

1

Introdução

Dado um modelo STR-Tree (*Tree-Structured Smooth Transition Regression*) proposto em (7), que é composto pela fusão do algoritmo CART (*Classification and Regression Tree*), apresentado em (2), por sua praticidade e interpretabilidade; e o modelo STR (*Smooth Transition Regression*) encontrado e amplamente debatido em (10), que garante a utilização dos métodos da Inferência Estatística clássica, permitindo-nos lançar mão de testes de hipóteses, intervalos de confiança e garantindo a significância estatística dos parâmetros.

Contudo, no trabalho citado inicialmente, o STR-Tree é utilizado para fazer regressão e não, como será proposto neste estudo, para classificação, sendo confrontado com alguns dos usuais métodos existentes como: *Regressão Logística, Análise Discriminante, CART, k-Vizinhos e GAM*.

O processo de Classificação nada mais é do que, por meio do conjunto de dados previamente classificados, gerar classificadores que descrevam ou distingam classes de dados ou conceitos mediante um rótulo, que não é mais do que um valor de um atributo. Em posse do classificador, poderemos testar seu poder classificatório através de dados que saibamos previamente a qual classe estes pertencem e, com isso possamos confrontar com o resultado obtido por aquele. Após tais testes, podemos utilizar o método para prever qual será a classificação de outros dados dos quais a classe é desconhecida.

Iremos propor uma adaptação feita ao modelo STR-Tree, inserindo o mesmo em um contexto de Regressão Logística para a aplicação em dados reais. Levando em consideração que algumas mudanças na estrutura e na forma de estimação foram feitas, passaremos a chamá-lo de STLR-Tree (*Smooth Transition Logistic Regression-Tree*). As mudanças mencionadas são: não considerar mais a possibilidade de que as variáveis sejam correlacionadas no tempo, como no caso das séries temporais, não admitindo a utilização de variáveis defasadas dentro do conjunto de variáveis explicativas tão pouco dentro das variáveis de transição. Além disso, a estimação dos parâmetros lineares não será mais feita através do usual método de Mínimos Quadrados Não Lineares (MQNL), dado o tipo de variáveis dependentes em que se aplica a Regressão Logística, variáveis categóricas. Assim será desenvolvida a função de

Máxima Verossimilhança para o modelo adaptado, a fim de se fazer a estimação dos parâmetros. A parte não-linear é estimada por *grid* e, os testes para a inclusão de novos regimes não são mais aqueles utilizados com a estatística F, onde medimos a soma dos quadrados dos resíduos (SQR) dos modelos testados, pois em Regressão Logística não se tem uma estrutura sistemática de resíduos e os testes, que utilizamos para a divisão dos nós, são os testes de Razão de Verossimilhança.

A capacidade de classificação do modelo STLR-Tree será comparada com outros métodos tais como: *Generalized Additive Models (GAM)*, *Regressão Logística*, *Redes Neurais*, *Análise Discriminante*, *k-Nearest Neighbor (k-NN)* e *Classification and Regression Trees (CART)*.