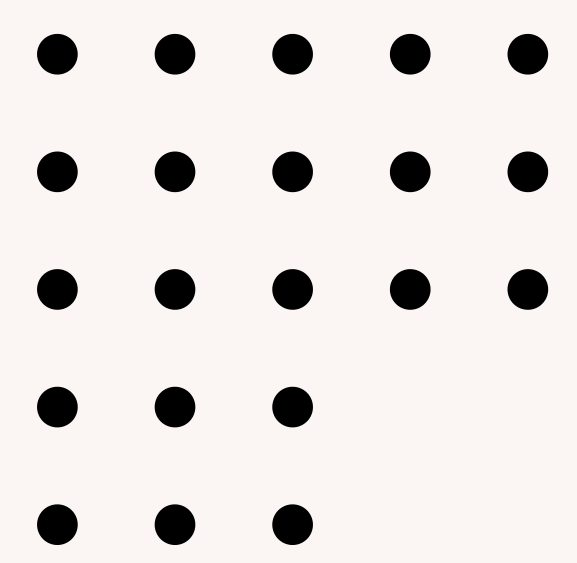
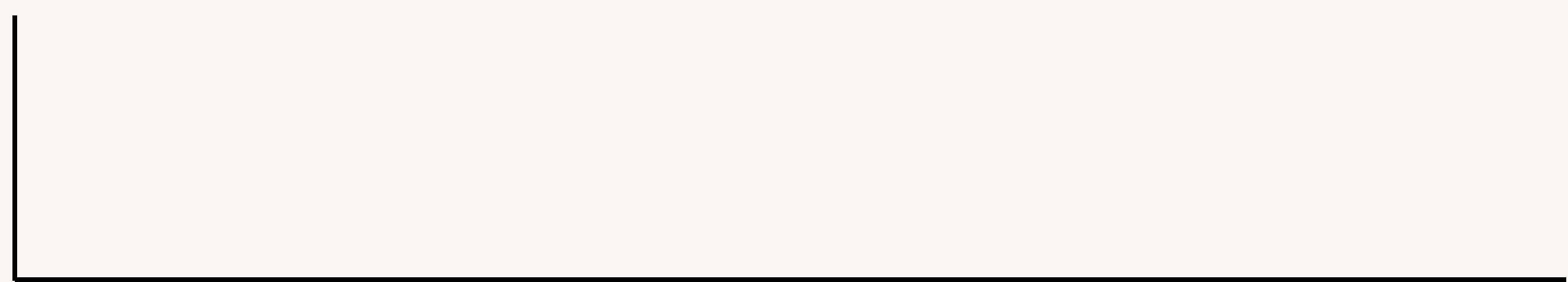
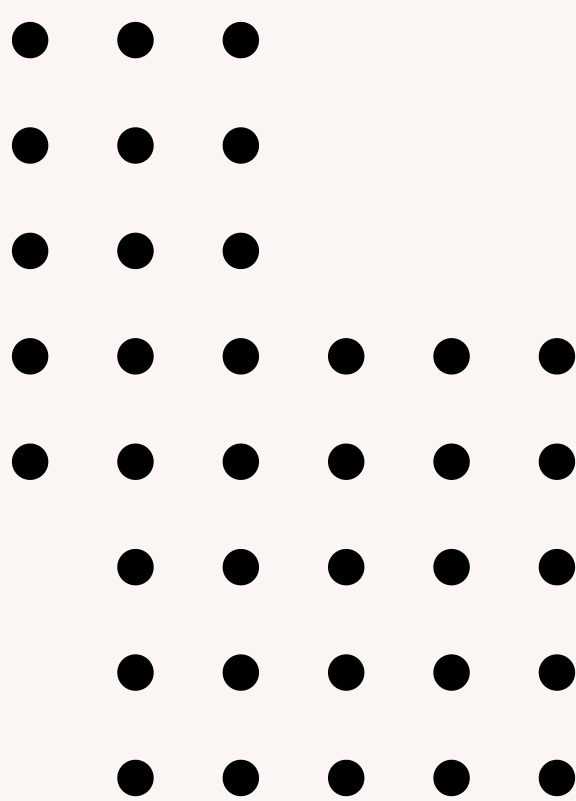


