Pontifícia Universidade Católica
do Rio de Janeiro

Jessica Quintanilha Kubrusly

Regressão Construtiva por Regiões Definidas Implicitamente

Tese de Doutorado

Tese apresentada ao Programa de Pós–graduação em Matemática do Departamento de Matemática da PUC–Rio como requisito parcial para obtenção do título de Doutor em Matemática

Orientador: Prof. Hélio Côrtes Vieira Lopes

Rio de Janeiro
junho de 2009

Pontifícia Universidade Católica
do Rio de Janeiro

Jessica Quintanilha Kubrusly

Regressão Construtiva por Regiões Definidas Implicitamente

Tese apresentada ao Programa de Pós–graduação em Matemática do Departamento de Matemática do Centro Técnico Científico da PUC–Rio como requisito parcial para obtenção do título de Doutor em Matemática. Aprovada pela comissão examinadora abaixo assinada.

Prof. Hélio Côrtes Vieira Lopes
Orientador
Departamento de Matemática — PUC–Rio

Prof. Beatriz Vaz de Melo Mendes
Departamento de Métodos Estatísticos do Instituto de Matemática — UFRJ

Prof. Dirce Uesu Pesco
Departamento de Geometria — UFF

Prof. Luiz Henrique de Figueiredo
IMPA

Prof. Cristiano Fernandes
Departamento de Engenharia Elétrica — PUC–Rio

Prof. Sinesio Pesco
Departamento de Matemática — PUC–Rio

Prof. Thomas Lewiner
Departamento de Matemática — PUC–Rio

Prof. José Eugênio Leal
Coordenador Setorial do Centro Técnico Científico — PUC–Rio

Rio de Janeiro, 22 de junho de 2009

Jessica Quintanilha Kubrusly

No ano de 2003 graduou-se em Engenharia Elétrica pela Pontifícia Universidade Católica do Rio de Janeiro, com ênfase em Sistemas de Apoio à Decisão. Em 2005 concluiu seu mestrado em Matemática Aplicada, também pela Pontifícia Universidade Católica do Rio de Janeiro, onde desenvolveu junto com seus orientadores uma ferramenta estatística para o cálculo de reservas.

Ficha Catalográfica

Kubrusly, Jessica

Regressão Construtiva por Regiões Definidas Implicitamente / Jessica Quintanilha Kubrusly; orientador: Hélio Côrtes Vieira Lopes. — Rio de Janeiro : PUC–Rio, Departamento de Matemática, 2009.

v., 89 f: il. ; 29,7 cm

1. Tese (doutorado) - Pontifícia Universidade Católica do Rio de Janeiro, Departamento de Matemática.

Inclui referências bibliográficas.

1. Matemática – Tese. 2. Árvores de Regressão. 3. Estatística. 4. Modelos não Lineares. 5. Modelos não Paramétricos. 6. IBNR. I. Lopes, Hélio Côrtes Vieira. II. Pontifícia Universidade Católica do Rio de Janeiro. Departamento de Matemática. III. Título.

CDD: 510

Agradecimentos

Agradeço à Funenseg, pela bolsa concedida, à PUC-Rio, pelo suporte em todos os aspectos e ao professor Luiz Roberto Cunha, pela confiança e incentivo durante esse período. Sem dúvida foram ajudas fundamentais para que esse trabalho fosse concretizado.

Gostaria muito de agradecer ao professor Hélio pela sua dedicação, não só nesses últimos quatro anos de doutorado, mas também por todos os oito anos de convivência. Sua simpatia e profissionalismo contribuíram muito para que esta tese acontecesse e para que a relação entre aluno e orientador fosse a melhor possível.

Não poderia deixar de agradecer a todos da PUC-Rio, funcionários, alunos e professores, que fizeram parte desse agradável ambiente acadêmico e por isso são totais responsáveis pela enorme saudade que sinto desde já.

Por último, quero agradecer aos familiares e amigos, que nesses últimos meses foram muito compreensíveis com a minha constante ansiedade e falta de tempo por causa da tese.

Resumo

Kubrusly, Jessica; Lopes, Hélio Côrtes Vieira. **Regressão Construtiva por Regiões Definidas Implicitamente**. Rio de Janeiro, 2009. 89p. Tese de Doutorado — Departamento de Matemática, Pontifícia Universidade Católica do Rio de Janeiro.

Os métodos de regressão baseados em árvores são modelos não lineares e não paramétricos, estudados desde a década de 80, quando houve a criação do algoritmo CART. Até hoje há muita pesquisa nessa área e cada vez mais novos métodos são apresentados com o objetivo de aperfeiçoar os modelos já existentes. Esse trabalho propõe um novo método chamado de Regressão Construtiva em Regiões Implícitas (RCRI). Sua principal diferença, com relação aos demais métodos baseados em árvores, está na forma como o domínio é particionado. Até o momento essa partição era formada por retângulos com arestas paralelas aos eixos, porém o algoritmo RCRI permitiu que as partições fossem formadas por regiões mais flexíveis definidas implicitamente. Além disso, o trabalho também propõe uma extensão intervalar para o modelo. Duas diferentes aplicações desse novo método também são sugeridas. A primeira em atuária, que busca melhorar a estimativa da reserva IBNR fornecida pelo já usual modelo *Chain Ladder*. A segunda em geologia, que utiliza as informações existentes nos poços para realizar inferências sobre dados faltantes.

Palavras–chave

Árvores de Regressão. Estatística. Modelos não Lineares. Modelos não Paramétricos. IBNR.

Abstract

Kubrusly, Jessica; Lopes, Hélio Côrtes Vieira (Adviser). **Constructive Regression on Implicity Defined Regions**. Rio de Janeiro, 2009. 89p. D.Sc. Thesis — Departamento de Matemática, Pontifícia Universidade Católica do Rio de Janeiro.

Tree-based methods are playing an important role in non-linear and non-parametric regression. They have been studied since the 80's, when the CART algorithm was proposed. Nowadays there is a lot of research in this area and new methods are being created, aiming at improving existing models. This work proposes a new tree-based method called Constructive Regression on Implicit Regions. Its main difference, with respect to other tree-based methods, is how the domain is partitioned. The proposed algorithm allows partitions formed by flexible regions whose borders are implicitly defined. Moreover, the work also proposes an interval extension to the model. Two different applications of this new method are also proposed. The first one is in actuary, which looks for improvements in the estimation of IBNR reserves, already provided by the usual Chain Ladder model. The second one is in geology, which uses the well data to perform inferences about the missing data in the well itself.

Keywords

Regression Trees. Statistics. Non Linear Models. Non Parametric Models. IBNR.

Sumário

Lista de figuras

Lista de tabelas