Sam Taylor

Professor Lukasz Kurgan, Ph.D.

CMSC 435

9/7/2023

Q1.0

I really liked the article. I found the article to be an engaging exploration of some of the most popular data mining algorithms. It summarized the defining aspects of the algorithm while remaining concise. I liked the way it covered how every algorithm could be used correctly or incorrectly. Mentioning overfitting and not diversifying the methods that you use in particular could be useful for learning. On the other hand, I didn't like that it switched between third person and first-person plural at times because I found it hard to focus on the results of the algorithms. It also uses a lot of jargon which some people may not be familiar with.

Q2.1

C4.5 and CART and their different iterations both generate decision trees. CART was created in 1984 and C4.5 was created in 1993 so CART was published first.

Q2.2

It assumes that the components of x are independent and reduces the *p* dimensional multivariate problems to *p* univariate estimation problems. Univariate estimation is familiar, simple, and requires smaller training set sizes to obtain accurate estimates while still allowing estimation to be simple, very quick, and not requiring complicated iterative estimation schemes. The word bayes refers to a mathematician who developed a method of determining conditional probability, which is used in the Naïve Bayes method.

Q2.3

K is the number of objects that are close to the test object that the algorithm will use to count the prominence of certain things in each object. K could be set to 1 but it would have to be for a very specific circumstance because having such a small k would make it sensitive to noise points. Since there are more occurrences of class 2 than all the others the majority vote prediction for the new data point would be class 2.

Q2.4

PageRank was created by Sergey Brin and Larry Page as part of founding Google. It uses directed graphs to represent the internet. The vertices make up pages to different websites and the edges make up the hyperlinks that connect them to each other.