Aprendizado de máquina

Métodos probabilísticos

V.C.Parro

Março - 2020

INSTITUTO MAUÁ DE TECNOLOGIA

Introdução

Tomada de decisão

Uma decisão era sensata, mesmo que levasse a conseqüências 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

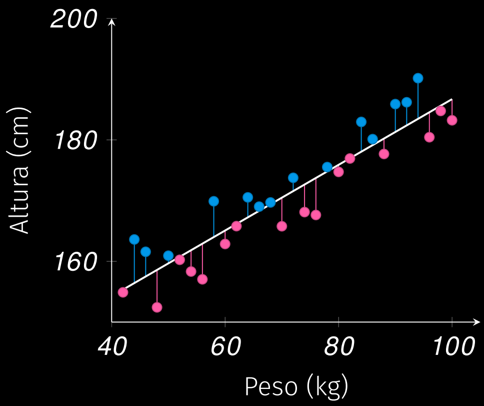


Figura 1: Regressão linear - erros aleatórios e sistemáticos.

Separação

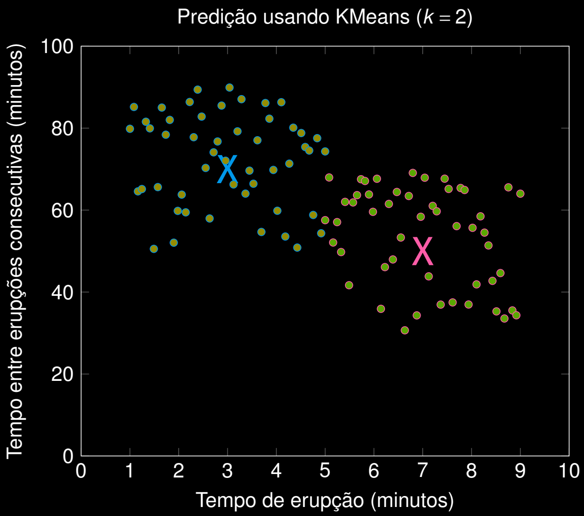


Figura 2: Separação de dados.

A variável tempo - processo estocástico

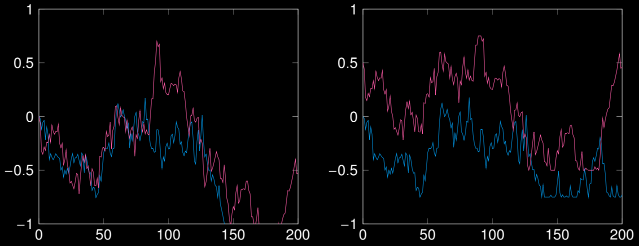


Figura 3: Valores das ações em função do tempo.

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.

Vocabulário

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

Explorando o senso comum

Espaço amostral

S = { ⚀, ⚁, ⚂, ⚃, ⚄, ⚅ }

Viés - lançando o dado seis vezes consecutivas

Evento

E_1 = {⚂, ⚁, ⚄, ⚁, ⚀, ⚃}

Evento

E_3 = {⚅, ⚅, ⚅, ⚅, ⚅, ⚅}

Evento

E_2 = {⚃, ⚀, ⚁, ⚄, ⚁, ⚂}

Evento

E_4 = {⚁, ⚁, ⚂, ⚂, ⚃, ⚃}

Armadilha:o que se pode afirmar acerca das sequências?

Evento

E_1 = {⚂, ⚁, ⚄, ⚁, ⚀, ⚃}

Evento

E_2 = {⚃, ⚂, ⚅, ⚂, ⚁, ⚄}

Evento

E_3 = {⚁, ⚀, ⚃, ⚀, ⚅, ⚂}

Evento

E_4 = {⚃, ⚀, ⚁, ⚄, ⚁, ⚂}

Viés - lançando o dado seis vezes consecutivas

Evento

E_1 = {⚅, ⚅, ⚅, ⚅, ⚅, ⚅}

Evento

E_2 = {⚅, ⚅, ⚅, ⚅, ⚅, ⚅}

Evento

E_3 = {⚅, ⚅, ⚅, ⚅, ⚅, ⚅}

Evento

E_4 = {⚅, ⚅, ⚅, ⚅, ⚅, ⚅}

O que podemos dizer acerca da honestidade deste dado?

