

**UNIVERSIDADE ESTADUAL DE MARINGÁ
CENTRO DE CIÊNCIAS EXATAS
CURSO DE ESTATÍSTICA**

**AVALIAÇÃO DE MÉTODOS NÃO PARAMÉTRICOS
PARA PREDIÇÃO EM MODELOS ADITIVOS**

Marco Aurelio Valles Leal

Maringá
2022

MARCO AURELIO VALLES LEAL

**AVALIAÇÃO DE MÉTODOS NÃO PARAMÉTRICOS
PARA PREDIÇÃO EM MODELOS ADITIVOS**

Trabalho de conclusão de curso apresentado
como requisito parcial para a obtenção do título
de bacharel em Estatística pela Universidade
Estadual de Maringá.

Orientador: Profº Drº George Lucas Moraes Pezzot

Coorientador: Profº Drº Willian Luís de Oliveira

Maringá
2022

AVALIAÇÃO DE MÉTODOS NÃO PARAMÉTRICOS PARA PREDIÇÃO EM MODELOS ADITIVOS

MARCO AURELIO VALLES LEAL

Trabalho de conclusão de curso apresentado como requisito parcial para a obtenção do título de bacharel em Estatística pela Universidade Estadual de Maringá.

Aprovado em: ____/____/____.

BANCA EXAMINADORA

Orientador

Profº Drº George Lucas Moraes Pezzot
Universidade Estadual de Maringá

Membro da banca

Nome do professor membro da banca
Instituição do professor membro da banca

Membro da banca

Nome do professor membro da banca
Instituição do professor membro da banca

RESUMO

É comum, nas mais diversas áreas, investigar e modelar a relação entre variáveis. O modelo mais simples é denominado modelo de regressão linear simples e assume que a média da variável resposta é modelada como uma função linear das variáveis explicativas, supondo erros aleatórios com média zero, variância constante e não correlacionados. Entretanto, nem sempre a relação existente é perfeitamente linear. Neste contexto, é possível flexibilizar o modelo de regressão linear modelando a dependência da variável resposta com cada uma das variáveis explicativas em um contexto não paramétrico. Esta nova classe de modelos é dita modelos aditivos e mantém a característica dos modelos de regressão lineares de serem aditivos nos efeitos preditivos. Portanto, este projeto visa apresentar os modelos aditivos, além de técnicas de suavização utilizadas para ajustar modelos no contexto não paramétrico. Por fim, a metodologia é aplicada em dados artificiais (simulados) e em dados reais, dando enfoque à qualidade das predições.

Palavras-chave : Regressão. Modelo aditivo. Suavizadores.

Sumário

0.1	Aplicações	1
0.1.1	Aplicação 1	1
0.2	Aplicação 2	4
1	Referências	6

0.1 Aplicações

Nesta seção

0.1.1 Aplicação 1

Para esta aplicação serão empregadas as técnicas de suavização em dados reais. Os dados foram retirados do site NIST Standard Reference Database 140 ¹. É um estudo referente a expansão térmica de cobre. A variável resposta é o coeficiente de expansão térmica e a variável preditora é a temperatura em graus kelvin. Neste trabalho sera abordado um modelo com apenas uma covariável neste caso, sendo o modelo aditivo da seguinte forma

$$y = \alpha + f(X) + \epsilon,$$

onde os erros ϵ são independentes, com $E(\epsilon) = 0$ e $var(\epsilon) = \sigma^2$. A $f(X)$ é uma função univariada arbitrária, que será suavizada pelos métodos vistos até o momento.

