

Avaliação de métodos não paramétricos para predição em modelos aditivos

Orientador(^a): Prof^o Dr^o George Lucas Moraes Pezzot

Co-orientador(^a): Prof^o Dr^o Willian Luís de Oliveira

Aluno(^a) : Marco Aurelio Valles Leal

RA: 103159

Maringá, 3 de maio de 2022

Conteúdo

Introdução

Objetivo

Metodologia

Resultados e Discussão

Considerações Finais

Introdução

Introdução

Objetivo

Objetivos

Metodologia

Metodologia

Resultados e Discussão

Cenário 1 - Introdução

- ▶ Valores de X uma sequência de 0 a 5;
- ▶ $y = 10 + 5\text{sen}\pi\frac{x}{24} + \varepsilon$;
- ▶ ε é um termo aleatório, normalmente, distribuído com média zero e variância constante;
- ▶ Tamanos amostrais utilizados serão iguais a 150, 250 e 350;
- ▶ Valores de desvio padrão, utilizado nos erros, 0.5, 1 e 2.

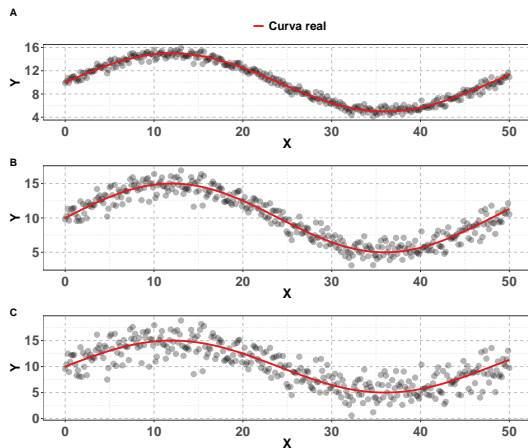


Figura 1: Gráfico de dispersão dos dados simulados e curva real.

Cenário 1 - Ajustes suavizados

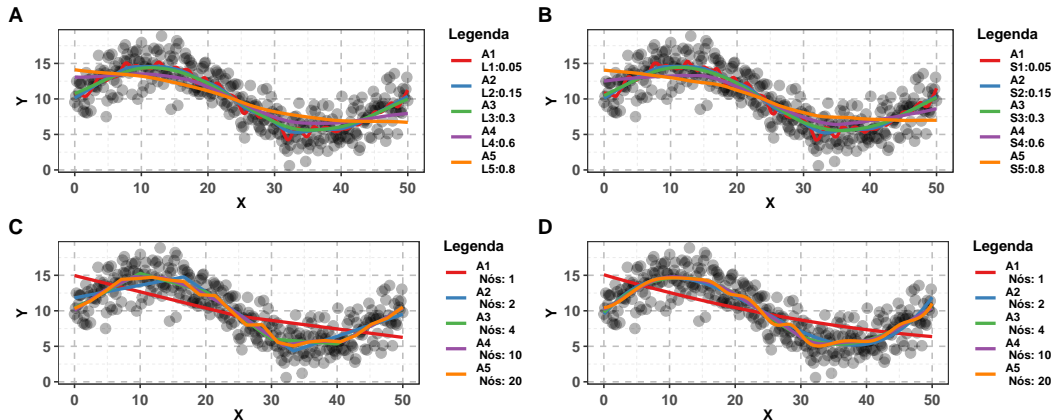


Figura 2: Comparação entre diferentes ajustes, com parâmetros de suavização distintos, considerando os suavizadores (A) Kernel, (B) Loess, (C) Splines de Regressão Grau 1 e (D) Splines de Regressão Grau 3.

Cenário 1 - Leave One Out Cross-Validation

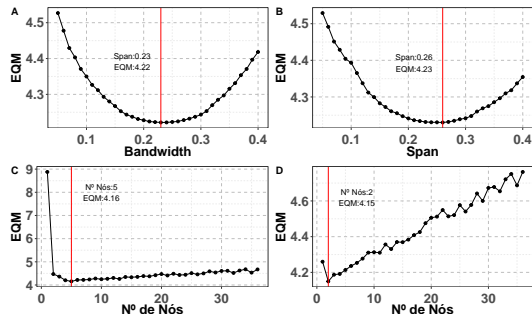
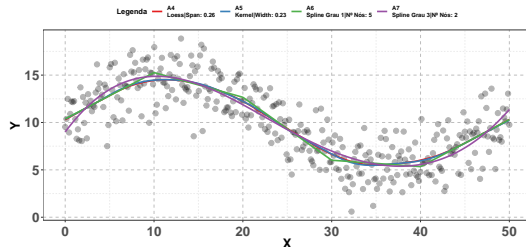


Figura 3: Erro quadrático médio versus parâmetro de suavização pós aplicação do Leave One Out Cross-Validation. (A) Kernel, (B) Loess, (C) Splines de Regressão Grau 1 e (D) Splines de Regressão Grau 3.

