

Aprendizagem automática

Sessão 9 - T

**Modelos baseados em instâncias e
probabilísticos**

2023/2024

Aprendizagem baseada em instâncias

- Os modelos baseados em instâncias são uma classe de algoritmos de aprendizagem automática que **efectuam previsões com base em semelhanças entre instâncias**.
- Ideia: **exemplos semelhantes têm rótulos semelhantes**.
 - O que é que se entende por semelhante?
 - Semelhante significa "próximo" / **valores de características semelhantes**.
- **Algoritmo:**
 - Dado um novo exemplo X para o qual queremos prever o rótulo Y;
 - Encontrar os exemplos de formação mais semelhantes (mais próximos);
 - Prever a etiqueta Y de X com base nas etiquetas dos exemplos mais semelhantes.



Aprendizagem baseada em instâncias

- Perguntas:

- Como determinar a semelhança?
- Que medidas de semelhança utilizar?
- Como é que o modelo aprende?
- Quantos exemplos semelhantes há a considerar?
- Como resolver as incoerências entre os exemplos semelhantes?

Aprendizagem baseada em instâncias

- Perguntas:

- **Como determinar a semelhança?**

- A semelhança é determinada utilizando uma medida de semelhança, que quantifica a proximidade entre instâncias no espaço de características.

- **Que medidas de semelhança utilizar?**

- As medidas comuns incluem as distâncias euclidianas e de Manhattan, a semelhança de cosseno e o coeficiente de correlação de Pearson, dependendo do tipo e das características dos dados.

- **Como é que o modelo aprende?**

- O modelo não "aprende" (lazy-learner), armazena todo o conjunto de dados de treino e faz previsões com base nas semelhanças entre as novas instâncias e os exemplos armazenados.

Aprendizagem baseada em instâncias

- Perguntas:

- **Quantos exemplos semelhantes há a considerar?**
 - É um parâmetro que precisa de ser afinado com base nas características específicas do conjunto de dados e do problema em causa.
- **Como resolver as incoerências entre os exemplos semelhantes?**
 - Para problemas de classificação, o rótulo previsto é determinado por votação maioritária.
 - Para problemas de regressão, o rótulo previsto é determinado pela média (ou média ponderada) dos valores-alvo dos exemplos semelhantes.

Aprendizagem baseada em instâncias

- **Vantagens:**

- **Flexibilidade:** Os modelos baseados em instâncias podem lidar com relações complexas e não-linearidades nos dados.
- **Sem formação de modelos:** Estes modelos não requerem uma formação explícita tornando-as fáceis de implementar e atualizar.
- **Eficiência computacional:** Embora o tempo de previsão possa ser lento com grandes conjuntos de dados, a fase de formação é normalmente rápida, uma vez que não há formação explícita de modelos envolvida.
- **Interpretável:** As previsões podem muitas vezes ser explicadas através da análise das instâncias mais próximas nos dados de treino.

Aprendizagem baseada em instâncias

- **Limitações:**

- **Complexidade computacional:** As previsões podem ser lentas, especialmente com grandes conjuntos de dados, uma vez que envolvem o cálculo de distâncias entre a nova instância e todas as instâncias de treino.
- **Sensibilidade ao ruído:** Os modelos baseados em instâncias podem ser sensíveis a ruídos ou características irrelevantes no conjunto de dados.
- **Requisitos de memória:** O armazenamento de todo o conjunto de dados de treino pode ser impraticável para conjuntos de dados muito grandes.

