

Bayes

Lendo os dados

V.C.Parro

Primavera - 2020

INSTITUTO MAUÁ DE TECNOLOGIA



Objetivos do módulo

O que preciso conhecer além dos comandos de *machine learning*?

Objetivos

1. **Capacidade de análise** dos principais algoritmos.
2. **Tradução** das equações em pseudo códigos.
3. **Tradução** dos pseudo códigos em códigos Python.
4. Capacidade de **modelagem probabilística**.
5. Aplicação de **testes estatísticos** para avaliação de resultados.
6. **Repertório** de algoritmos e aplicações.

1. **Análise teórica** dos fundamentos.
2. **Codificação** dos algoritmos resultantes.
3. **Visualização** e interpretação de resultados.
4. **Repertório** de conceitos probabilísticos.
5. **Repertório** de testes estatísticos.
6. **Repertório** de métodos de análise.
7. Uso de base de dados **consolidadas e experimentais**.

Metáfora que define o curso

Cenário preliminar:

Suponha que uma pessoa não sabia qual é realmente sua capacidade de poupança. Isso a impede de planejar como conduzir os anos futuros. Tal falta de informação é compensada pela capacidade desta pessoa obter rendimentos variáveis para seu trabalho: precisa mais, trabalha mais, o que depende da demanda - **risco**.

Organização:

A organização da força de trabalho e das despesas e investimentos proporciona a possibilidade de estimativa da capacidade de poupança. Uma conclusão possível é que a poupança "real" não atenda o objetivo desta pessoa quando analisada a luz do desejo de trabalho: redução de risco e estabelecimento expectativa. **Como compatilizar?**

Introdução

Uma decisão era sensata, mesmo que levasse a consequências desastrosas, se as evidências disponíveis indicassem que era a melhor decisão a se tomar; e uma decisão foi tola, mesmo que tenha levado às consequências mais felizes possíveis, se não fosse razoável esperar essas consequências.

Herodotus

Regressão linear

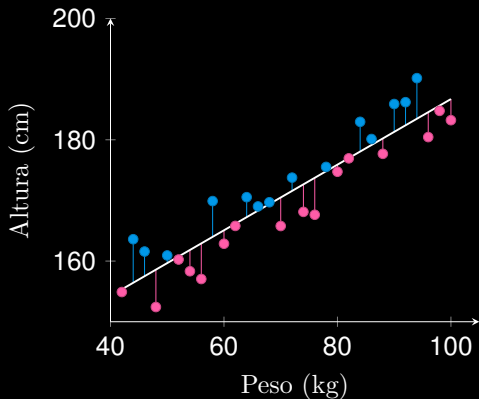


Figura 1: Regressão linear - erros aleatórios e sistemáticos.

Separação

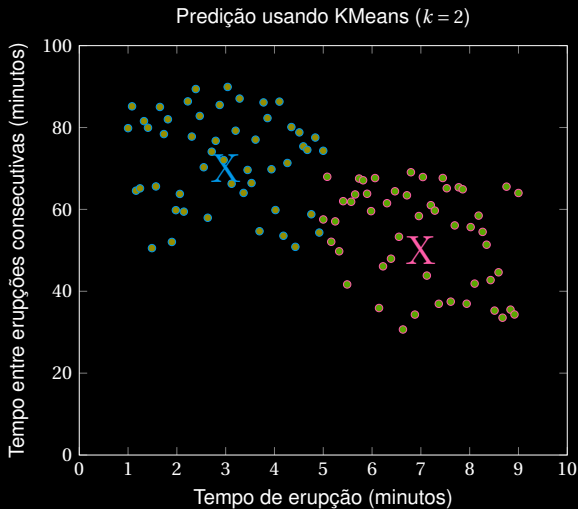


Figura 2: Separação de dados.

A variável tempo - processo estocástico

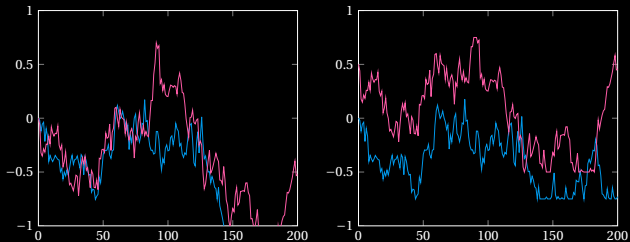


Figura 3: Valores das ações em função do tempo.

