherwise Green

therefore (0.1 + 0.15 +0.2 + 0.2 + 0.55 + 0.6 + 0.6 + 0.65 + 0.7 + 0.75)/ 10 = 0.45 0.45 < 0.5 therefore, the final classification is Green.