Aprendizagem baseada em instâncias

- **Melhores práticas:**

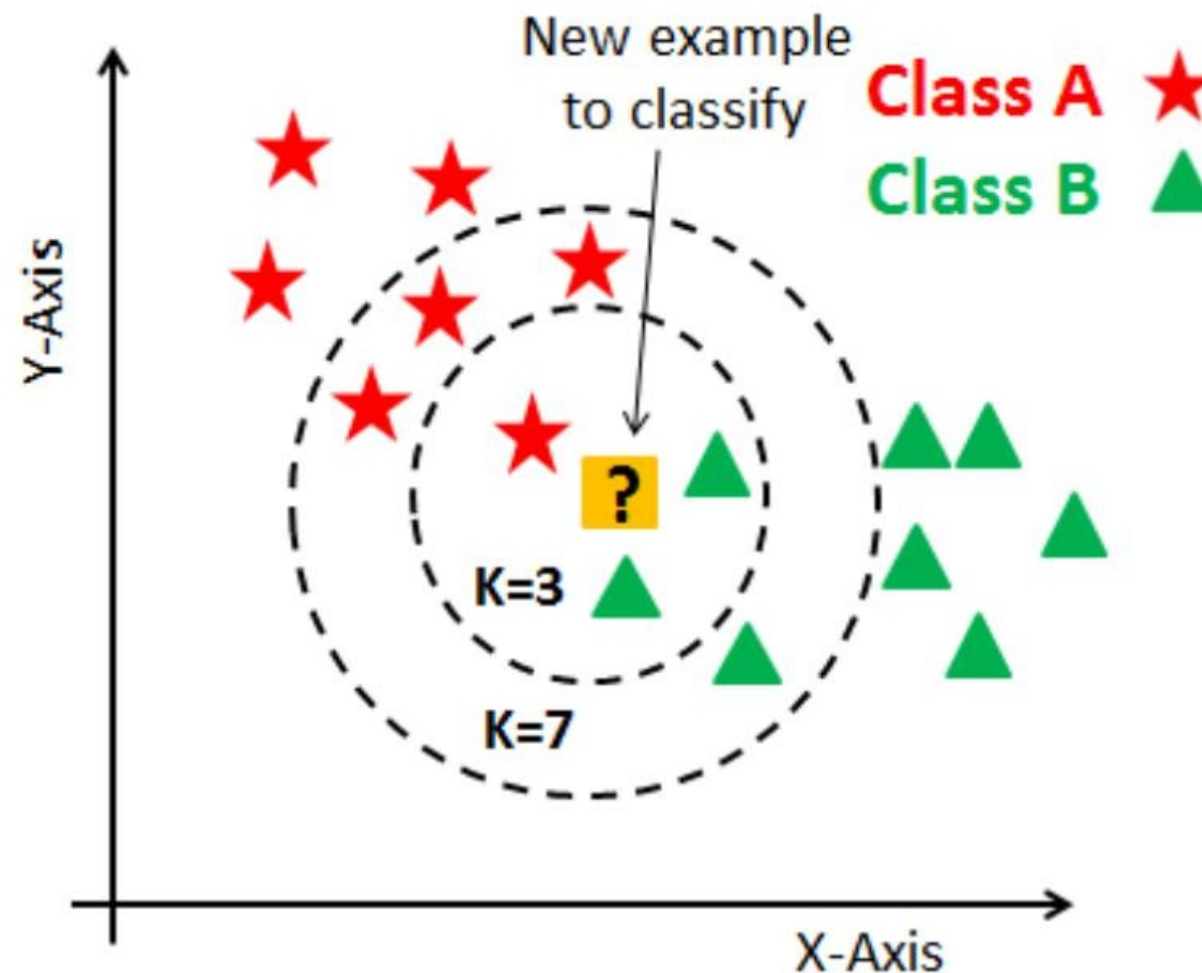
- **Dimensionamento de elementos:** O dimensionamento das características é importante para garantir que todas as contribuem igualmente para o cálculo da distância.
- **Validação cruzada:** Avaliar o desempenho do modelo utilizando técnicas como a validação cruzada k-fold para escolher hiperparâmetros óptimos (por exemplo, número de exemplos semelhantes a utilizar) e avaliar o desempenho da generalização.
- **Tratamento de dados desequilibrados:** Resolver os desequilíbrios de classe ajustando o ponderação das instâncias.

K-Nearest Neighbors (KNN)

- O KNN é um algoritmo de aprendizagem simples baseado em instâncias, utilizado para tarefas de classificação e regressão.
- Na fase de treino, **armazena todo o conjunto de dados de treino**;
- Durante a previsão, **calcula a distância** entre os pontos de dados de entrada e todos os exemplos de treino utilizando uma métrica de distância (por exemplo, a distância euclidiana).
- O algoritmo identifica os K vizinhos mais próximos dos pontos de dados de entrada e utiliza as suas etiquetas para efetuar previsões. Utiliza a **votação por maioria** para a classificação e o **cálculo da média** para a regressão.

