

Modelagem Preditiva

Filipe J. Zabala

Escola Politécnica
PUCRS

2019-11-30

- 1 Minibio
- 2 Sobre modelagem preditiva
- 3 Inferência bayesiana
- 4 Exemplos
- 5 Para saber mais

Filipe J. Zabala · filipe.zabala@pucrs.br

- 2000-2004 Bacharel em Estatística (IME-UFRGS)
- 2006-2009 Mestre em Estatística (IME-USP)
- 2007-2009 Analista do Banco Itaú S.A.
- 2009-2019 Sócio da ZN Consultoria Estatística
- 2010-2019 Professor da Escola Politécnica da PUCRS

- Do Latim *praedicere*, anunciar antecipadamente

- Do Latim *praedicere*, anunciar antecipadamente
- Métodos para prever novos valores de X
 - X : variável de interesse
 - θ : parâmetro associado a X

- Do Latim *praedicere*, anunciar antecipadamente
- Métodos para prever novos valores de X
 - X : variável de interesse
 - θ : parâmetro associado a X
- *As duas culturas* de Leo Breiman (2001):
 - interpretar θ vs prever X

- Do Latim *praedicere*, anunciar antecipadamente
- Métodos para prever novos valores de X
 - X : variável de interesse
 - θ : parâmetro associado a X
- *As duas culturas* de Leo Breiman (2001):
 - interpretar θ vs prever X
- Debabrata Basu (1982): *Information is what information does. It changes opinion (about θ).*

- Do Latim *praedicere*, anunciar antecipadamente
- Métodos para prever novos valores de X
 - X : variável de interesse
 - θ : parâmetro associado a X
- *As duas culturas* de Leo Breiman (2001):
 - interpretar θ vs prever X
- Debabrata Basu (1982): *Information is what information does. It changes opinion (about θ).*
- Bruno de Finetti (1974): *Probability, like beauty, exists only in the mind.*

- Do Latim *praedicere*, anunciar antecipadamente
- Métodos para prever novos valores de X
 - X : variável de interesse
 - θ : parâmetro associado a X
- *As duas culturas* de Leo Breiman (2001):
 - interpretar θ vs prever X
- Debabrata Basu (1982): *Information is what information does. It changes opinion (about θ).*
- Bruno de Finetti (1974): *Probability, like beauty, exists only in the mind.*
- George Box (1979): *All models are wrong but some are useful.*

- Priori: opinião antes de observar os dados

$$\pi(\theta)$$

- Priori: opinião antes de observar os dados

$$\pi(\theta)$$

- Verossimilhança: informação dos dados

$$L(X|\theta)$$

- Priori: opinião antes de observar os dados

$$\pi(\theta)$$

- Verossimilhança: informação dos dados

$$L(X|\theta)$$

- Posteriori: opinião depois de observar os dados

$$\pi(\theta|X)$$

- 'A posteriori de hoje é a priori de amanhã'

- 'A posteriori de hoje é a priori de amanhã'
- Independência condicional (em relação a θ)

- 'A posteriori de hoje é a priori de amanhã'
- Independência condicional (em relação a θ)
- Operação bayesiana: calibrar a opinião à luz dos dados

$$\pi(\theta|X) = \frac{\pi(\theta)L(X|\theta)}{\int_{\theta} \pi(\theta)L(X|\theta)d\theta} \propto \pi(\theta)L(X|\theta)$$

- 'A posteriori de hoje é a priori de amanhã'
- Independência condicional (em relação a θ)
- Operação bayesiana: calibrar a opinião à luz dos dados

$$\pi(\theta|X) = \frac{\pi(\theta)L(X|\theta)}{\int_{\theta} \pi(\theta)L(X|\theta)d\theta} \propto \pi(\theta)L(X|\theta)$$

- Preditiva: distribuição de X

$$Pr(X = x) = \int_{\theta} \pi(\theta)L(X|\theta)d\theta$$

- 'A posteriori de hoje é a priori de amanhã'
- Independência condicional (em relação a θ)
- Operação bayesiana: calibrar a opinião à luz dos dados

$$\pi(\theta|X) = \frac{\pi(\theta)L(X|\theta)}{\int_{\theta} \pi(\theta)L(X|\theta)d\theta} \propto \pi(\theta)L(X|\theta)$$

- Preditiva: distribuição de X

$$Pr(X = x) = \int_{\theta} \pi(\theta)L(X|\theta)d\theta$$

- A probabilidade de o próximo resultado da moeda ser 'cara'

$$Pr(X_{n+1} = \text{cara}) = \frac{r+1}{n+2}$$

- Variáveis permutáveis: a ordem das observações é indiferente

$$Pr(X_1 = x_1, \dots, X_N = x_N) = Pr(X_{\pi(1)} = x_1, \dots, X_{\pi(N)} = x_N)$$

