



Métodos de Regressão Quantílica Não Paramétrico utilizando inteligência de exames para dados simbólicos do tipo intervalo

NEVES JUNIOR, R. B.¹, rbnj@ecomp.poli.br
FAGUNDES, R. A. A.², roberta.fagundes@upe.br

Análise de regressão é uma ferramenta estatística que investiga a relação entre duas ou mais variáveis. As variáveis podem apresentar-se em sua representação clássica ou através dos dados simbólicos do tipo intervalo (DSTI) que fazem parte do contexto da Análise de Dados Simbólicos (ADS) (BILLARD, 2006). A regressão quantílica (RQ) introduzida por Koenker (KOENKER, 2006), destaca-se em relação aos demais tipos de regressão, enquanto as regressões usuais permitem a interpretação do modelo com apenas uma curva, a RQ permite a utilização de várias curvas (ou quantís) para que se possa obter uma visão mais completa da relação entre as variáveis estudadas. Na utilização modelo de regressão quantílica não paramétrica (RQNP), o kernel gaussiano pode ser aplicado mediante ao ajuste do parâmetro largura de banda (i.e., h), que controla o grau de suavidade da função estimada. Atualmente na literatura o modelo existente de RQNP é utilizado para dados clássicos pontuais, entretanto existem situações onde há necessidade de aplicar modelos de regressão para DSTI. O *Particle Swarm Optimization* (PSO) é um algoritmo da área de Inteligência de Enxames, inspirado no comportamento dos pássaros, capaz de resolver problemas de otimização em espaços de busca multidimensionais (KENNEDY, 1995). Neste contexto, é proposto: (i) desenvolver um modelo de RQNP para DSTI; (ii) desenvolver dois modelos de RQNP híbridos com PSO, capazes de estimar o parâmetro largura de banda do kernel gaussiano; sendo um para dados clássicos pontuais e outro para DSTI. Como resultado preliminar, foi desenvolvido o modelo de regressão quantílica intervalar não paramétrica (RQINP) citado como primeiro objetivo e comparado com o modelo regressão quantílica intervalar (RQI) introduzida por Fagundes et al., (FAGUNDES, 2016). Como metodologia adotada, os modelos RQINP e RQI foram submetidos à experimentos utilizando dados reais e simulados em que estes dados possuíam um comportamento predominantemente não linear com ruídos e *outliers*. Após a realização dos experimentos, foi utilizado o teste estatístico de Wilcoxon com 5% de significância para avaliar se o modelo RQINP é estatisticamente superior ao RQI. Concluiu-se que o modelo RQINP obteve melhores resultados porque o modelo não paramétrico é mais robusto na presença de dados não lineares, com ruídos e *outliers*.

Palavras-chave: *Análise de Regressão; Análise de Dados simbólicos; Inteligência de Enxame;*

Referências

- L. BILLARD AND E. DIDAY. **Symbolic Data Analysis: Conceptual Statistics and Data Mining**. Wiley, West Sussex, England, 2006.
- KOENKER, ROGER. **Quantile regression**. No. 38. Cambridge university press, 2005.
- J. KENNEDY E R. EBERHART. **Particle swarm optimization**. vol. 4, pp 4945-4948.



MOSTRA POLI 2017

FAGUNDES, ROBERTA AA, RENATA MCR DE SOUZA, AND YANNE MG SOARES. **Quantile regression of interval-valued data.** Pattern Recognition (ICPR), 2016 23rd International Conference on. IEEE, 2016.