Por que agora?

1. 1952 - *Stochastic Gradient Descent*.
2. 1956 - *Perceptron* - pesos ajustáveis.
3. 1986 - *Backpropagation* - multicamadas.
4. 1995 - Redes convolucionais "profundas".

Probabilidade

"...a teoria das probabilidades é basicamente, o senso comum que se reduz ao cálculo: faz-nos apreciar com precisão o que as mentes percebem por um tipo de instinto, sem que elas frequentemente sejam capazes de percebê-lo."

Pierre Simon de Laplace, *Sur les probabilités*

Moedas, dados, urnas, bolas e cartas

1. A Sequência premiada da mega sena.
2. Previsão do tempo.
3. Prefeitos das capitais em 2020.
4. Campeão brasileiro de 2020.
5. Número de infectados com o corona virus em Dezembro de 2020.

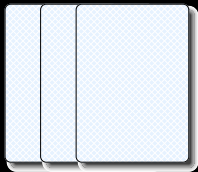


O problema de Monty-hall^a

^aO problema de Monty Hall, também conhecido por paradoxo de Monty Hall é um problema matemático e paradoxo que surgiu a partir de um concurso televisivo dos Estados Unidos chamado Lets Make a Deal, exibido na década de 1970 - <https://towardsdatascience.com/solving-the-monty-hall-problem-with-bayes-theorem-893289953e16>

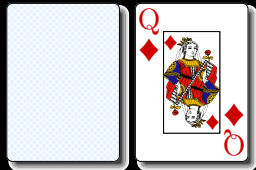
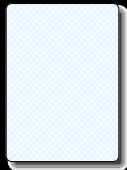
Escolhemos a primeira carta

- Existe apenas um Às entre as três cartas.
- Se virarmos um Às seremos ganhadores, caso contrário perdemos.
- **Qual a nossa chance de ganhar?**

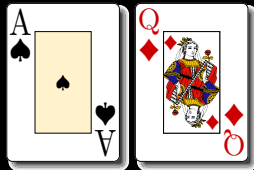
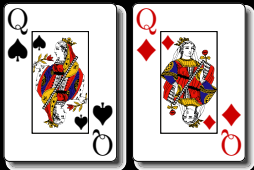
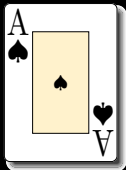


Trocamos?

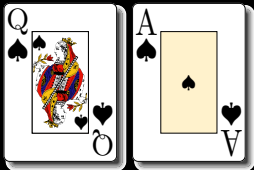
- A mesa vira a última carta.
- A mesa sabe onde está o Às.
- **Qual a nossa chance se trocamos?**



Possibilidades



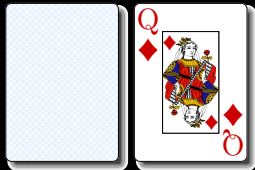
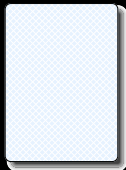
Possibilidades



A mesa sabe onde está o Às.

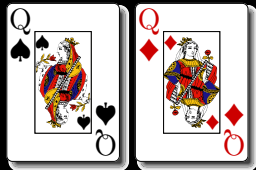
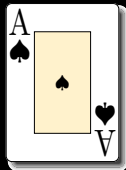
Trocamos?

- A mesa vira a última carta.
- A mesa **não** sabe onde está o Às.
- **Qual a nossa chance se trocamos?**



Sistematizar

- Combinatória.
- Permutas.
- Condicionais.



Linguagem natural^a

^aExemplo do livro:

<https://web.stanford.edu/~jurafsky/slp3/4.pdf>

frase: *predictable with no fun*

**Esta frase transmite uma sensação positiva
ou negativa?**

Probabilidade

Vocabulário

Conjuntos	Probabilidade	Notação
Universo	Espaço amostral	S
Elemento	Resultado	s
Subconjunto	Evento	E
Conjunto nulo	Evento impossível	\emptyset
Conjunto simples	Evento isolado	$E = \{s\}$

Axiomas

1. $P(E) \geq 0 \forall E$
2. $P(S) = 1$
3. $P(E_1 \cup E_2) = P(E_1) + P(E_2)$, considerando eventos mutuamente exclusivos: $E_1 \cap E_2 = \emptyset$.