- Variáveis permutáveis: a ordem das observações é indiferente

$$Pr(X_1 = x_1, \dots, X_N = x_N) = Pr(X_{\pi(1)} = x_1, \dots, X_{\pi(N)} = x_N)$$

- Teorema da representação de de Finetti (1931)

$$Pr(X_1 = x_1, \dots, X_N = x_N) = \int_{\theta} \theta^a (1 - \theta)^b \mu(d\theta)$$

- Variáveis permutáveis: a ordem das observações é indiferente

$$Pr(X_1 = x_1, \dots, X_N = x_N) = Pr(X_{\pi(1)} = x_1, \dots, X_{\pi(N)} = x_N)$$

- Teorema da representação de de Finetti (1931)

$$Pr(X_1 = x_1, \dots, X_N = x_N) = \int_{\theta} \theta^a (1 - \theta)^b \mu(d\theta)$$

- Flexibiliza a suposição de independência

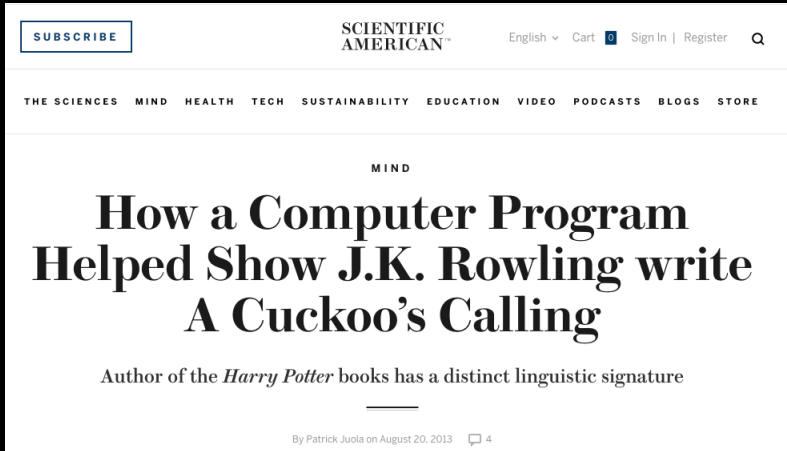
- Variáveis permutáveis: a ordem das observações é indiferente

$$Pr(X_1 = x_1, \dots, X_N = x_N) = Pr(X_{\pi(1)} = x_1, \dots, X_{\pi(N)} = x_N)$$

- Teorema da representação de de Finetti (1931)

$$Pr(X_1 = x_1, \dots, X_N = x_N) = \int_{\theta} \theta^a (1 - \theta)^b \mu(d\theta)$$



- Flexibiliza a suposição de independência
- Trata θ apenas como uma variável de integração



The image shows a screenshot of the Scientific American website. At the top left is a 'SUBSCRIBE' button. The main header features the 'SCIENTIFIC AMERICAN' logo, followed by navigation links for 'English', 'Cart', 'Sign In', and 'Register', and a search icon. Below the header is a horizontal menu with categories: 'THE SCIENCES', 'MIND', 'HEALTH', 'TECH', 'SUSTAINABILITY', 'EDUCATION', 'VIDEO', 'PODCASTS', 'BLOGS', and 'STORE'. The main content area is titled 'MIND' and features a large, bold headline: 'How a Computer Program Helped Show J.K. Rowling write A Cuckoo's Calling'. Below the headline is a sub-headline: 'Author of the *Harry Potter* books has a distinct linguistic signature'. At the bottom of the article preview, it says 'By Patrick Juola on August 20, 2013' and shows a comment icon with the number '4'.

SUBSCRIBE

SCIENTIFIC AMERICAN™


English ▾ Cart  Sign In | Register 

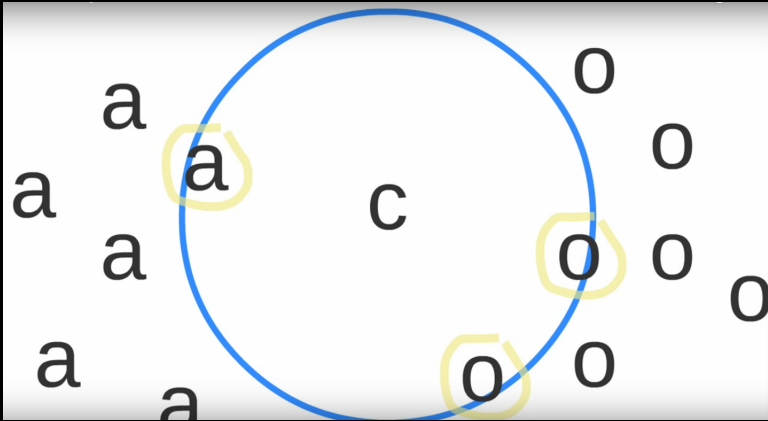
THE SCIENCES MIND HEALTH TECH SUSTAINABILITY EDUCATION VIDEO PODCASTS BLOGS STORE

