## **Naive Bayes**

José Macedo

#### **Aplicações**

- Análise de Sentimentos
- Classificação de Documentos
- Filtros de Spam

#### Motivações:

- Funciona bem a partir de um conjunto razoável de dados
- Trabalha bem com dados categóricos
- Atinge bons resultados em diferentes domínios, além de ser um algoritmo rápido para classificar

## **Monty Hall Problem**







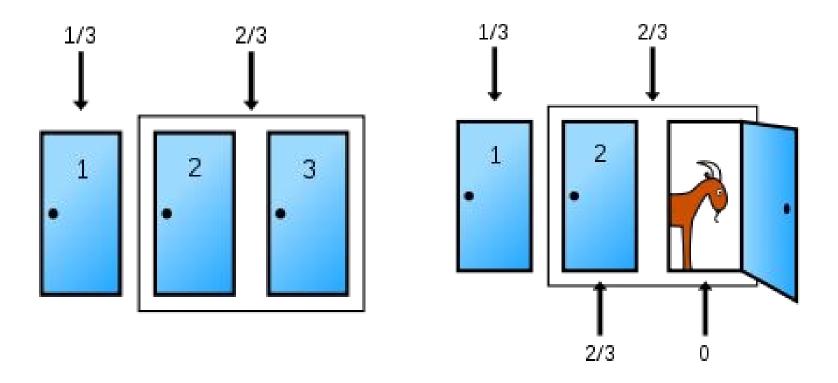
O problema de Monty Hall é um problema matemático que surgiu a partir de um concurso televisivo dos Estados Unidos da América chamado Let's Make a Deal, exibido na década de 1970.

O jogo consiste no seguinte: Monty Hall (o apresentador) apresentava 3 portas aos concorrentes, sabendo que atrás de uma delas está um carro (prêmio bom) e que as outras têm prêmios de pouco valor.

- Na 1ª etapa o concorrente escolhe uma porta (que ainda não é aberta).
- 2 Em seguida, Monty abre uma das outras duas portas que o concorrente não escolheu, sabendo que o carro não se encontra nela.
- 3 Agora, com duas portas apenas para escolher e sabendo que o carro está atrás de uma delas, o concorrente tem que se decidir se permanece com a porta que escolheu no início do jogo e abre-a ou se muda para a outra porta que ainda está fechada para então a abrir.

Neste caso, existe uma estratégia mais lógica? Ficar com a porta escolhida inicialmente ou mudar de porta? Será que com alguma das portas ainda fechadas o concorrente tem mais probabilidades de ganhar? Por que?

#### **Monty Hall Problem**



## **Monty Hall Problem**

Behind door 1	Behind door 2	Behind door 3	Result if staying at door #1	Result if switching to the door offered
Car	Goat	Goat	Wins car	Wins goat
Goat	Car	Goat	Wins goat	Wins car
Goat	Goat	Car	Wins goat	Wins car

#### Teorema de Bayes

O teorema de Bayes pode ser derivado a partir da definição de probabilidade condicional:

$$P(A\mid B)=rac{P(A\cap B)}{P(B)}, ext{ se } P(B)
eq 0,$$

$$P(B \mid A) = rac{P(B \cap A)}{P(A)}, ext{ se } P(A) 
eq 0,$$

pois 
$$P(B \cap A) = P(A \cap B)$$
.

Então,

$$P(A \cap B) = P(A \mid B) P(B) = P(B \mid A) P(A)$$

Logo, ajustando-se os termos, tem-se:

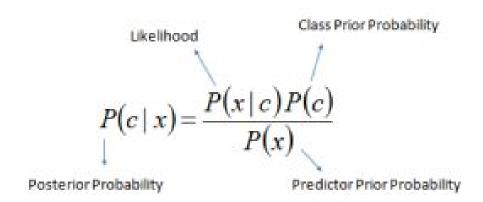
$$P(A\mid B)=rac{P(B\mid A)\,P(A)}{P(B)}, ext{ se } P(B)
eq 0.$$

#### **Teorema Bayes**

$$P(h|D) = \frac{P(D|h)P(h)}{P(D)}$$

- P(h | D) = Probabilidade a posteriori de h dado D
- P(D|h) = Probabilidade a posteriori de D dado h
- o P(h) = Probabilidade a priori de h
- o P(D) = Probabilidade a priori de D