Probabilidade: eventos discretos e contáveis

$$\begin{aligned} P(E) &= \sum_{(i,j): s_{ij} \in E} \sum P[s_{ij}] \\ &= \frac{N_E}{N_S} \\ &= \frac{\text{Número de resultados no evento } E}{\text{Número total de resultados em } S} \end{aligned}$$

Para o caso de um dado com faces **equiprováveis e dois lançamentos**:

$$P(s_{ij}) = \frac{1}{36}$$

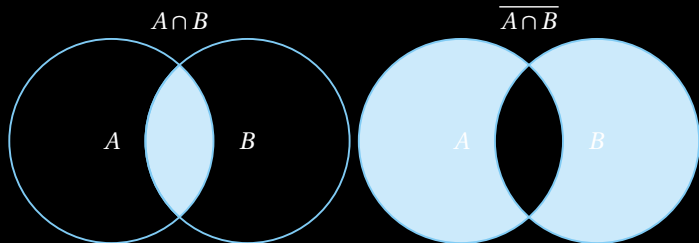


Figura 5: Operações lógicas.

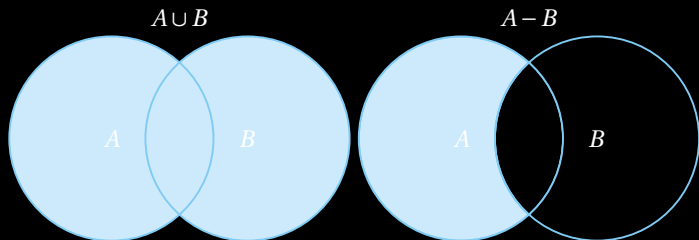


Figura 6: Operações lógicas.

Eventos não mutuamente exclusivos

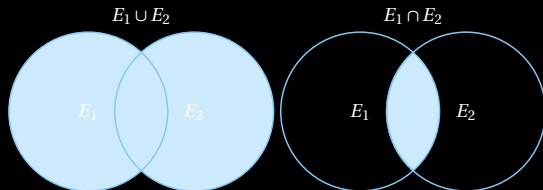


Figura 7: Eventos não mutuamente exclusivos.

$$P(E_1 \cup E_2) = P(E_1) + P(E_2) - P(E_1 \cap E_2) \quad (1)$$

Probabilidade condicional

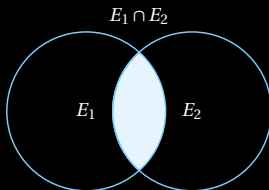


Figura 8: Eventos não mutuamente exclusivos.

$$P(E_1|E_2) = \frac{P(E_1, E_2)}{P(E_2)} \quad (2)$$

No caso das moedas a intersecção é nula. Os eventos são independentes.

Probabilidade condicional

Dos pacientes que chegam na emergência 25% são homens hipertensos.

$$P(Hipertenso|Homem) = \frac{P(Hipertenso, Homem)}{P(Homem)} \quad (3)$$

Qual a probabilidade de um paciente ser hipertenso sabendo-se que é homem?

$$P(E_1, E_2) = P(E_1|E_2)P(E_2) \quad (4)$$

No caso das moedas a intersecção é nula. Os eventos são independentes.

$$P(E_1|E_2) = P(E_1) \quad (5)$$

O caso do dado é um bom exemplo.

