

MÓDULO 2

Aprendizagem supervisionada

Especialização em Machine Learning

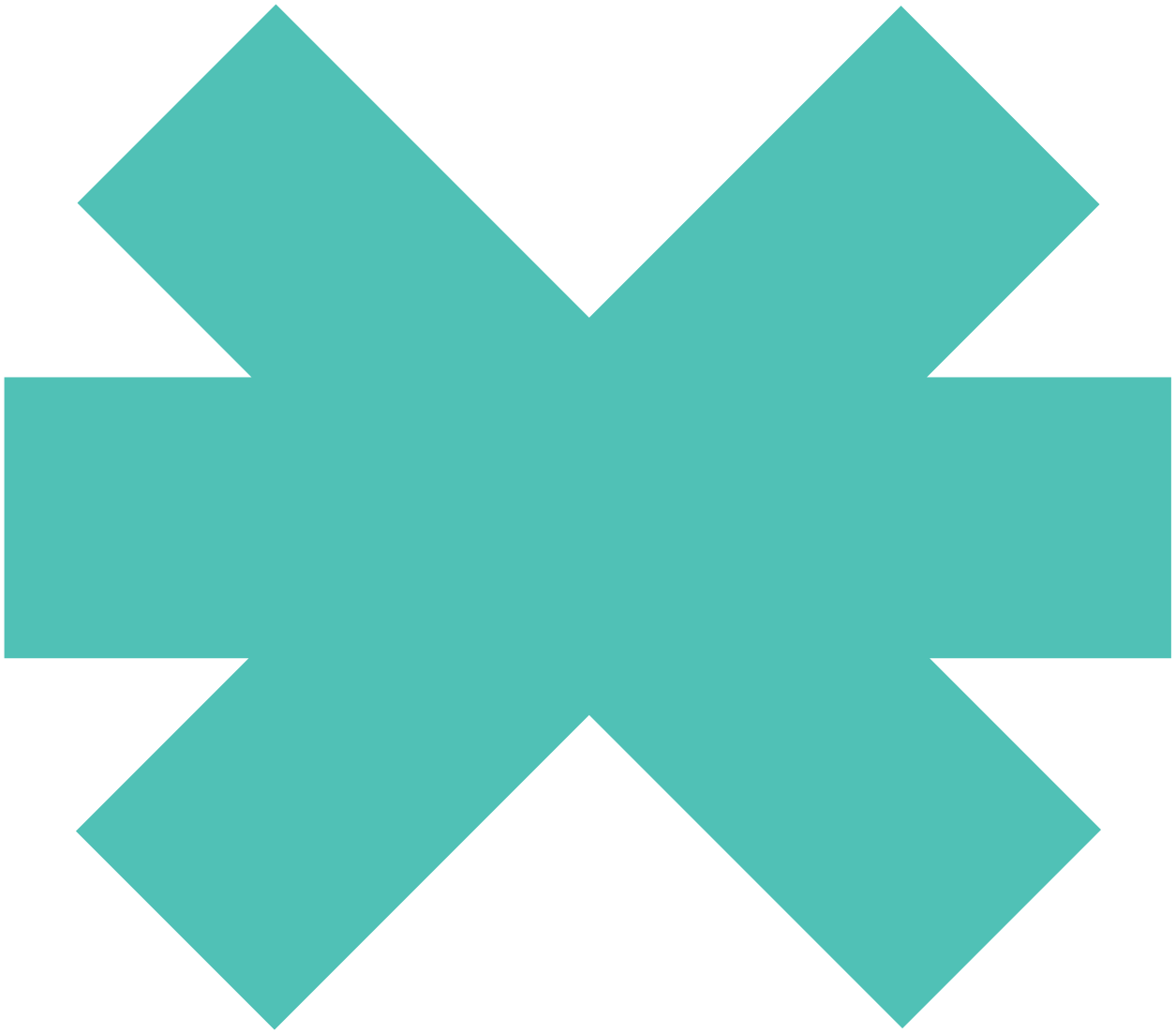
6

# Teorema de Bayes



New  
Technology  
School

Tokio.



# 6 Teorema de Bayes

## Sumário

6.1	Exemplo: classificação de células cancerígenas	05
6.2	Sensibilidade e especificidade	06
6.3	Conclusões	07

O teorema de Bayes permite-nos relacionar a probabilidade de se produzir um evento, sabendo que previamente ocorreu outro e conhecendo a probabilidade de ambos se produzirem em simultâneo.

Expressa, assim, a probabilidade condicional de que se dê um evento aleatório  $A$ , dado outro evento aleatório  $B$ , e previamente conhecida a probabilidade de se produzir  $B$  tendo ocorrido  $A$ .

Assim, o teorema de Bayes permite-nos realizar estimativas ou classificações probabilísticas, tratando-se de um método estatístico não incluído em *machine learning*.

$$P(A_i|B) = \frac{P(B|A_i)P(A_i)}{P(B)}$$
$$P(B) = \sum_{k=1}^n P(B|A_k)P(A_k)$$

Figura 17

Código Latex:

```
P(A_i|B) = \frac{P(B|A_i)P(A_i)}{P(B)} \\
```

```
P(B) = \sum_{k=1}^n P(B|A_k)P(A_k)
```

## 6.1 Exemplo: classificação de células cancerígenas

Neste exemplo queremos classificar amostras de células para detetar se poderiam ser ou não cancerígenas, em função de determinadas provas disponíveis.

Ao avaliar o procedimento, obtemos resultados que parecem promissores; a pergunta é: qual é a utilidade do sistema na deteção do cancro?

$$P(C) = 0,01; P(\bar{C}) = 0,99$$
$$P(+|C) = 0,9; P(+|\bar{C}) = 0,2$$

Figura 18

Código Latex:

$P(C) = 0,01; P(\bar{C}) = 0,99$

$P(+|C) = 0,9; P(+|\bar{C}) = 0,2$

Os resultados desse sistema são:

- $P(C) = 0,01$
- $P(+|C) = 0,9$
- $P(+|\bar{C}) = 0,2$
- $P(+) = 0,207$
- **$P(C|+) = 0,0435$**

Conclusão: o teste apenas aumenta as probabilidades de detetar cancro, até 4.35% desde o 1% inicial.

*Qual é, então, a utilidade do sistema na deteção do cancro?*

---

## 6.2 Sensibilidade e especificidade

A sensibilidade e a especificidade são duas das métricas-chave a ter em consideração na avaliação de um sistema de classificação.

A sensibilidade constitui a capacidade de detetar verdadeiros positivos, por exemplo, classificando corretamente as células cancerígenas como tais e não como não cancerígenas, o que implicaria um falso negativo.

A especificidade é a capacidade de não errar, por exemplo, classificando células não cancerígenas como cancerígenas, o que implicaria um falso positivo.

---

## 6.3 Conclusões

Assim, não podemos considerar bom um sistema ou modelo unicamente porque apresenta uma alta precisão ou sensibilidade, mas precisamos de o avaliar ao pormenor, usando mais de uma métrica.

Da mesma forma, devemos estimar que valor mínimo necessitamos para cada métrica, em função de cada caso. Assim, para classificar células cancerígenas, preferimos ter mais certeza de que um teste é positivo ou de que é negativo? Preferimos errar no que respeita à saúde ou não dar uma má notícia ao paciente, se não tivermos tanta certeza?