#### Teorema de Bayes



$$P(c \mid X) = P(x_1 \mid c) \times P(x_2 \mid c) \times \cdots \times P(x_n \mid c) \times P(c)$$

#### Justificativa pelo Teorema de Bayes: Consideremos os eventos A<sub>1</sub> = "Carro está na primeira porta", A<sub>2</sub> = "Carro está na segunda porta", A<sub>3</sub> = "Carro está na terceira porta" e • C = "O apresentador abre a terceira porta". Naturalmente, iremos assumir $P(C \mid A_1) = 0.5$ , $P(C \mid A_2) = 1$ e $P(C \mid A_3) = 0$ . Assim, pelo teorema da probabilidade total, temos

$$P(C) = P(C|A_1)P(A_1) + P(C|A_2)P(A_2) + P(C|A_3)P(A_3) =$$

al

$$= \frac{1}{2} \cdot \frac{1}{3} + 1 \cdot \frac{1}{3} + 0 \cdot \frac{1}{3} = \frac{1}{2} = 0.5$$

$$P(C) = P(C|A_1)P(A_1) + P(C)$$

$$= \frac{1}{2} \cdot \frac{1}{2} + 1 \cdot \frac{1}{2} + 0 \cdot \frac{1}{2}$$

$$= \frac{1}{\cdot \cdot \cdot \cdot \cdot + 1 \cdot \cdot \cdot \cdot + 0 \cdot \frac{1}{\cdot \cdot \cdot \cdot \cdot \cdot \cdot + 0}}$$

$$\frac{1}{2} \cdot \frac{1}{3} + 1 \cdot \frac{1}{3} + 0 \cdot \frac{1}{3} = \frac{1}{2} = 0.5$$

$$\cdot \frac{1}{3} + 0 \cdot \frac{1}{3} = \frac{1}{2} = 0.5$$

 $P(A_1 \mid C) = \frac{P(C \mid A_1)P(A_1)}{P(C)} = \frac{\frac{1}{2} \times \frac{1}{3}}{\frac{1}{2}} = \frac{1}{3},$ 

 $P(A_2 \mid C) = \frac{P(C \mid A_2)P(A_2)}{P(C)} = \frac{1 \times \frac{1}{3}}{\frac{1}{2}} = \frac{2}{3}$  e

 $P(A_3 \mid C) = \frac{P(C \mid A_3)P(A_3)}{P(C)} = \frac{0 \times \frac{1}{3}}{\frac{1}{2}} = 0.$ 

$$\frac{1}{3} + 0 \cdot \frac{1}{3} = \frac{1}{2} = 0.5$$

$$+0\cdot\frac{1}{3}=\frac{1}{2}=0.5$$

$$P(A_2) P(A_2) + P(C|A_3) P(A_3) =$$

$$C \mid A_3) = 0$$
. Assim, pelo

$$(A_3) = 0$$
. Assim, pelo

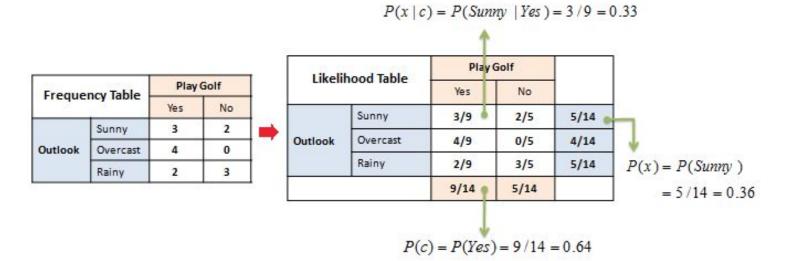
$$A_3$$
) = 0. Assim, pelo

$$|A_3| = 0$$
. Assim, pelo

#### Naive Bayes - Motivação

Dia	Panorama	Temperatura	Umidade	Vento	Jogar Tênis
D1	Ensolarado	Quente	Alta	Fraco	Não
D2	Ensolarado	Quente	Alta	Forte	Não
D3	Nublado	Quente	Alta	Fraco	Sim
D4	Chuvoso	Intermediária	Alta	Fraco	Sim
D5	Chuvoso	Fria	Normal	Fraco	Sim
D6	Chuvoso	Fria	Normal	Forte	Não
D7	Nublado	Fria	Normal	Forte	Sim
D8	Ensolarado	Intermediária	Alta	Fraco	Não
D9	Ensolarado	Fria	Normal	Fraco	Sim
D10	Chuvoso	Intermediária	Normal	Fraco	Sim
D11	Ensolarado	Intermediária	Normal	Forte	Sim
D12	Nublado	Intermediária	Alta	Forte	Sim
D13	Nublado	Quente	Normal	Fraco	Sim
D14	Chuvoso	Intermediária	Alta	Forte	Não