Teorema de Bayes

$$P(E_1|E_2) = \frac{P(E_2|E_1)P(E_1)}{p(E_2)} \quad (6)$$

1. $P(E_1|E_2)$ *posterior* ou "a posteriori".
2. $P(E_2|E_1)$ *likelihood* ou verossimilhança.
3. $P(E_1)$ *prior* ou "a priori", "crença" acerca da probabilidade de ocorrência de ocorrer o evento E_1 .
4. $P(E_2)$ *evidence* ou evidência.

$$P(A_S|Q) = \frac{P(Q|A_S)P(A_S)}{p(Q)} \quad (7)$$

Como adaptar para um classificador?

$$P(\textit{Label}|\textit{Feature}) = \frac{P(\textit{Feature}|\textit{Label})P(\textit{Label})}{P(\textit{Feature})} \quad (8)$$

$$P(\textit{Causa}|\textit{Efeito}) = \frac{P(\textit{Efeito}|\textit{Causa})P(\textit{Causa})}{P(\textit{Efeito})} \quad (9)$$

$$P(\textit{Patologia}|\textit{Sintoma}) = \frac{P(\textit{Sintoma}|\textit{Patologia})P(\textit{Patologia})}{P(\textit{Sintoma})} \quad (10)$$

Formalizando o classificador

Como estimar a qual classe um evento está associado?

Para uma determinada categoria c , em um conjunto de categoria \mathcal{C} , desejamos estimar a categoria \hat{c} a qual o documento d pertence.

$$\hat{c} = \max_{c \in \mathcal{C}} P(c/d) \quad (11)$$

Utilizando Bayes:

$$\hat{c} = \max_{c \in \mathcal{C}} \frac{\overbrace{P(d/c)}^{\text{verossimilhana}} \overbrace{P(c)}^{\text{apriori}}}{P(d)} \quad (12)$$

$p(d)$ para o conjunto de categorias \mathcal{C} é constante.

Eventos independentes

O documento d é formado por um conjunto de *features*.

$$d \rightarrow f_1, f_2, f_3, \dots, f_n \quad (13)$$

Utilizando Bayes para um conjunto de "features" \mathcal{F}

$$C_{NB} = \max_{c \in \mathcal{C}} P(c) \prod_{f \in \mathcal{F}} P(f|c) \quad (14)$$

Forma final

O conjunto de "features" \mathcal{F} pode ser representado por um conjunto de palavras \mathcal{W} que representem o que desejamos capturar.

$$C_{NB} = \max_{c \in \mathcal{C}} P(c) \prod_{i \in \mathcal{W}} P(w_i | c) \quad (15)$$

No formato $\log()$:

$$C_{NB} = \max_{c \in \mathcal{C}} \left(\log(P(c)) + \log \left(\sum_{i \in \mathcal{W}} \log(P(w_i | c)) \right) \right) \quad (16)$$

Voltando ao problema inicial

frase: *predictable with no fun*

**Esta frase transmite uma sensação positiva
ou negativa?**

frase: *predictable with no fun*

Categoria	Documentos
N	just plain boring
N	entirely predictable and lacks energy
N	no surprises and very few laughs
P	very powerful
P	the most fun film of the summer

$$\begin{aligned}\hat{P}(c) &= \frac{N_c}{N_D} \\ \hat{P}(w_i/vertc) &= \frac{contagem(w_i, c)}{\sum_{w \in V} contagem(w, c)}\end{aligned}\tag{17}$$

$$\begin{aligned}\hat{P}(c) &= \frac{N_c}{N_D} \\ \hat{P}(w_i/vertc) &= \frac{contagem(w_i, c) + 1}{\sum_{w \in V} contagem(w, c) + 1}\end{aligned}\tag{18}$$

Regressor NB

Como adaptar para um regressor?

$$P(y|\theta_0, \theta_1, \dots, \theta_p) = \frac{P(\theta_0, \theta_1, \dots, \theta_p|y)P(y)}{P(\theta_0, \theta_1, \dots, \theta_p)} \quad (19)$$

considerando eventos independentes:

$$P(y|\theta_0, \theta_1, \dots, \theta_p) = \frac{\overbrace{P(\theta_0|y)}^{\theta \text{ independente}} P(\theta_1|y)(\theta_p|y)P(y)}{P(\theta_0, \theta_1, \dots, \theta_p)} \quad (20)$$

Literatura

*o sinal e o ruído e o
e o ruído e o ruído e
o ruído e o ruído e o
por que tantas e o r
previsões falham e
e outras não e o ruí
ruído e o ruído e o r
e o ruído e o ruído e
nate silver e o ruído
e o ruído e o ruído e
o ruído e intrínseca*