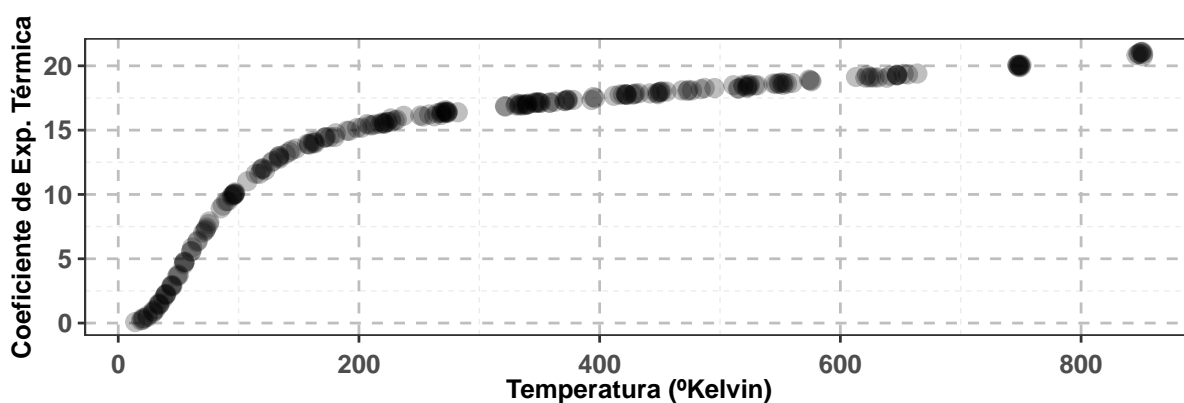


Figura 1: Gráfico de dispersão referente ao estudo de expansão térmica do cobre versus a temperatura em graus Kelvin.

Na Tabela 2 é apresentado a os EQM's proveniente do processo de *leave one out cross-validation* para os suavizadores *kernel*, *loess*, *splines* de regressão consdierando grau 1 e grau 3.

¹<https://www.itl.nist.gov/div898/strd/index.html>

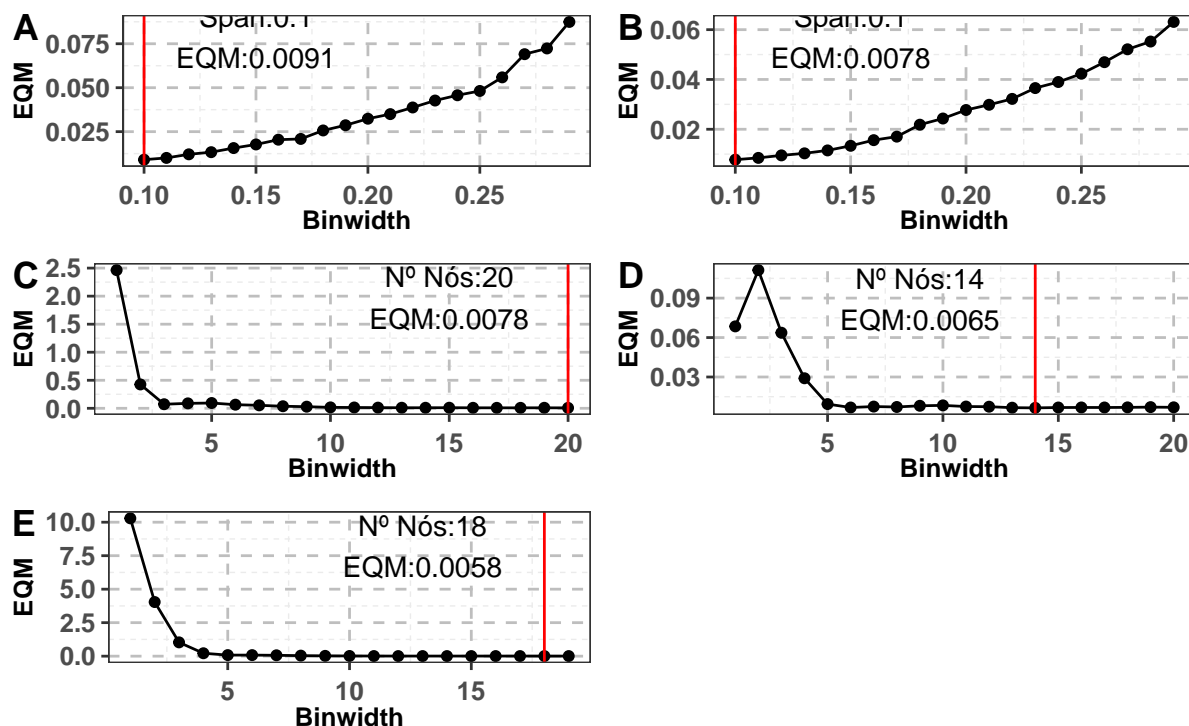


Figura 2: Comparação entre diferentes ajustes para valores de parâmetros distintos, considerando os suavizadores (A) Kernel, (B) Loess, (C) Splines de Regressão Grau 1 e (D) Splines de Regressão Grau 3.

