ANÁLISE INTELIGENTE DE DADOS (COB 754)

NAÏVE BAYES

LETÍCIA MARTINS RAPOSO

NAÏVE BAYES

CARACTERÍSTICAS

- CLASSIFICADORES PROBABILÍSTICOS BASEADOS NA APLICAÇÃO DO TEOREMA DE BAYES.
- PRESSUPOSTOS DE INDEPENDÊNCIA ENTRE AS VARIÁVEIS.
- CLASSIFICAÇÃO BINÁRIA E MULTICLASSE.

NAÏVE BAYES

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

USA O TEOREMA DE BAYES COMO SEU PRINCÍPIO DE FUNCIONAMENTO.

O TEOREMA DE BAYES TRATA DE PROBLEMAS EM QUE SE DESEJA DETERMINAR A PROBABILIDADE DE UM EVENTO OCORRER DADA UMA CONDIÇÃO (A PROBABILIDADE DE OCORRER A, NA CONDIÇÃO DE QUE B JÁ TENHA OCORRIDO).

NAÏVE BAYES

EX: DETERMINAR SE UMA PESSOA ESTÁ GRIPADA OU NÃO

VARIÁVEIS EXPLICATIVAS:

CORIZA, TOSSE, FEBRE, DOR MUSCULAR, DOR DE GARGANTA

VARIÁVEL RESPOSTA:

GRIPE (SIM/NÃO)

NAÏVE BAYES

Coriza = Sim
Tosse = Frequente
Febre = Alta
Dor muscular = Sim
Dor de garganta = Não

$$P(y_i|x) = \frac{P(x|y_i)P(y_i)}{P(x)}$$

PROBABILIDADE A POSTERIORI:
PROBABILIDADE DO EXEMPLO PERTENCER À CLASSE y_i

Probabilidade de Gripe | Coriza = Sim, Tosse = Frequente, Febre = Alta, Dor muscular = Sim, Dor de garganta = Não

NAÏVE BAYES

Coriza = Sim
Tosse = Frequente
Febre = Alta
Dor muscular = Sim
Dor de garganta = Não

$$P(y_i|x) = \frac{P(x|y_i)P(y_i)}{P(x)}$$

PROBABILIDADE CONDICIONAL (VEROSSIMILHANÇA):
PROBABILIDADE DE OBSERVAR VÁRIOS OBJETOS QUE PERTENCEM À CLASSE

Probabilidade de Coriza = Sim, Tosse = Frequente, Febre = Alta, Dor muscular = Sim, Dor de garganta = Não | Gripe

NAÏVE BAYES

Coriza = Sim
Tosse = Frequente
Febre = Alta
Dor muscular = Sim
Dor de garganta = Não

$$P(y_i|x) = \frac{P(x|y_i)P(y_i)}{P(x)}$$

PROBABILIDADE A PRIORI DA CLASSE

Probabilidade de Gripe

NAÏVE BAYES

Coriza = Sim
Tosse = Frequente
Febre = Alta
Dor muscular = Sim
Dor de garganta = Não

$$P(y_i|x) = \frac{P(x|y_i)P(y_i)}{P(x)}$$

PROBABILIDADE DE OCORRÊNCIA DOS OBJETOS

Probabilidade de Coriza = Sim, Tosse = Frequente, Febre = Alta, Dor muscular = Sim, Dor de garganta = Não

NAÏVE BAYES

Coriza = Sim
Tosse = Frequente
Febre = Alta
Dor muscular = Sim
Dor de garganta = Não