⚃ ⚂ ⚁ ⚀ ⚀ ⚅ ⚄ ⚅ ⚅ ⚀ ⚃ ⚃ ⚃ ⚄ ⚁ ⚂ ⚅ ⚅ ⚅ ⚂ ⚄ ⚅

⚂ ⚅ ⚄ ⚀ ⚂ ⚀ ⚄ ⚅ ⚅ ⚄ ⚁ ⚄ ⚄ ⚄ ⚂ ⚀ ⚀ ⚁ ⚁ ⚁ ⚅ ⚅

⚁ ⚂ ⚅ ⚅ ⚅ ⚂ ⚄ ⚀ ⚁ ⚀ ⚂ ⚅ ⚂ ⚁ ⚀ ⚅ ⚅ ⚁ ⚂ ⚅ ⚄ ⚅

⚄ ⚅ ⚃ ⚂ ⚅ ⚂ ⚄ ⚅ ⚅ ⚄ ⚅ ⚅ ⚁ ⚁ ⚀ ⚂ ⚅ ⚅ ⚅ ⚄ ⚅ ⚅

⚅ ⚀ ⚅ ⚅ ⚂ ⚅ ⚅ ⚅ ⚅ ⚅ ⚄ ⚄ ⚀ ⚁ ⚅ ⚅ ⚅ ⚅ ⚃ ⚁ ⚅ ⚅

⚂ ⚂ ⚅ ⚅ ⚅ ⚅ ⚅ ⚅ ⚅ ⚃ ⚁ ⚁ ⚃ ⚂ ⚁ ⚀ ⚅ ⚅ ⚅ ⚅ ⚅ ⚄

⚅ ⚅ ⚅ ⚀ ⚁ ⚂ ⚃ ⚄ ⚅ ⚁ ⚀ ⚂ ⚂ ⚃ ⚄ ⚂ ⚂ ⚅ ⚂ ⚃ ⚃ ⚅

⚄ ⚄ ⚀ ⚃ ⚄ ⚅ ⚅ ⚅ ⚅ ⚀ ⚅ ⚀ ⚃ ⚃ ⚂ ⚄ ⚅ ⚅ ⚁ ⚂ ⚁ ⚀

⚂ ⚀ ⚀ ⚀ ⚁ ⚁ ⚁ ⚅ ⚅ ⚂ ⚅ ⚅ ⚄ ⚂ ⚄ ⚂ ⚅ ⚅ ⚂ ⚄ ⚅ ⚂

⚀ ⚂ ⚁ ⚁ ⚅ ⚀ ⚅ ⚃ ⚅ ⚀ ⚃ ⚀ ⚁ ⚅ ⚂ ⚅ ⚄ ⚅ ⚀ ⚄ ⚂ ⚄

⚅ ⚅ ⚅ ⚁ ⚁ ⚄ ⚅ ⚃ ⚅ ⚃ ⚀ ⚀ ⚀ ⚅ ⚂ ⚄ ⚄ ⚅ ⚄ ⚅ ⚂ ⚂

Formalizando

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 mutualmente exclusivos: $E_1 \cap E_2 = \emptyset$.

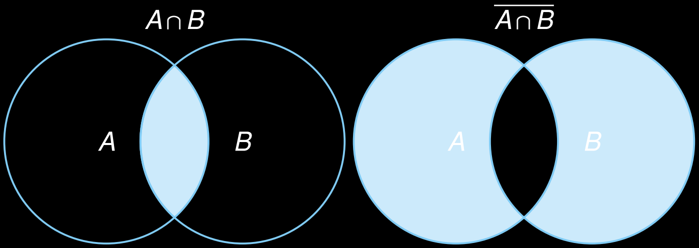


Figura 5: Operações lógicas.