Tabela 1: Erro Quadrático Médio para os suavizadores Loess, Kernel e Spline Cúbico

Smoother	EQM
Kernel	0.0090683
Loess	0.0077765
Splines Grau 1	0.0077643
Splines Grau 3	0.0065196
Poly	0.0058110

Na Figura 6 é apresentado os ajustes para cada técnica considerando os melhores parâmetros obtidos por meio da validação cruzada.

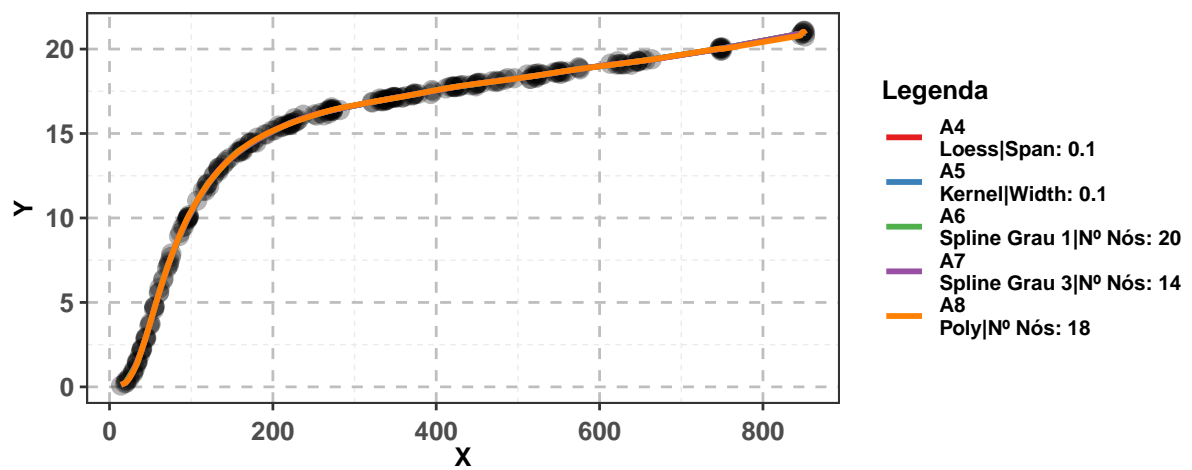


Figura 3: Comparação dos ajustes entre os métodos

0.2 Aplicação 2

site : <https://www.itl.nist.gov/div898/strd/lls/data/Filip.shtml>

site : <https://www.itl.nist.gov/div898/strd/lls/data/LINKS/DATA/Filip.dat>

NIST/ITL StRD Dataset Name: Filip (Filip.dat)

File Format: ASCII Certified Values (lines 31 to 55) Data (lines 61 to 142)

Procedure: Linear Least Squares Regression

Reference: Filippelli, A., NIST.

Data: 1 Response Variable (y) 1 Predictor Variable (x) 82 Observations Higher Level of Difficulty Observed Data

Model: Polynomial Class 11 Parameters (B0,B1,...,B10)

$$y = B_0 + B_1 \cdot x + B_2 \cdot (x^{**2}) + \dots + B_9 \cdot (x^{**9}) + B_{10} \cdot (x^{**10}) + e$$