$$P(y_i|x) = \frac{P(x|y_i)P(y_i)}{P(x)}$$

$$Posterior = \frac{Likelihood \times Prior}{Evidence}$$

NAÏVE BAYES

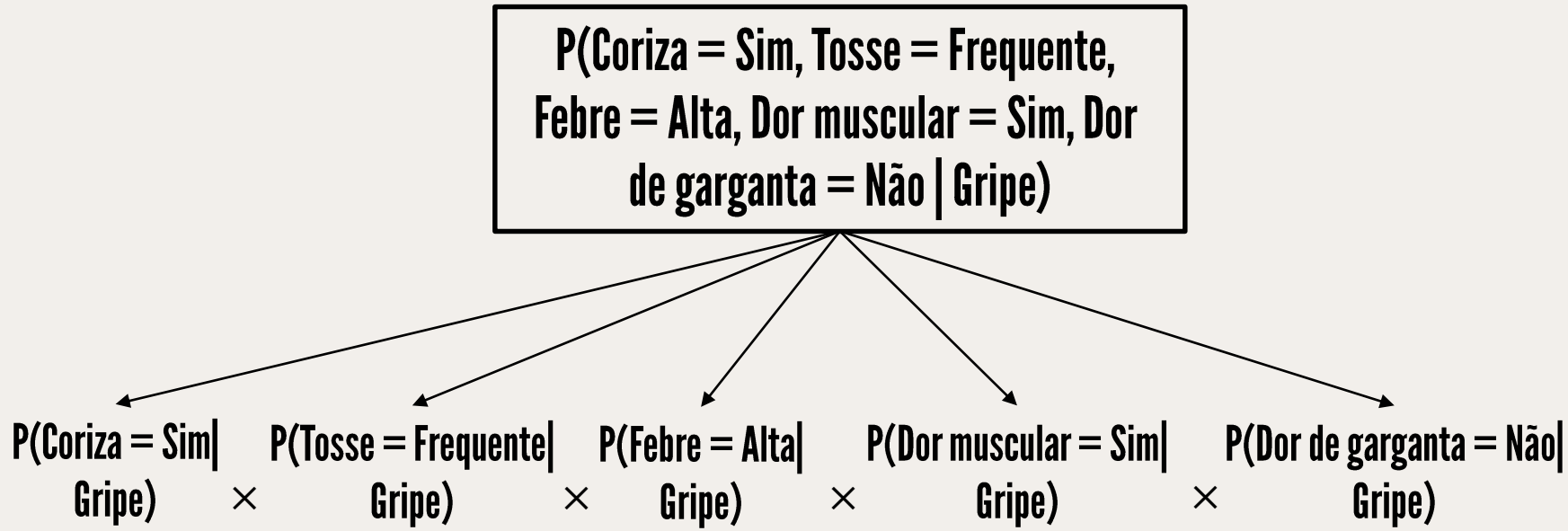
INGÊNUO (NAÏVE)

Assume que todos os valores dos atributos de um exemplo são independentes entre si, tornando o cálculo mais fácil.

EVENTOS INDEPENDENTES

- A ocorrência do evento A em nada interfere na probabilidade de ocorrência do outro evento, B.
- A probabilidade de ambos ocorrerem é igual ao produto de suas probabilidades.

NAÏVE BAYES


P(Coriza = Sim, Tosse = Frequente, Febre = Alta, Dor muscular = Sim, Dor de garganta = Não | Gripe)
P(Coriza = Sim| Gripe)
×
P(Tosse = Frequente| Gripe)
×
P(Febre = Alta| Gripe)
×
P(Dor muscular = Sim| Gripe)
×
P(Dor de garganta = Não| Gripe)

NAÏVE BAYES

- Dada a classe, $P(\boldsymbol{x}|y_i)$ pode ser decomposta em $P(x^1|y_i) \times \cdots \times P(x^d|y_i)$.

$$P(y_i|\boldsymbol{x}) = \frac{P(\boldsymbol{x}|y_i)P(y_i)}{P(\boldsymbol{x})}$$

$$P(y_i|\boldsymbol{x}) \propto \frac{P(x^1|y_i) \times \cdots \times P(x^d|y_i)P(y_i)}{P(\boldsymbol{x})} \propto \frac{\prod_{j=1}^{d} P(x^j|y_i)P(y_i)}{P(\boldsymbol{x})}$$

- Essa hipótese (independência entre atributos) é quase sempre violada.
  - Mas, na prática, o classificador naïve Bayes se mostra bastante robusto.

NAÏVE BAYES

Além disso, o denominador pode ser ignorado, uma vez que é o mesmo para todas as classes.

$$P(y_i|\boldsymbol{x}) \propto \frac{\prod_{j=1}^{d} P\left(x^j|y_i\right)P(y_i)}{P(\boldsymbol{x})}$$

$$P(y_i|\boldsymbol{x}) \propto \prod_{j=1}^{d} P\left(x^j|y_i\right)P(y_i)$$

NAÏVE BAYES

- Depois de calcular a probabilidade a posteriori para várias classes diferentes, pode-se selecionar a classe com a maior probabilidade.
- Este método é denominado estimativa por MAP (do inglês, *Maximum A Posteriori*).

$$\mathrm{MAP}(y_i) = \max_{y_i \in \{\text{classes}\}} \left(P(y_i) \prod_{j=1}^{d} P\left(x^j | y_i\right) \right)$$

NAÏVE BAYES

- Se tivermos o mesmo número de observações em cada classe em nossos dados de treinamento, a probabilidade de cada classe $P(y_i)$ será igual.
- Mais uma vez, este seria um termo constante em nossa equação e poderíamos descartá-lo:

$$\text{MAP}(y_i) = \max_{y_i \in \{\text{classes}\}} \left(\prod_{j=1}^{d} P(x^j | y_i) \right)$$

NAÏVE BAYES

Podemos aplicar log nas probabilidades, pois só precisamos saber qual classe tem maior probabilidade e não a específica.

