Variable selection via information gain

Ting-Li Chen

Institute of Statistical Science, Academia Sinica, Taiwan

Abstract

Variable selection is a critical step in building statistical models, as it helps to identify the most important predictors for explaining the response variable. In this talk, we will focus on an information gain-based variable selection technique that can effectively handle non-linear relationships between the response and the covariates. We will introduce the information gain criterion, which compares the entropy of the response variable to the conditional entropy of the response variable given an explanatory variable, to identify the variables that best explain the response. Finally, we will demonstrate the strength of this technique through examples, and compare it with other variable selection methods.