K-Nearest Neighbors (KNN)

- Como classificarias o novo exemplo?
 - Com **k=3**?
 - Com **k=7**?

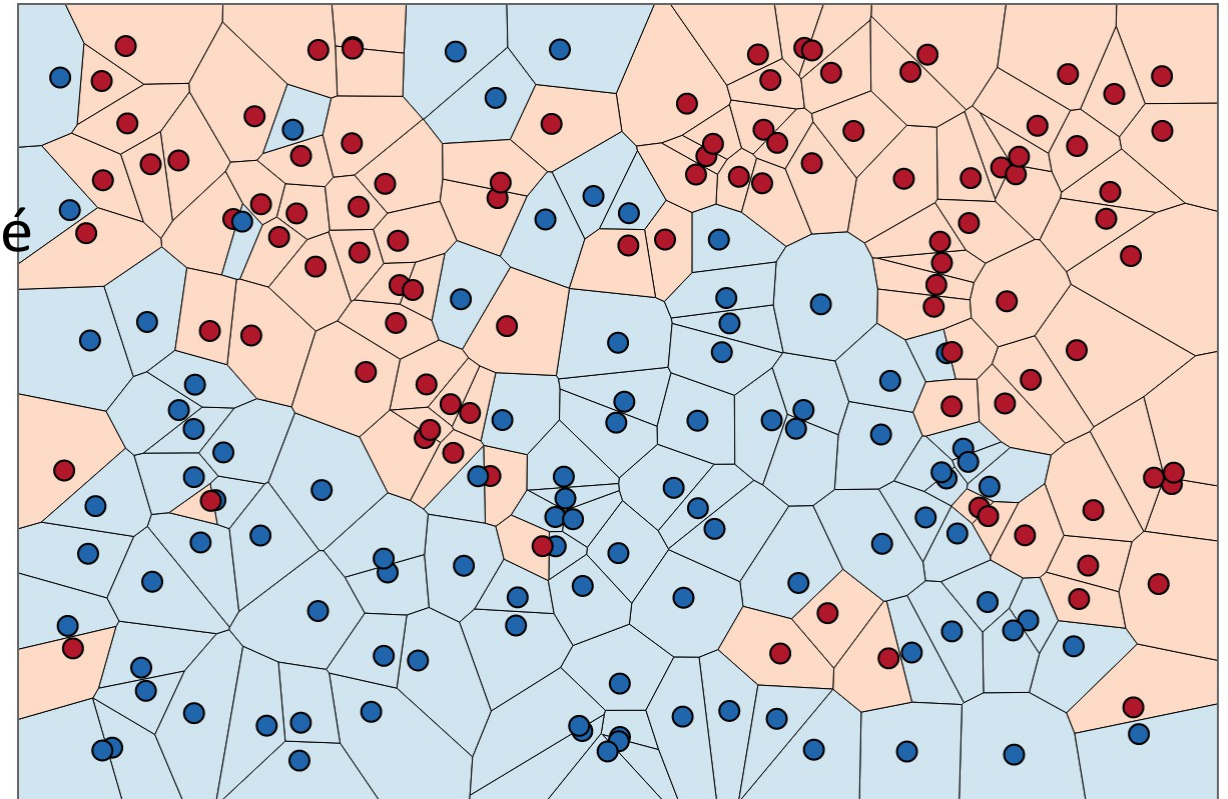


KNN - Limites de decisão

- O algoritmo do vizinho mais próximo não calcula diretamente os limites de decisão; no entanto, estes podem ser inferidos a partir dos dados de treino.

- **Diagrama de Voronoi:**

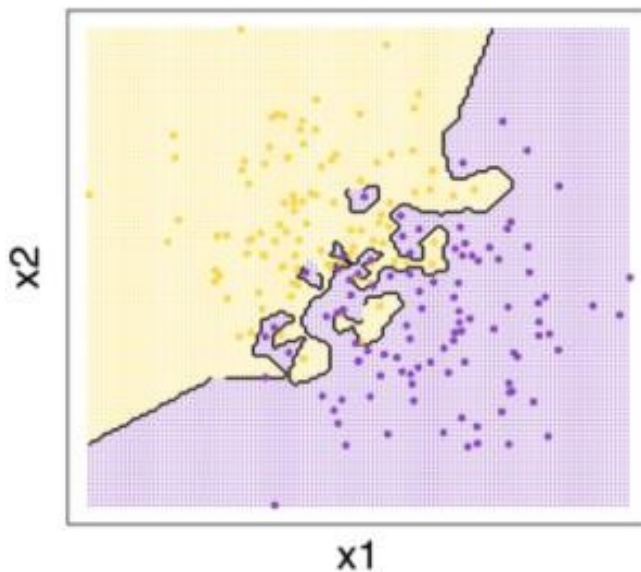
- Mostrar como o espaço de entrada é dividido em classes
- Cada linha é equidistante a pontos de diferentes classes