- Isso evita que probabilidades pequenas sejam multiplicadas entre si gerando valores menores ainda.

$$\log P(y_i|\boldsymbol{x}) \propto \log\left[P(y_i)\prod_{j=1}^{d} P\left(x^j|y_i\right)\right]$$

$$\log P(y_i|\boldsymbol{x}) \propto \log P(y_i) + \sum_{j=1}^{d} \log P\left(x^j|y_i\right)$$

NAÏVE BAYES

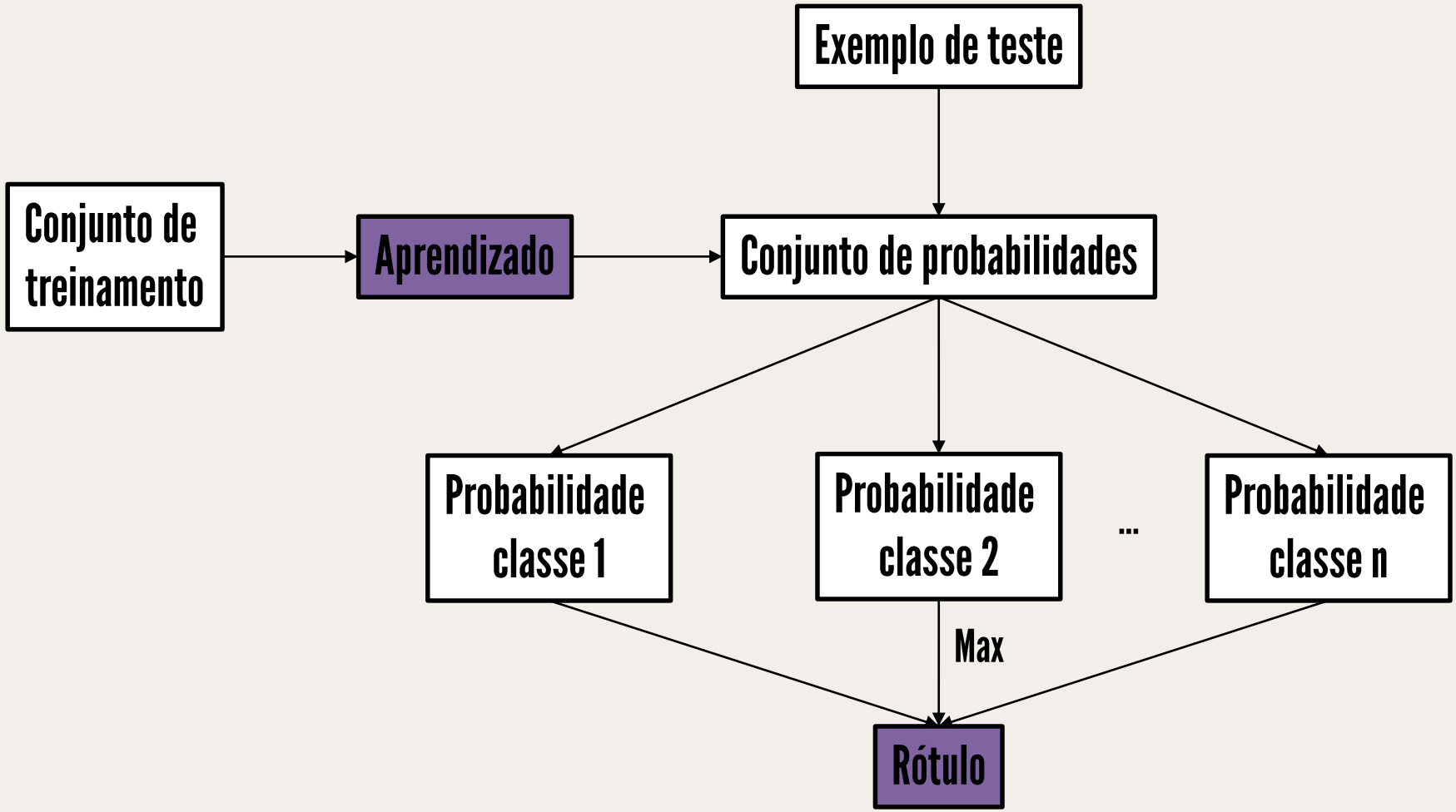
No caso de duas classes:

$$\log\left(\frac{P(y_1|\boldsymbol{x})}{P(y_2|\boldsymbol{x})}\right) \propto \log\left(\frac{P(y_1)}{P(y_2)}\right) + \sum_{\boldsymbol{j=1}}^{\boldsymbol{d}} \log\left(\frac{P\left(x^j|y_1\right)}{P\left(x^j|y_2\right)}\right)$$