MIND

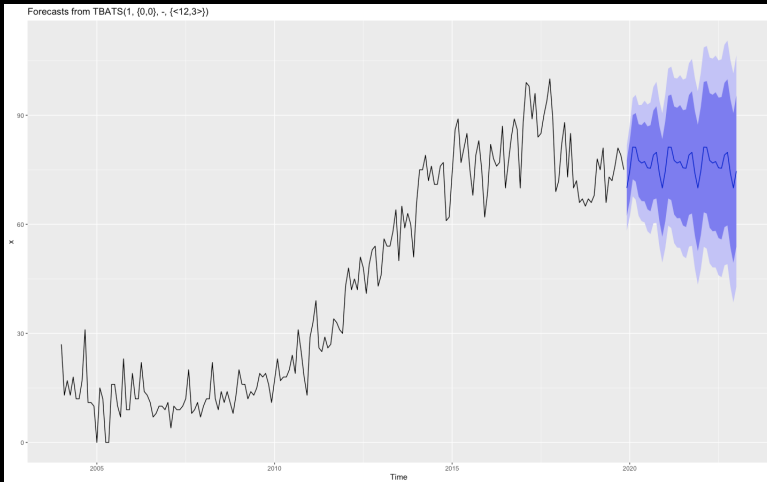
How a Computer Program Helped Show J.K. Rowling write A Cuckoo's Calling

Author of the *Harry Potter* books has a distinct linguistic signature

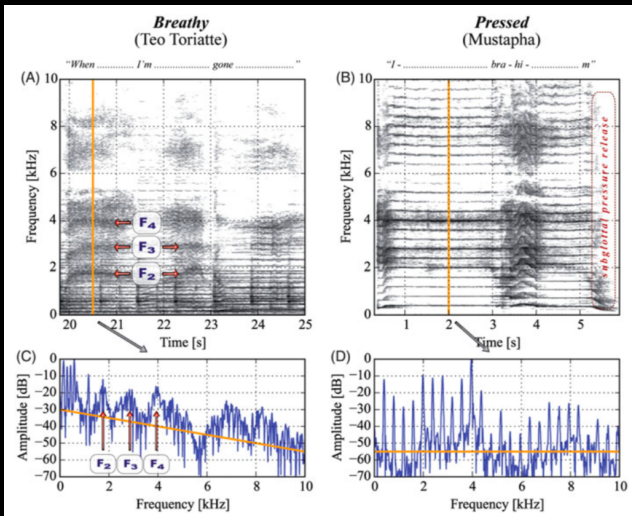
By Patrick Juola on August 20, 2013  4



Exemplo 2: Predição de séries temporais



Exemplo 3: Classificação de falantes



- Zabala (2009): O empate técnico é uma falácia

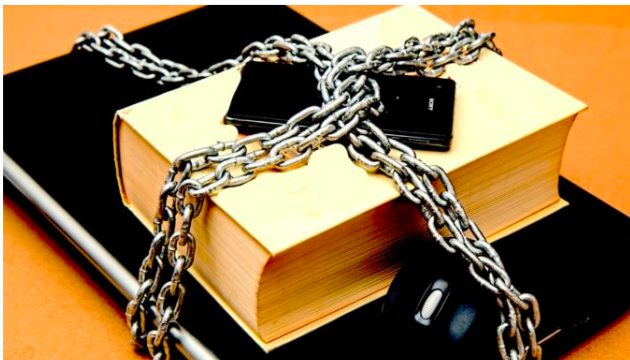
- Zabala (2009): O empate técnico é uma falácia
 - A metodologia dos institutos de pesquisa ignora que a soma dos percentuais deve ser 100%

- Zabala (2009): O empate técnico é uma falácia
 - A metodologia dos institutos de pesquisa ignora que a soma dos percentuais deve ser 100%
 - **Desempate técnico**

- Zabala (2009): O empate técnico é uma falácia
 - A metodologia dos institutos de pesquisa ignora que a soma dos percentuais deve ser 100%
 - [Desempate técnico](#)
 - [Polling Data](#)

France Bans Judge Analytics, 5 Years In Prison For Rule Breakers

🕒 4th June 2019 👤 artificiallawyer 📁 Litigation Prediction 💬 17



- Aitchison & Dunsmore (1975) - Statistical Prediction Analysis
- Seymour Geisser (1993) - Predictive Inference - An Introduction
- Breiman (2001) - Statistical Modeling: The Two Cultures
- Zabala (2009) - Desempate Técnico
- Clarke & Clarke (2018) - Predictive Statistics - Analysis and Inference Beyond Models
- Hyndman & Athanasopoulos (2018) - Forecasting: Principles and Practice
- Izbicki & Santos (2019) - Machine Learning sob a ótica da Estatística