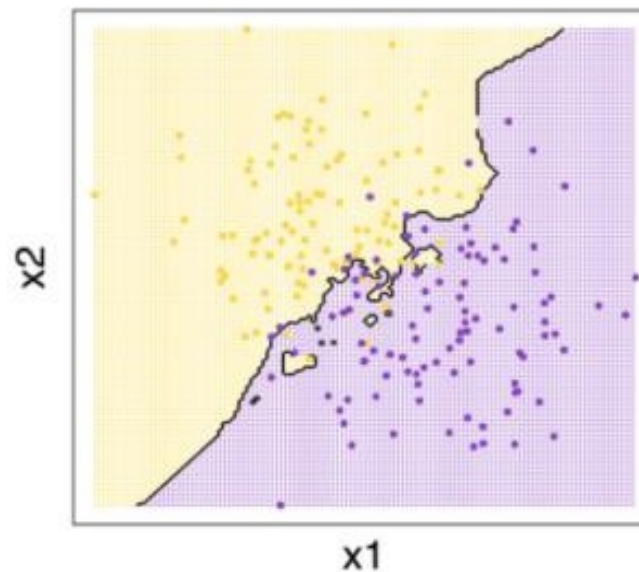
KNN - Limites de decisão

- Impacto de k

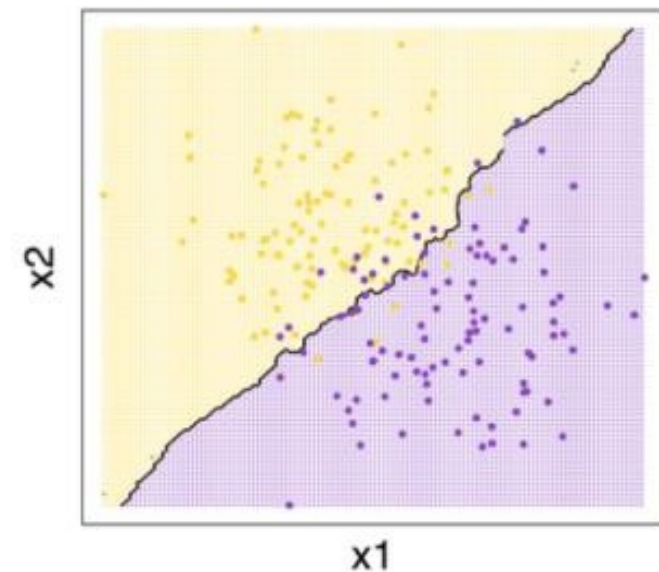
Binary kNN Classification ($k=1$)



Binary kNN Classification ($k=5$)