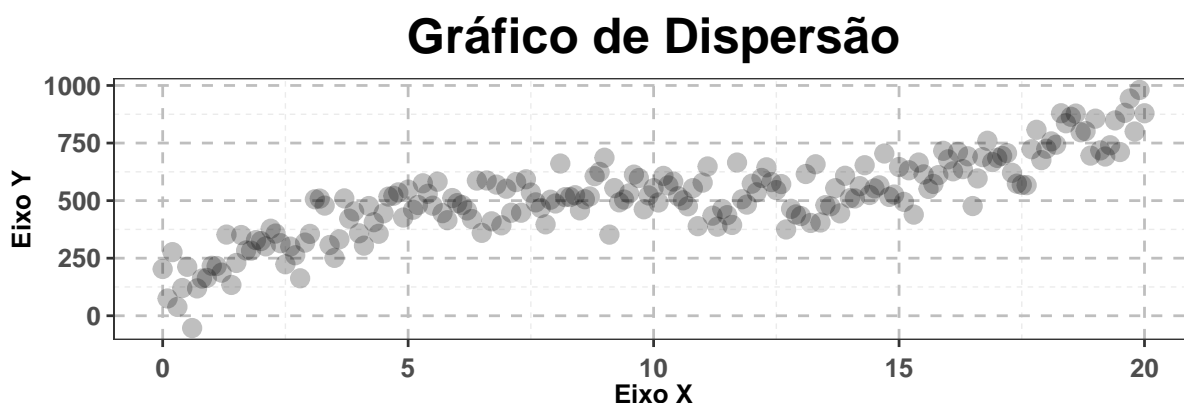


Figura 4: Gráfico de dispersão dos dados gerados para o estudo de simulação

Na Tabela 2 é apresentado a os EQM's proveniente do processo de *leave one out cross-validation* para os suavizadores *kernel*, *loess*, *splines* de regressão consdierando grau 1 e grau 3.

Tabela 2: Erro Quadrático Médio para os suavizadores Loess, Kernel e Spline Cúbico

Smoother	EQM
Kernel	6582.860
Loess	6590.736
Splines Grau 1	6487.688
Splines Grau 3	6528.207
Poly	5480.951

Na Figura 6 é apresentado os ajustes para cada técnica considerando os melhores parâmetros obtidos por meio da validação cruzada.

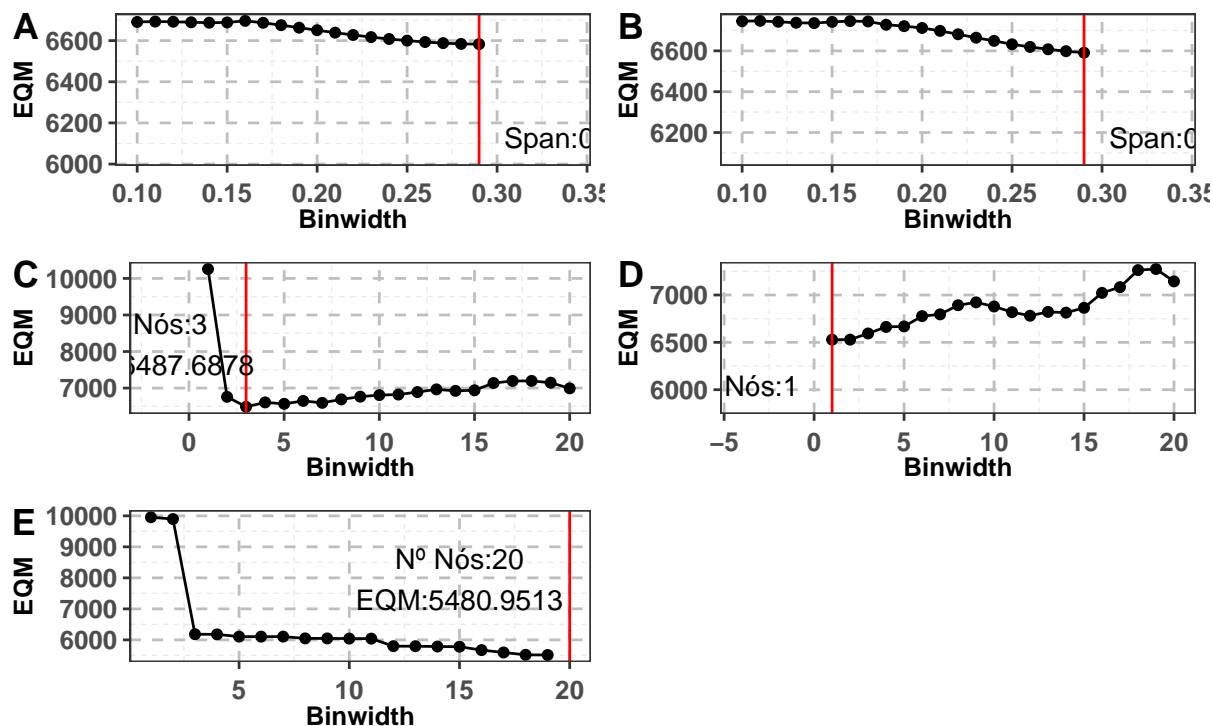


Figura 5: Comparação entre diferentes ajustes para valores de parâmetros distintos, considerando os suavizadores (A) Kernel, (B) Loess, (C) Splines de Regressão Grau 1 e (D) Splines de Regressão Grau 3.