Tabela 1: Erro Quadrático Médio para os suavizadores Loess, Kernel, Splines de Regressão Linear e Splines de Regressão Cúbico

Smoother	EQM
Kernel	4.2216
Loess	4.2303
Splines Grau 1	4.1564
Splines Grau 3	4.1497



Cenário 1 - Simulação MCMC

Tabela 2: Percentual do Erro quadrático mínimo para cada suavizador em relação a 1000 amostras.

TAMANHO	VAR	Kernel	Loess	Sp. Reg. 1	Sp. Reg. 3
150	0.5	0.9%	5.3%	19.7%	74.1%
150	1.0	1.4%	8.4%	31.1%	59.1%
150	2.0	1.4%	10.8%	48%	39.8%
250	0.5	0.7%	3.8%	16.9%	78.6%
250	1.0	1.1%	7%	23.5%	68.4%
250	2.0	1.7%	13.1%	35.8%	49.4%
350	0.5	0.8%	3.7%	15.8%	79.7%
350	1.0	1.3%	7%	20.5%	71.2%
350	2.0	1.8%	11.7%	33.5%	53%

Cenário 1 - Comportamento dos EQM's

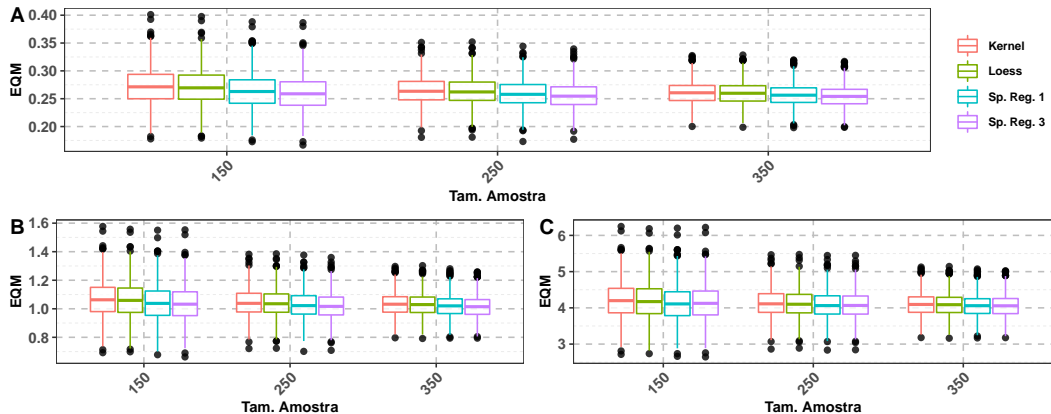


Figura 5: Comparação do erro quadrático para as 1000 amostras para cada suavizador por (A) $DP = 0.5$, (B) $DP = 1$ e (C) $DP = 2$

Cenário 2 - Introdução

- ▶ Valores de X uma sequência de 0.1 a 2;
- ▶ $y = f(x) + \varepsilon$, com $f(x) \sim \text{Gamma}(6, 10)$;
- ▶ $\varepsilon \sim N(0, \sigma^2)$;
- ▶ Tamanhos amostrais utilizados serão iguais a 150, 250 e 350;
- ▶ Valores de desvio padrão, utilizado nos erros, 0.05, 0.1 e 0.15.

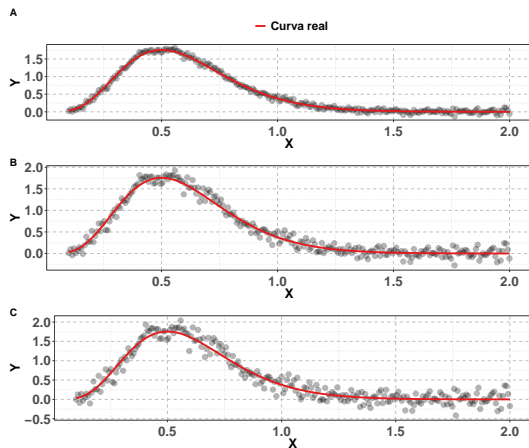


Figura 6: Gráfico de dispersão dos dados gerados para o estudo de simulação.