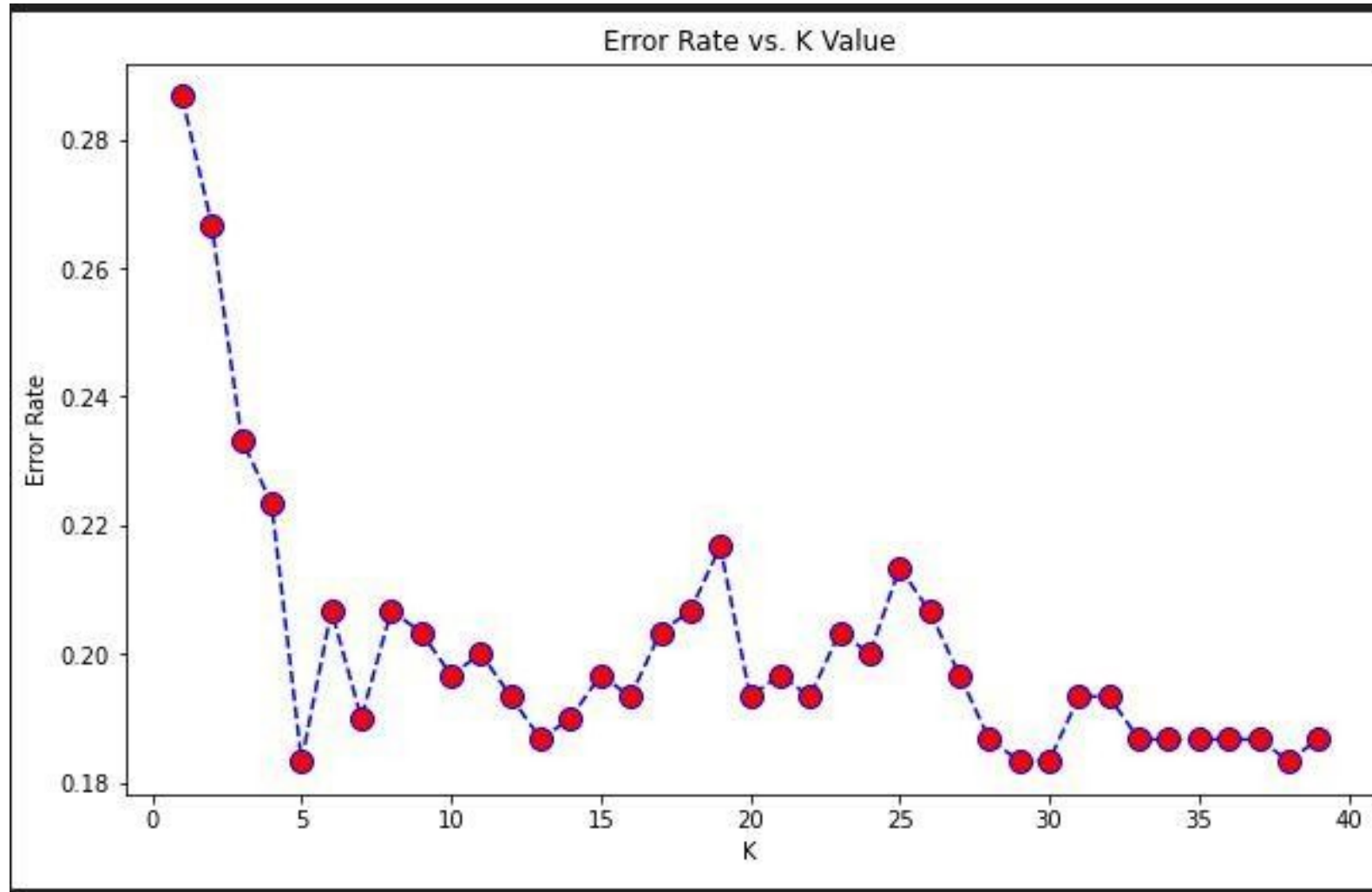
Binary kNN Classification ($k=25$)



KNN - Como escolher k?

- Um **k maior** pode **melhorar potencialmente o desempenho**.
- No entanto, a definição de k excessivamente grande pode implicar a consideração de amostras que **não** são **verdadeiros vizinhos**, levando a uma diminuição da exatidão.
- Os métodos de estimativa de erros (como o holdout e a validação cruzada) podem ajudar a encontrar o **k ótimo**.
- É comum utilizar $k = \sqrt{n}$, onde n é o número de exemplos de treino.

KNN - Limites de decisão



Modelos probabilísticos

- Os modelos probabilísticos são quadros matemáticos utilizados para representar **a incerteza**.
- Estes modelos têm por objetivo captar as **distribuições de probabilidade** subjacentes aos **dados**.
- **O Naive Bayes** é um classificador probabilístico baseado no **teorema de Bayes**. Assume que as características são condicionalmente independentes, dada a etiqueta da classe.