Conjuntos e lógica

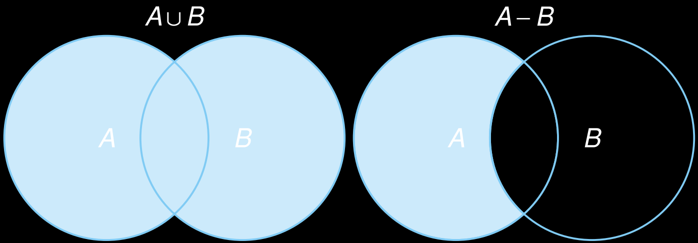


Figura 6: Operações lógicas.

Eventos não mutuamente exclusivos

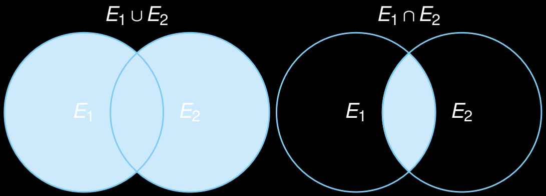


Figura 7: Eventos não mutuamente exclusivos.

$$P[E_1 \cup E_2] = P[E_1] + P[E_2] - P[E_1 \cap E_2] \tag{2}$$

Moedas

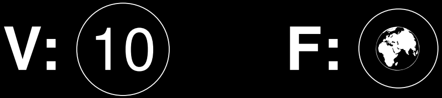


Figura 8: Tabela verdade.

s_1	s_2
V	V
V	F
F	V
F	F

Árvore: moedas

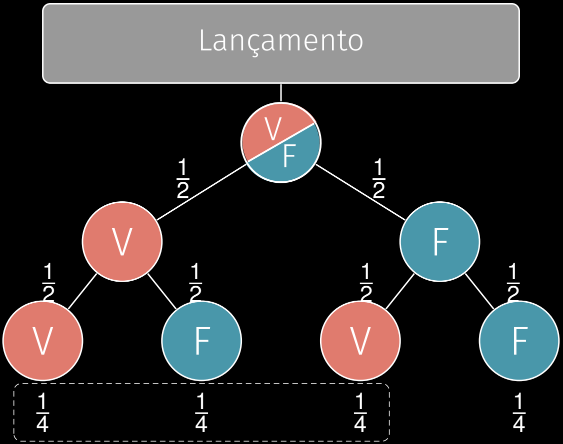


Figura 9: Assumindo que as faces são **equiprováveis**.

Árvore: urnas e bolas

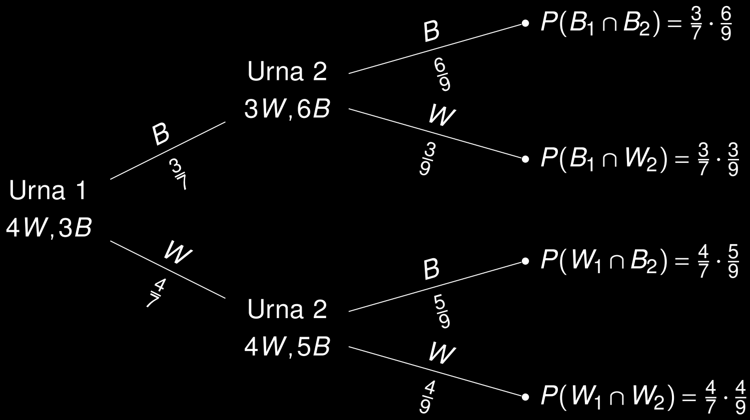


Figura 10: Duas urnas - eventos dependentes.

Árvore: urnas e bolas

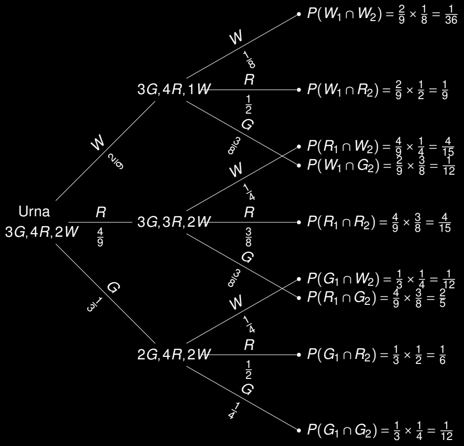


Figura 11: Única urna - eventos dependentes.