Cenário 2 - Leave One Out Cross-Validation

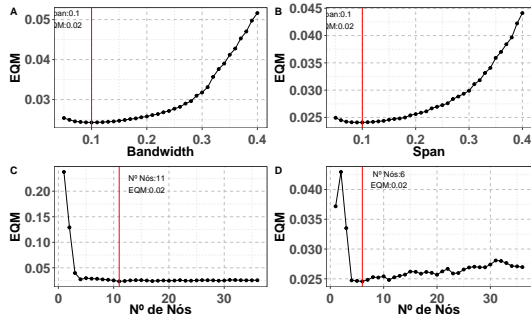
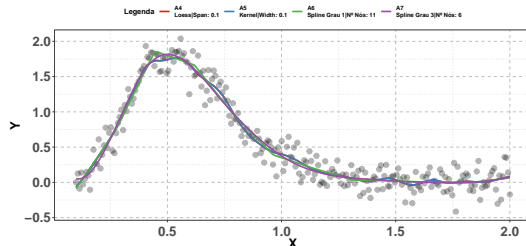


Figura 7: Erro quadrático médio versus parâmetro de suavização pós aplicação do Leave One Out Cross-Validation. (A) Kernel, (B) Loess, (C) Splines de Regressão Grau 1 e (D) Splines de Regressão Grau 3.

Tabela 3: Erro Quadrático Médio para os suavizadores Loess, Kernel, Splines de Regressão Linear e Splines de Regressão Cúbico

Smoother	EQM
Kernel	0.0243
Loess	0.0241
Splines Grau 1	0.0234
Splines Grau 3	0.0246



Cenário 2 - Simulação MCMC

Tabela 4: Percentual do Erro quadrático mínimo para cada suavizador em relação a 1000 amostras.

TAMANHO	VAR	Kernel	Loess	Sp. Reg. 1	Sp. Reg. 3
150	0.5	0.6%	2.6%	7.4%	89.4%
150	1.0	0.7%	6.3%	13.5%	79.5%
150	2.0	0.7%	9.1%	20.6%	69.6%
250	0.5	0.4%	1.7%	8.2%	89.7%
250	1.0	0.7%	3.4%	10.3%	85.6%
250	2.0	0.8%	6.5%	12.7%	80%
350	0.5	0.2%	2%	5.9%	91.9%
350	1.0	0.6%	2.5%	8.9%	88%
350	2.0	1%	4.7%	11.4%	82.9%

Cenário 2 - Comportamento dos EQM's

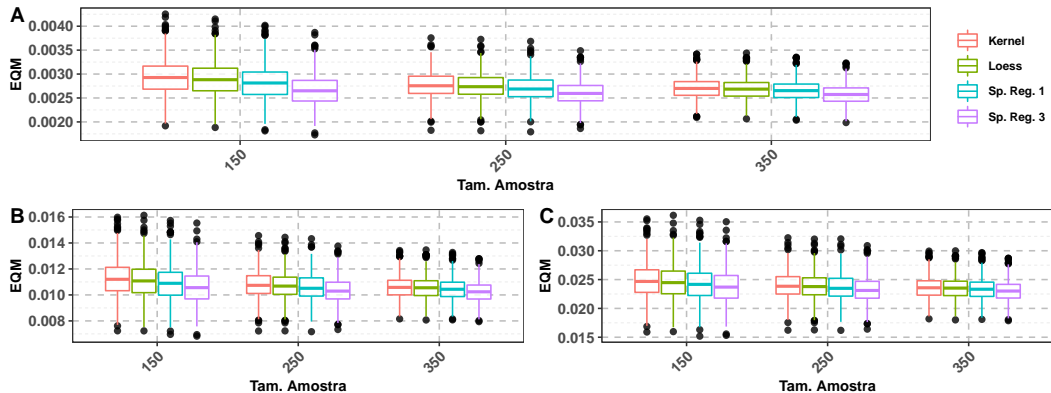


Figura 9: Comparação do erro quadrático para as 1000 amostras para cada suavizador por (A) $DP = 0.05$, (B) $DP = 0.1$ e (C) $DP = 0.15$

Considerações Finais

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 scatterplots.** 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

Referências

- 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).

Fim

Obrigado!

”Sem números, não há vantagem nem probabilidades; sem vantagens e probabilidades, o único meio de lidar com o Risco é apelar para os deuses e o destino. Sem números, o RISCO é uma questão de pura CORAGEM.” (Peter L. Bernstein).