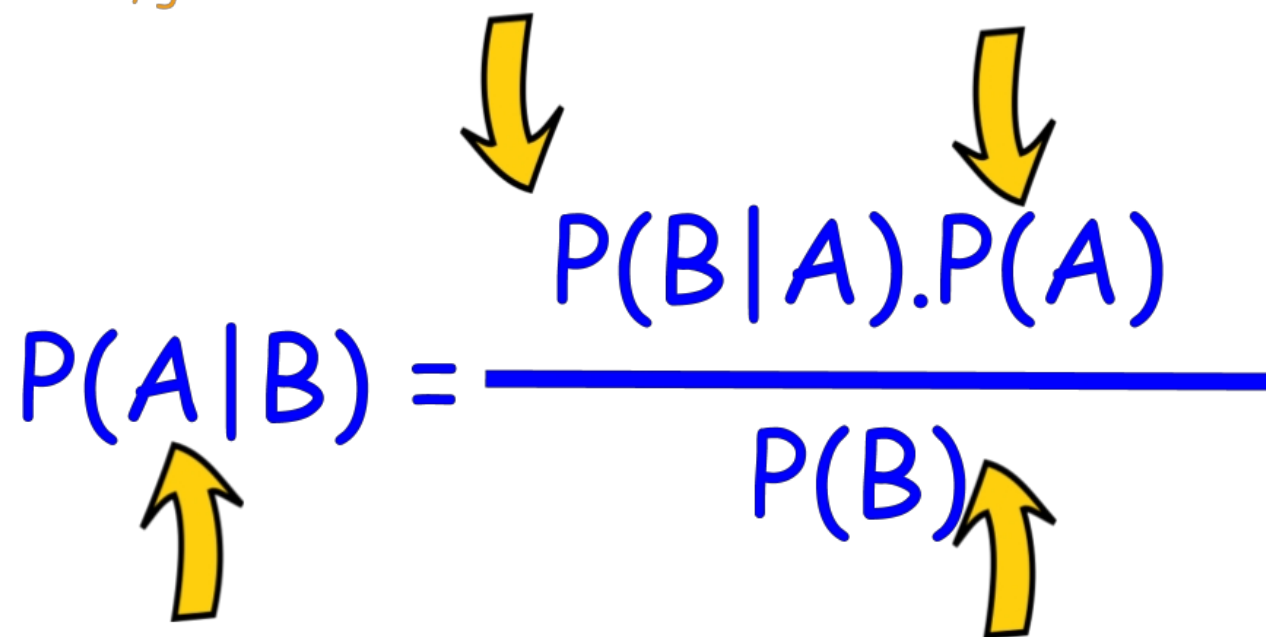
Teorema de Bayes

LIKELIHOOD

The probability of "B" being True, given "A" is True

PRIOR

The probability "A" being True. This is the knowledge.


$$P(A|B) = \frac{P(B|A).P(A)}{P(B)}$$

POSTERIOR

The probability of "A" being True, given "B" is True

MARGINALIZATION

The probability "B" being True.

Naive- Bayes

- Baseado em **probabilidades condicionais** (Teorema de Bayes);
- Calcula as probabilidades associadas à pertença de um exemplo a cada classe possível;
- **Pressupostos** (que raramente se verificam na realidade):
 - Todas as características têm a mesma importância;
 - Os valores das várias características ocorrem de forma independente.

Naive- Bayes

- Tipos de Naive Bayes:
 - **Multinomial Naive Bayes:** Adequado para classificação com características discretas.
 - **Gaussian Naive Bayes:** assume que as características seguem uma distribuição normal.
 - **Bernoulli Naive Bayes:** funciona com características binárias.
- Vantagens:
 - **Simplicidade:** Fácil de compreender e implementar.
 - **Eficiência:** Computacionalmente eficiente, especialmente com dados de elevada dimensão.
 - **Robustez:** Tem um bom desempenho mesmo com pequenos conjuntos de dados e na presença de dados irrelevantes características.
- Limitações:
 - **Pressuposto de independência:** O pressuposto "ingénuo" da independência das características pode não se manter em conjuntos de dados do mundo real, conduzindo potencialmente a classificações incorrectas.
 - **Sensível à qualidade dos dados de entrada:** O Naive Bayes pode ser sensível a

Naive- Bayes

características irrelevantes ou a características com elevada correlação, o que pode degradar o desempenho da classificação.