Se $\log\left(\frac{P(y_1|\boldsymbol{x})}{P(y_2|\boldsymbol{x})}\right)$ for positivo, os atributos contribuem para a predição da classe y_1.

log positivo, $\frac{P(y_1|\boldsymbol{x})}{P(y_2|\boldsymbol{x})} > \mathbf{1}$, logo $P(y_1|\boldsymbol{x}) > P(y_2|\boldsymbol{x})$.

FUNCIONAMENTO



EXEMPLO

DIA	APARÊNCIA	TEMPERATURA	UMIDADE	VENTO	JOGAR TÊNIS
1	Ensolarado	Quente	Alta	Fraco	Não
2	Ensolarado	Quente	Alta	Forte	Não
3	Nublado	Quente	Alta	Fraco	Sim
4	Chuva	Moderada	Alta	Fraco	Sim
5	Chuva	Fria	Normal	Fraco	Sim
6	Chuva	Fria	Normal	Forte	Não
7	Nublado	Fria	Normal	Forte	Sim
8	Ensolarado	Moderada	Alta	Fraco	Não
9	Ensolarado	Fria	Normal	Fraco	Sim
10	Chuva	Moderada	Normal	Fraco	Sim
11	Ensolarado	Moderada	Normal	Forte	Sim
12	Nublado	Moderada	Alta	Forte	Sim
13	Nublado	Quente	Normal	Fraco	Sim
14	Chuva	Moderada	Alta	Forte	Não

EXEMPLO

Jogar tênis ou não, dado que:
Aparência = Ensolarado,
Temperatura = Fria,
Umidade = Alta e
Vento = Forte?

EXEMPLO

Quais probabilidades precisamos?

$$\text{MAP}(y_i) = \max_{y_i \in \{\text{Sim,Não}\}} P(y_i) \times P(\text{Aparência}=\text{Ensolarado}|y_i) \times P(\text{Temperatura} = \text{Fria}|y_i) \times P(\text{Umidade}=\text{Alta}|y_i) \times P(\text{Vento} = \text{Forte}|y_i)$$

- Probabilidade das classes:
  P(Sim) e P(Não)
- Cada probabilidade para as duas possíveis classes:
  P(Aparência=Ensolarado | Sim) e P(Aparência=Ensolarado | Não)
  P(Temperatura=Fria | Sim) e P(Temperatura=Fria | Não)
  P(Umidade=Alta | Sim) e P(Umidade=Alta | Não)
  P(Vento=Forte | Sim) e P(Vento=Forte | Não)

EXEMPLO

Probabilidade das classes:

P(Jogar = Sim) = 9/14

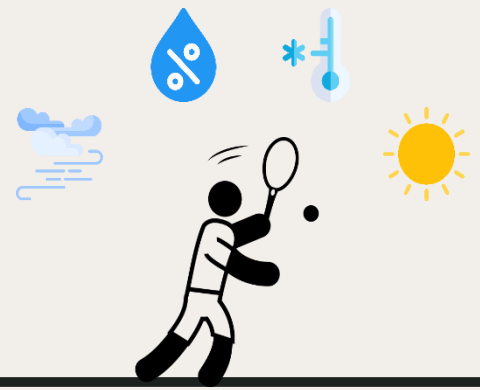
P(Jogar = Não) = 5/14

DIA	APARÊNCIA	TEMPERATURA	UMIDADE	VENTO	JOGAR TÊNIS
1	Ensolarado	Quente	Alta	Fraco	Não
2	Ensolarado	Quente	Alta	Forte	Não
3	Nublado	Quente	Alta	Fraco	Sim
4	Chuva	Moderada	Alta	Fraco	Sim
5	Chuva	Fria	Normal	Fraco	Sim
6	Chuva	Fria	Normal	Forte	Não
7	Nublado	Fria	Normal	Forte	Sim
8	Ensolarado	Moderada	Alta	Fraco	Não
9	Ensolarado	Fria	Normal	Fraco	Sim
10	Chuva	Moderada	Normal	Fraco	Sim
11	Ensolarado	Moderada	Normal	Forte	Sim
12	Nublado	Moderada	Alta	Forte	Sim
13	Nublado	Quente	Normal	Fraco	Sim
14	Chuva	Moderada	Alta	Forte	Não