NAIVE BAYES



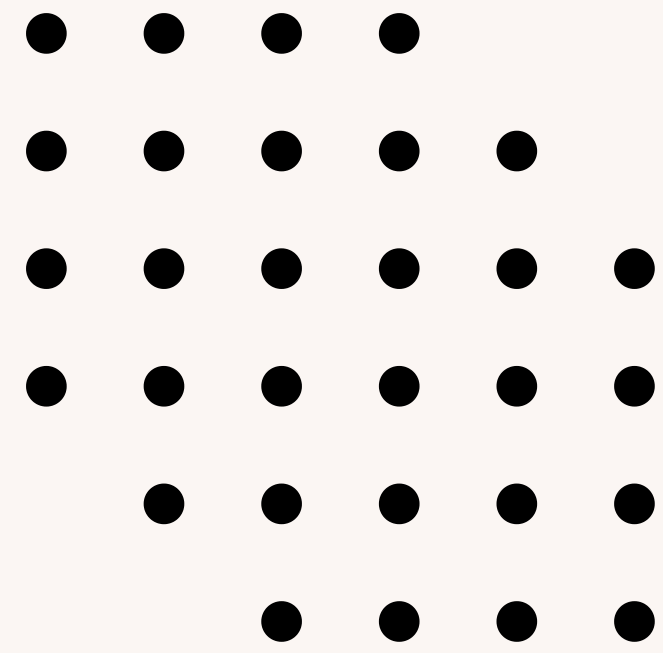


O QUE É?

- O Naive Bayes é um classificador probabilístico baseado no Teorema de Bayes.
 - Ele assume que a presença de uma determinada característica em uma classe não está relacionada com a presença de outras características.
 - Essa suposição ingênua é chamada de "naive" (ingênuo) e é a base para o nome do algoritmo.
- 
-

TEOREMA DE BAYES

- O Teorema de Bayes é uma fórmula que descreve como atualizar a probabilidade de uma hipótese em função de novas evidências.
- A fórmula é: $P(A|B) = (P(B|A) * P(A)) / P(B)$
- O Naive Bayes utiliza o Teorema de Bayes para calcular a probabilidade de uma instância pertencer a uma determinada classe.





FUNCIONAMENTO PASSO A PASSO

1. Coleta de dados: O primeiro passo é obter um conjunto de dados rotulados para treinar o classificador.
2. Treinamento: O algoritmo calcula as probabilidades a priori de cada classe e as probabilidades condicionais de cada atributo dado cada classe.
3. Preparação dos dados de teste: Os dados de teste são preparados para o classificador, garantindo que estejam no mesmo formato do conjunto de treinamento.
4. Cálculo das probabilidades: O classificador usa as probabilidades calculadas durante o treinamento para calcular a probabilidade de cada classe para os dados de teste.
5. Classificação: Com base nas probabilidades calculadas, o Naive Bayes classifica a instância de teste atribuindo-a à classe com a maior probabilidade.

1 . COLETA DE DADOS

CALAFRIOS	CORIZA	CEFALEIA	FEBRE	GRIPE
sim	não	média	sim	NÃO
sim	sim	não	não	SIM
sim	não	forte	sim	SIM
não	sim	média	sim	SIM
não	não	não	não	NÃO
não	sim	forte	sim	SIM
não	sim	forte	não	NÃO
sim	sim	média	sim	SIM

2. TREINAMENTO

- Calculamos a probabilidade de cada classe aparecer no conjunto de treinamento:

$$P(\text{Gripe}=\text{Sim}) = 5/8 = 0.625$$

$$P(\text{Gripe}=\text{Não}) = 3/8 = 0.375$$

- Para cada atributo, calculamos a probabilidade condicional de cada valor dado cada classe.