Eventos independentes

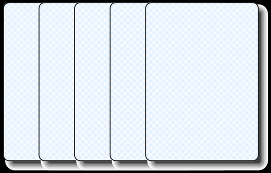
Independência

1. **Para o primeiro caso:** $P(W_2|W_1) = P(W_2)$ - a probabilidade de retirar uma bola branca da segunda urna, sendo que uma branca foi retirada da primeira é igual a probabilidade de retirar uma bola branca da segunda urna.
2. **Para o segundo caso:** $P(W_2|W_1) \neq P(W_2)$ - a probabilidade de retirar uma bola branca da urna, sendo que uma branca foi retirada na primeira tentativa é diferente da probabilidade de retirar uma bola branca na primeira tentativa.

Estudo de caso: cartas

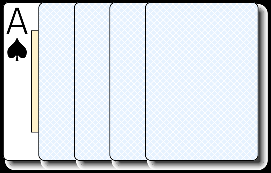
Poker

- Qual a probabilidade de recebermos um Às?
- Qual a probabilidade de recebermos um Às de espadas?



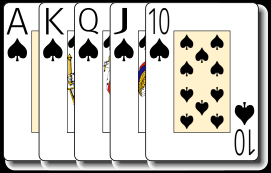
Royal Flush ?

- Qual a probabilidade de obtermos um *Royal Flush*?
- Qual a probabilidade recebermos um novo Ás?



Permutação e combinatória

- Qual a probabilidade de obtermos um *Royal Flush*?
- Qual a probabilidade recebermos um novo Ás?



Espaços discretos e finitos

Espaço amostral: dois lançamentos de dados

$$S = (i, j) : i = 1 \dots 6; j = 1 \dots 6$$

O espaço amostral ($6^2 = 36 \text{número de elementos}$) pode ser determinado como segue:

$$\begin{aligned} S = \{&(1,1),(1,2),\dots,(1,6),\\ &(2,1),(2,2),\dots,(2,6),\\ &\quad\vdots \qquad \vdots \qquad \vdots \qquad \vdots\\ &(6,1),(6,2),\dots,(6,6)\} \end{aligned}$$

Probabilidade: para um determinado evento

$$P[E] = \sum\sum_{(i,j):s_{ij}\in E} P[s_{ij}]$$

$$= \frac{N_E}{N_S}$$

$$= \frac{\text{Número de resultados no evento } \mathbf{E}}{\text{Número total de resultados em } \mathbf{S}}$$

Para o caso **equiprovável:**

$$P[s_{ij}] = \frac{1}{36}$$

Modelagem matemática

Permutação: arranjos possíveis dos resultados.

Para o caso de três "*labels*"que podem acontecer:

l_1	l_2	l_3
A	B	C
A	C	B
B	A	C
B	C	A
C	A	B
C	B	A

Temos um total de 6 permutações possíveis:

$$\text{Permutações} = 3! = 3x2x1 = 6$$

Permutação: generalizando.

Retirada sem reposição:

$$(N)_r = N(N-1)\ldots(N-r+1) = \frac{\mathbf{N}!}{(\mathbf{N}-\mathbf{r})!}$$

considerando r-tuplas

Para o caso do espaço amostral S :

$$(N)_N = N(N-1)\ldots 1 = \mathbf{N}!$$

Combinatória: k arranjos em N elementos.

Retirada sem reposição:

$$\binom{N}{k} = \frac{(N)_k}{k!} = \frac{N!}{(N-k)!k!}$$

Um exemplo: quatro bolas, duas vermelhas e duas pretas

1. Supondo que o sucesso seja atingido quando tenhamos uma bola vermelha **seguida** de uma bola preta.
2. Neste caso temos $\binom{4}{2} = \frac{(4)_2}{2!} = \frac{4!}{(4-2)!2!} = 6 \text{ combinações possíveis}$.
3. Para cada evento temos duas permutações possíveis: $2!$.
4. Desta forma, se as bolas não forem repostas, teremos 12 tuplas de dois elementos no total.