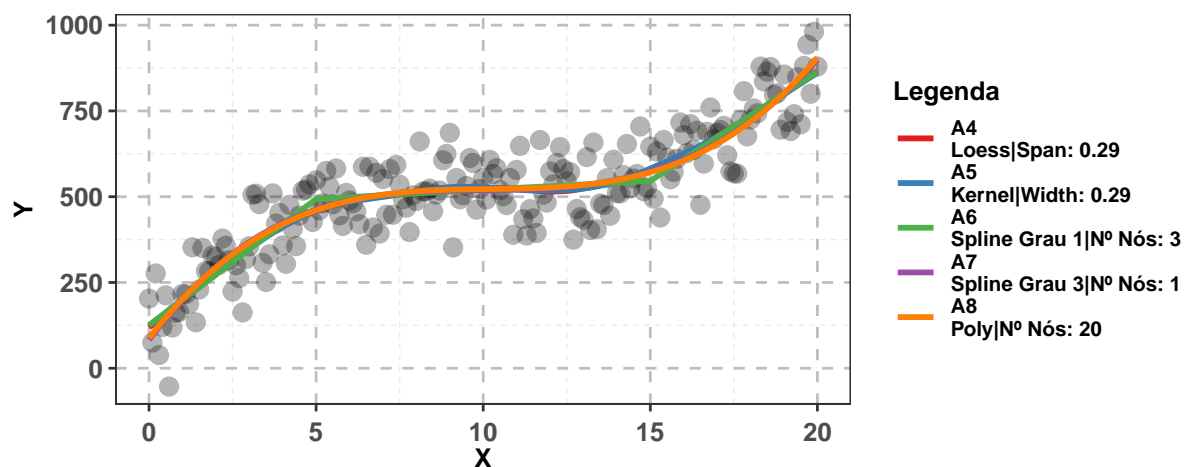


Figura 6: Comparação dos ajustes entre os métodos

1 Referências

- BUJA, A., HASTIE, T. & TIBSHIRANI, R. (1989). **Linear smoothers and additive models**. The Annals of Statistics, 17, 453-510.
- CLEVELAND, W. S. (1979). **Robust locally weighted regression and smoothing scatter-plots**. Journal of the American Statistical Association, 74, 829-836.
- DELICADO, P., 2008 **Curso de Modelos no Paramétricos** p. 200.
- EUBANK, R. L(1999) **Nonparametric Regression and Spline Smoothing**. Marcel Dekker, 2o edição. Citado na pág. 1, 2, 29
- FAHRMEIR, L. & TUTZ, G. (2001) **Multivariate Statistical Modelling Based on Generalized Linear Models**. Springer, 2o edição. Citado na pág. 15
- GREEN, P. J. & YANDELL, B. S. (1985) **Semi-parametric generalized linear models**. Lecture Notes in Statistics, 32:4455. Citado na pág. 15
- GREEN P. J. & SILVERMAN B. W. (1994). **Nonparametric regression and generalized linear models: a roughness penalty approach**. Chapman & Hall, London.
- HASTIE, T. J. & TIBSHIRANI, R. J. (1990). **Generalized additive models**, volume 43. Chapman and Hall, Ltd., London. ISBN 0-412-34390-8.
- MONTGOMERY, D. C. & PECK, E. A. & VINING, G. G. **Introduction to Linear Regression Analysis**. 5th Edition. John Wiley & Sons, 2012.
- IZBICK, r & SANTOS, T. M. **Aprendizado de máquina: uma abordagem estatística**. ISBN 978-65-00-02410-4.
- TEAM, R. CORE. R: **A language and environment for statistical computing**. (2013).