Calafrios:

$$P(\text{Calafrios}=\text{Sim} \mid \text{Gripe}=\text{Sim}) = 3/5 = 0.6$$

$$P(\text{Calafrios}=\text{Não} \mid \text{Gripe}=\text{Sim}) = 2/5 = 0.4$$

$$P(\text{Calafrios}=\text{Sim} \mid \text{Gripe}=\text{Não}) = 1/3 = 0.33$$

$$P(\text{Calafrios}=\text{Não} \mid \text{Gripe}=\text{Não}) = 2/3 = 0.66$$

Coriza:

$$P(\text{Coriza}=\text{Sim} \mid \text{Gripe}=\text{Sim}) = 4/5 = 0.8$$

$$P(\text{Coriza}=\text{Não} \mid \text{Gripe}=\text{Sim}) = 1/5 = 0.2$$

$$P(\text{Coriza}=\text{Sim} \mid \text{Gripe}=\text{Não}) = 1/3 = 0.33$$

$$P(\text{Coriza}=\text{Não} \mid \text{Gripe}=\text{Não}) = 2/3 = 0.66$$

Cefaleia:

$$P(\text{Cefaleia=Média} \mid \text{Gripe=Sim}) = 2/5 = 0.4$$

$$P(\text{Cefaleia=Forte} \mid \text{Gripe=Sim}) = 2/5 = 0.4$$

$$P(\text{Cefaleia=Não} \mid \text{Gripe=Sim}) = 1/5 = 0.1$$

$$P(\text{Cefaleia=Média} \mid \text{Gripe=Não}) = 1/3 = 0.33$$

$$P(\text{Cefaleia=Forte} \mid \text{Gripe=Não}) = 1/3 = 0.33$$

$$P(\text{Cefaleia=Não} \mid \text{Gripe=Não}) = 1/3 = 0.33$$

Febre:

$$P(\text{Febre=Sim} \mid \text{Gripe=Sim}) = 4/5 = 0.8$$

$$P(\text{Febre=Não} \mid \text{Gripe=Sim}) = 1/5 = 0.2$$

$$P(\text{Febre=Sim} \mid \text{Gripe=Não}) = 1/3 = 0.33$$

$$P(\text{Febre=Não} \mid \text{Gripe=Não}) = 2/3 = 0.66$$

3 . PREPARAÇÃO DOS DADOS DE TESTE

- Agora, vamos supor que queremos classificar se um paciente está gripado dado os sintomas a baixo. Vamos usar o classificador Naive Bayes para fazer esse cálculo.

CALAFRIOS	CORIZA	CEFALEIA	FEBRE	GRIPE
sim	não	média	não	?

4. CÁLCULO DAS PROBABILIDADES:

Precisamos calcular $P(\text{Gripe}=\text{Sim} \mid x)$ e $P(\text{Gripe}=\text{Não} \mid x)$

$x = (\text{Calafrios}=\text{sim}, \text{Coriza}=\text{não}, \text{Cefaleia}=\text{média}, \text{Febre}=\text{não})$

$$P(\text{Gripe}=\text{Sim} \mid x) = P(x \mid \text{Gripe}=\text{sim}) * P(\text{Gripe}=\text{sim}) / p(x)$$

$$P(\text{Gripe}=\text{não} \mid x) = P(x \mid \text{Gripe}=\text{não}) * P(\text{Gripe}=\text{não}) / p(x)$$

- $P(x \mid \text{Gripe}=\text{sim}) = P(\text{Calafrios}=\text{sim} \mid \text{Gripe}=\text{sim}) * P(\text{Coriza}=\text{não} \mid \text{Gripe}=\text{sim}) * P(\text{Cefaleia}=\text{média} \mid \text{Gripe}=\text{sim}) * P(\text{Febre}=\text{não} \mid \text{Gripe}=\text{sim}) = 0.6 * 0.2 * 0.4 * 0.2 = 0.0096$
- $P(x \mid \text{Gripe}=\text{não}) = P(\text{Calafrios}=\text{sim} \mid \text{Gripe}=\text{não}) * P(\text{Coriza}=\text{não} \mid \text{Gripe}=\text{não}) * P(\text{Cefaleia}=\text{média} \mid \text{Gripe}=\text{não}) * P(\text{Febre}=\text{não} \mid \text{Gripe}=\text{não}) = 0.33 * 0.66 * 0.33 * 0.66 = 0.0474$
- $P(x) = P(x \mid \text{Gripe}=\text{sim}) * P(\text{Gripe}=\text{sim}) + P(x \mid \text{Gripe}=\text{não}) * P(\text{Gripe}=\text{não}) = 0.0237$

Calculando as probabilidades finais....