# Qual a probabilidade de Jogar Golf quando estiver fazendo Sol?



Posterior Probability:

 $P(c \mid x) = P(Yes \mid Sunny) = 0.33 \times 0.64 \div 0.36 = 0.60$ 

#### **Naive Bayes - Exemplo**

Qual a probabilidade de Jogar Tênis, dado que a Umidade está Alta?

$$P(Jogar \ T\hat{e}nis = Sim | Umidade = Alta) = \frac{P(Jogar \ T\hat{e}nis = Sim \cap Umidade = Alta)}{P(Umidade = Alta)}$$

$$P(Jogar \ T\hat{e}nis = N\tilde{a}o | Umidade = Alta) = \frac{P(Jogar \ T\hat{e}nis = N\tilde{a}o \cap Umidade = Alta)}{P(Umidade = Alta)}$$

## **Naive Bayes**

• Exemplo:

Continuando...

$$P(Umidade = Alta) = \frac{7}{14} = 0.5$$

$$P(Jogar \ T\hat{e}nis = Sim \cap Umidade = Alta) = \frac{3}{14} = 0.214$$

Dia D1 D2

D5 D6 D7

D9 D10

D11 D12 D13

$$P(Jogar \ T\hat{e}nis = N\tilde{a}o \cap Umidade = Alta) = \frac{4}{14} = 0.286$$

Resultados:

$$P(Jogar \ T\hat{e}nis = Sim|Umidade = Alta) = \frac{\frac{3}{14}}{\frac{7}{14}} = 0.428$$
 
$$P(Jogar \ T\hat{e}nis = N\tilde{a}o|Umidade = Alta) = \frac{\frac{4}{14}}{\frac{7}{14}} = 0.571$$

Panorama	Temperatura	Umidade	Vento	Jogar Tênis
Ensolarado	Quente	Alta	Fraco	Não
Ensolarado	Quente	Alta	Forte	Não
Nublado	Quente	Alta	Fraco	Sim
Chuvoso	Intermediária	Alta	Fraco	Sim
Chuvoso	Fria	Normal	Fraco	Sim
Chuvoso	Fria	Normal	Forte	Não
Nublado	Fria	Normal	Forte	Sim
Ensolarado	Intermediária	Alta	Fraco	Não
Ensolarado	Fria	Normal	Fraco	Sim
Chuvoso	Intermediária	Normal	Fraco	Sim
Ensolarado	Intermediária	Normal	Forte	Sim
Nublado	Intermediária	Alta	Forte	Sim
Nublado	Quente	Normal	Fraco	Sim
Chuvoso	Intermediária	Alta	Forte	Não

#### **Naive Bayes**

Escolhendo a hipótese

$$\begin{aligned} h_{MAP} &= \arg\max_{h \in H} P(h|D) \\ &= \arg\max_{h \in H} \frac{P(D|h)P(h)}{P(D)} \\ &= \arg\max_{h \in H} P(D|h)P(h) \end{aligned}$$

 Observe que P(D) foi descartado, pois é constante

Classificador Naive Bayes

$$v_{MAP} = \underset{v_j \in V}{\operatorname{argmax}} \frac{P(a_1, a_2 \dots a_n | v_j) P(v_j)}{P(a_1, a_2 \dots a_n)}$$
$$= \underset{v_j \in V}{\operatorname{argmax}} P(a_1, a_2 \dots a_n | v_j) P(v_j)$$

$$v_{NB} = \underset{v_j \in V}{\operatorname{argmax}} P(v_j) \prod_i P(a_i | v_j)$$

- a1, a2, ..., aN são as features
- Vj a classe de teste
- Por que Naive? Considerado como ingênuo, pois assume independência entre as features.

## Prática