Um exemplo: abordagem exaustiva

Considerando para uma única tentativa T o espaço amostral $S = \{R_1, R_2, B_1, B_2\}$ - duas bolas vermelhas (R_i) e duas pretas ($_Pj$). [1]

T_1	T_2	T_1	T_2	T_1	T_2	T_1	T_2
R_1	B_1	R_1	R_2	B_1	R_1	R_1	R_1
R_1	B_2	R_2	B_1	B_1	R_2	R_2	R_2
R_2	B_1	B_1	B_2	B_2	R_1	B_1	B_1
R_2	B_2	B_2	B_1	B_2	R_2	B_2	B_2

[1] O caso em amarelo representa as possibildaides quando as bolas são repostas.

Generalizando

Considerando que o sucesso é alcançado quando a primeira bola é vermelha e a segunda é preta.

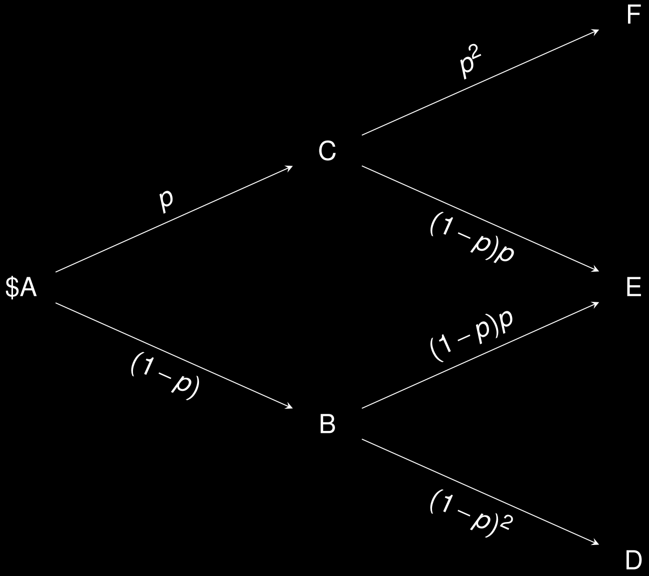
Com reposição (independente):

$$P[E] = \frac{N_E}{N_S} = \frac{k(N-k)}{N^2} = \frac{k}{N}\left(1-\frac{k}{N}\right)$$

Sem reposição:

$$P[E] = \frac{N_E}{N_S} = \frac{k(N-k)}{(N)_2} = \frac{(N-k)}{N}\frac{(k)}{N}\frac{(N)}{(N-1)} = p(1-p)\frac{(N)}{(N-1)}$$

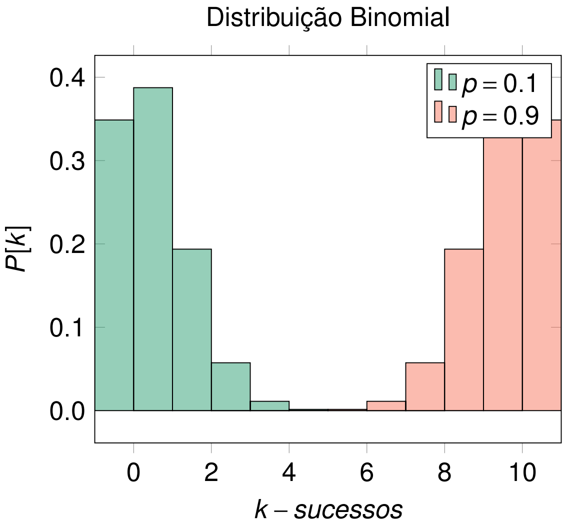
Lançamento de um moeda



Binomial: k sucessos em M tentativas independentes.

$$P[E] = P[X = k] = \binom{N}{k} p^k (1 - p)^k$$

Distribuição binomial



Função massa de probabilidade