UNIVERSIDADE
CATOLICA
PORTUGUESA

BRAGA

Multinomial Naive-Bayes

- Passo 1: **Obter os dados**

Outlook	Humidity	Wind	Run
Sunny	High	Weak	No
Overcast	High	Strong	No
Rainy	High	Weak	Yes
Rainy	Normal	Weak	No
Sunny	Normal	Weak	Yes
Sunny	High	Weak	Yes
Sunny	High	Weak	Yes
Rainy	Normal	Strong	No
Overcast	High	Weak	Yes
Sunny	High	Weak	Yes
Rainy	High	Weak	No
Overcast	Normal	Strong	No
Overcast	High	Weak	Yes
Sunny	High	Weak	Yes

Multinomial Naive-Bayes

- Passo 2: Converter os dados em tabelas de frequência

Frequency Table		Run	
		Yes	No
Outlook	Sunny	5	1
	Overcast	2	2
	Rainy	1	3

Frequency Table		Run	
		Yes	No
Humidity	High	7	3
	Normal	1	3

Frequency Table		Run	
		Yes	No
Wind	Strong	0	3
	Weak	9	2

Multinomial Naive-Bayes

- Passo 3: Calcular a probabilidade prévia e a verosimilhança

Likelihood Table		Run		
		Yes	No	
Outlook	Sunny	5/8	1/6	6/14
	Overcast	2/8	2/6	4/14
	Rainy	1/8	3/6	4/14
		8/14	6/14	

Likelihood Table		Run		
		Yes	No	
Humidity	High	7/8	3/6	10/14
	Normal	1/8	3/6	4/14
		8/14	6/14	

Likelihood Table		Run		
		Yes	No	
Wind	Strong	0/9	3/5	3/14
	Weak	9/9	2/5	11/14
		9/14	5/14	

Multinomial Naive-Bayes

- Passo 4: **Aplicar o Teorema de Bayes**
- Digamos que quer concentrar-se na probabilidade de ir correr, uma vez que está sol lá fora.

Likelihood Table		Run		
		Yes	No	
Outlook	Sunny	5/8	1/6	6/14
	Overcast	2/8	2/6	4/14
	Rainy	1/8	3/6	4/14
		8/14	6/14	

$P(\text{Sunny} | \text{Yes}) = 5/8 = 0.625$
 $P(\text{Sunny}) = 6/14 = 0.428$
 $P(\text{Yes}) = 8/14 = 0.571$

- $P(\text{Sim} | \text{Sol}) = P(\text{Sol} | \text{Sim}) * P(\text{Sim}) / P(\text{Sol}) = 0,625 * 0,571 / 0,428 = 0,834$

Multinomial Naive-Bayes

Likelihood Table		Run		
		Yes	No	
Outlook	Sunny	5/8	1/6	6/14
	Overcast	2/8	2/6	4/14
	Rainy	1/8	3/6	4/14
		8/14	6/14	

Likelihood Table		Run		
		Yes	No	
Humidity	High	7/8	3/6	10/14
	Normal	1/8	3/6	4/14
		8/14	6/14	

Cuidado com as probabilidades de 0. Geralmente, é adicionada uma constante a todas as contagens

Likelihood Table		Run		
		Yes	No	
Wind	Strong	0/9	3/5	3/14
	Weak	9/9	2/5	11/14
		9/14	5/14	

- Perspectivas: Chuvoso
- Humidade: Normal
- Vento: fraco
- **Executar: ?**

$$\text{Bayes Theorem} \longrightarrow P(A | x_1, \dots, x_n) = \frac{P(x_1, \dots, x_n | A) P(A)}{P(x_1, \dots, x_n)}$$

$$\text{Naïve Bayes} \longrightarrow P(A | x_1, \dots, x_n) = P(x_1 | A) \cdot P(x_2 | A) \cdot P(x_i | A) P(A)$$

Podemos eliminar o denominador da fórmula, assumindo a independência das características