$$\begin{aligned}P(\text{Gripe}=\text{Sim} \mid x) &= P(x \mid \text{Gripe}=\text{sim}) * P(\text{Gripe}=\text{sim}) / p(x) \\&= 0.0096 * 0.625 / 0.0237 \\&= 0.25 = 25\%\end{aligned}$$

$$\begin{aligned}P(\text{Gripe}=\text{n\~ao} \mid x) &= P(x \mid \text{Gripe}=\text{n\~ao}) * P(\text{Gripe}=\text{n\~ao}) / p(x) \\&= 0.0474 * 0.375 / 0.0237 \\&= 0.75 = 75\%\end{aligned}$$

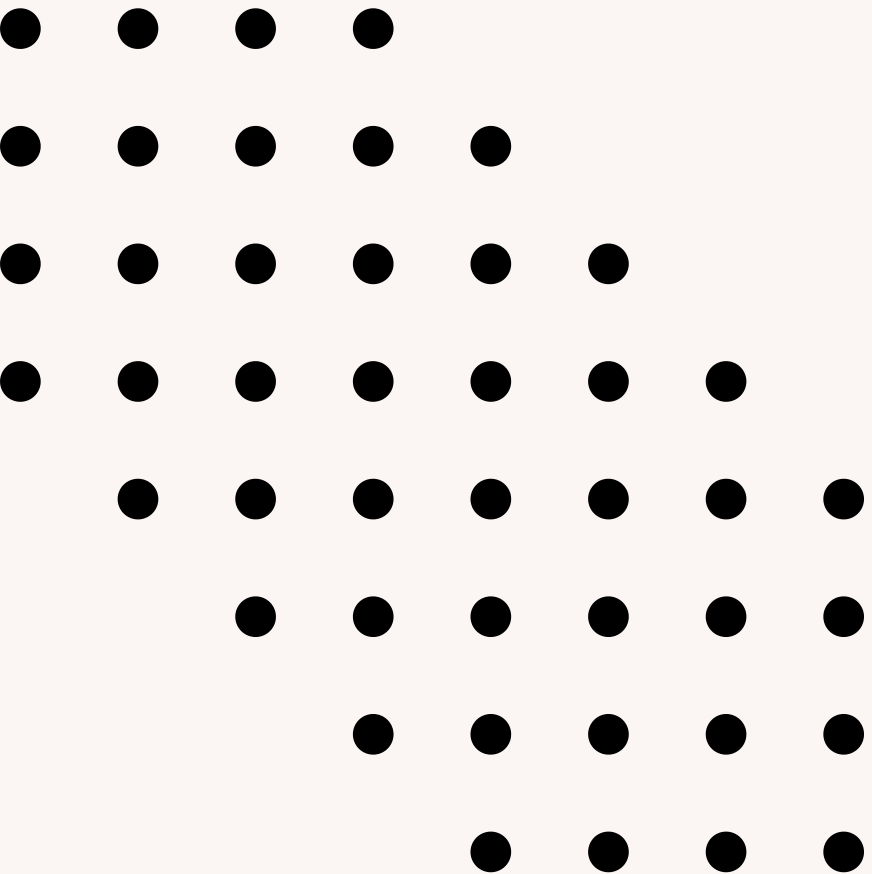
5. CLASSIFICAÇÃO

Com base nas probabilidades calculadas anteriormente e, como $P(\text{Gripe}=\text{Sim} \mid x) < P(\text{Gripe}=\text{Não} \mid x)$, podemos concluir que o paciente não está com gripe.

ATRIBUTOS CONTÍNUOS

- atributos contínuos são os dados que podem assumir diferentes valores dentro de um intervalo específico
- quando os atributos são contínuos, assumimos que estes possuem uma distribuição normal
- com isso em mente, podemos calcular a média e o desvio padrão para cada dado a partir dos dados de treinamento
- com os valores calculados, podemos estimar a probabilidade com a seguinte fórmula:

$$P(x_i | y) = \frac{1}{\sqrt{2\pi\sigma_y^2}} \exp \left(-\frac{(x_i - \mu_y)^2}{2\sigma_y^2} \right)$$



Alunos

Felipe Archanjo da Cunha Mendes

Thiago Gariani Quinto

Marcos Vinicius de Quadros

João Tiago Mielli