- $P(\text{Sim} | \text{Chuvoso}, \text{Normal}, \text{Fraco}) = P(\text{Chuvoso} | \text{Sim}) * P(\text{Normal} | \text{Sim}) * P(\text{Fraco} | \text{Sim}) * P(\text{Sim})$
 $= 1/8 \quad * 1/8 \quad * 9/9 * 8/14 = 0.0089$
- $P(\text{Não} | \text{Chuvoso}, \text{Normal}, \text{Fraco}) = P(\text{Chuvoso} | \text{Não}) * P(\text{Normal} | \text{Não}) * P(\text{Fraco} | \text{Não}) * P(\text{Não})$
 $= 3/6 \quad * 3/6 \quad * 2/5 * 6/14 = 0.042$

$$P(\text{Sim}) = 0,0089 / (0,0089 + 0,042) = 0,175$$

$$P(\text{Não}) = 0,042 / (0,0089 + 0,042) = 0,825$$

Multinomial Naive-Bayes

Likelihood Table		Run		
		Yes	No	
Outlook	Sunny	5/8	1/6	6/14
	Overcast	2/8	2/6	4/14
	Rainy	1/8	3/6	4/14
		8/14	6/14	

Likelihood Table		Run		
		Yes	No	
Humidity	High	7/8	3/6	10/14
	Normal	1/8	3/6	4/14
		8/14	6/14	

Likelihood Table		Run		
		Yes	No	
Wind	Strong	0/9	3/5	3/14
	Weak	9/9	2/5	11/14
		9/14	5/14	

$$\text{Bayes Theorem} \longrightarrow P(A | x_1, \dots, x_n) = \frac{P(x_1, \dots, x_n | A) P(A)}{P(x_1, \dots, x_n)}$$



$$\text{Naïve Bayes} \longrightarrow P(A | x_1, \dots, x_n) = P(x_1 | A) \cdot P(x_2 | A) \cdot P(x_i | A) P(A)$$

- Perspectivas: Sol
- Humidade: elevada
- Vento: fraco
- **Executar: ?**

Multinomial Naive-Bayes

Likelihood Table		Run		
		Yes	No	
Outlook	Sunny	5/8	1/6	6/14
	Overcast	2/8	2/6	4/14
	Rainy	1/8	3/6	4/14
		8/14	6/14	

Likelihood Table		Run		
		Yes	No	
Humidity	High	7/8	3/6	10/14
	Normal	1/8	3/6	4/14
		8/14	6/14	

Likelihood Table		Run		
		Yes	No	
Wind	Strong	0/9	3/5	3/14
	Weak	9/9	2/5	11/14
		9/14	5/14	

- Perspectivas: Sol
- Humidade: elevada
- Vento: fraco
- **Executar: ?**

$$\text{Bayes Theorem} \longrightarrow P(A | x_1, \dots, x_n) = \frac{P(x_1, \dots, x_n | A) P(A)}{P(x_1, \dots, x_n)}$$

$$P(\text{Não}) = 0,0143 / (0,3125 + 0,0143) = 0,044$$

$$\text{Naïve Bayes} \longrightarrow P(A | x_1, \dots, x_n) = P(x_1 | A) \cdot P(x_2 | A) \cdot P(x_i | A) P(A)$$

- $P(\text{Sim} | \text{Sol}, \text{Alta}, \text{Frac}) = P(\text{Sol} | \text{Sim}) * P(\text{Alta} | \text{Sim}) * P(\text{Frac} | \text{Sim}) * P(\text{Sim})$
 $= 5/8 \quad * 7/8 \quad * 9/9 \quad 8/14 = 0.3125$
- $P(\text{Não} | \text{Sol}, \text{Alta}, \text{Frac}) = P(\text{Sol} | \text{Não}) * P(\text{Alta} | \text{Não}) * P(\text{Frac} | \text{Não}) * P(\text{Não})$
 $= 1/6 \quad * 3/6 \quad * 2/5 \quad 6/14 = 0.0143$

$$P(\text{Sim}) = 0,3125 / (0,3125 + 0,0143) = 0,956$$

Multinomial Naive Bayes

Podemos eliminar o fenômeno de overfitting assumindo a independência das características



UNIVERSIDADE
CATOLICA
PORTUGUESA

BRAGA

- Webb, G. I. (2011). Naïve Bayes. Em Encyclopedia of Machine Learning (pp. 713-714). Springer US. https://doi.org/10.1007/978-0-387-30164-8_576
- Kramer, O. (2013). K-Nearest Neighbors. Em Dimensionality Reduction with Unsupervised Nearest Neighbors (Redução de dimensionalidade com vizinhos mais próximos não supervisionados) (pp. 13-23). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-38652-7_2
- <https://www.youtube.com/watch?v=HVXime0nQel>
- <https://www.youtube.com/watch?v=O2L2Uv9pdDA>