EXEMPLO

Aparência	Jogar = Sim	Jogar = Não
Ensolarado	2/9	3/5
Nublado	4/9	0/5
Chuva	3/9	2/5

Temperatura	Jogar = Sim	Jogar = Não
Quente	2/9	2/5
Moderada	4/9	2/5
Fria	3/9	1/5

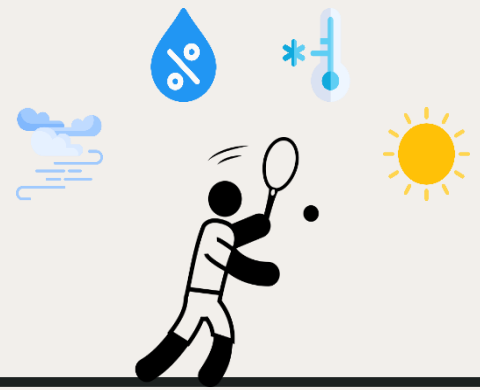


DIA	APARÊNCIA	TEMPERATURA	UMIDADE	VENTO	JOGAR TÊNIS
1	Ensolarado	Quente	Alta	Fraco	Não
2	Ensolarado	Quente	Alta	Forte	Não
3	Nublado	Quente	Alta	Fraco	Sim
4	Chuva	Moderada	Alta	Fraco	Sim
5	Chuva	Fria	Normal	Fraco	Sim
6	Chuva	Fria	Normal	Forte	Não
7	Nublado	Fria	Normal	Forte	Sim
8	Ensolarado	Moderada	Alta	Fraco	Não
9	Ensolarado	Fria	Normal	Fraco	Sim
10	Chuva	Moderada	Normal	Fraco	Sim
11	Ensolarado	Moderada	Normal	Forte	Sim
12	Nublado	Moderada	Alta	Forte	Sim
13	Nublado	Quente	Normal	Fraco	Sim
14	Chuva	Moderada	Alta	Forte	Não

EXEMPLO

Umidade	Jogar = Sim	Jogar = Não
Alta	3/9	4/5
Normal	6/9	1/5

Vento	Jogar = Sim	Jogar = Não
Forte	3/9	3/5
Fraco	6/9	2/5



DIA	APARÊNCIA	TEMPERATURA	UMIDADE	VENTO	JOGAR TÊNIS
1	Ensolarado	Quente	Alta	Fraco	Não
2	Ensolarado	Quente	Alta	Forte	Não
3	Nublado	Quente	Alta	Fraco	Sim
4	Chuva	Moderada	Alta	Fraco	Sim
5	Chuva	Fria	Normal	Fraco	Sim
6	Chuva	Fria	Normal	Forte	Não
7	Nublado	Fria	Normal	Forte	Sim
8	Ensolarado	Moderada	Alta	Fraco	Não
9	Ensolarado	Fria	Normal	Fraco	Sim
10	Chuva	Moderada	Normal	Fraco	Sim
11	Ensolarado	Moderada	Normal	Forte	Sim
12	Nublado	Moderada	Alta	Forte	Sim
13	Nublado	Quente	Normal	Fraco	Sim
14	Chuva	Moderada	Alta	Forte	Não

EXEMPLO

x'=(Aparência = Ensolarado, Temperatura = Fria, Umidade = Alta e Vento = Forte)

P(Aparência = Ensolarado|Jogar = Sim) = 2/9
P(Temperatura = Fria|Jogar = Sim) = 3/9
P(Umidade = Alta|Jogar = Sim) = 3/9
P(Vento = Forte|Jogar = Sim) = 3/9
P(Jogar = Sim) = 9/14

P(Aparência=Ensolarado|Jogar=Não) = 3/5
P(Temperatura=Fria|Jogar=Não) = 1/5
P(Umidade=Alta|Jogar=Não) = 4/5
P(Vento=Forte|Jogar=Não) = 3/5
P(Jogar=Não) = 5/14

$$P(Sim|x') \approx [P(Ensolarado|Sim)P(Fria|Sim)P(Alta|Sim)P(Forte|Sim)]P(Jogar=Sim) = 0.0053$$

$$P(Não|x') \approx [P(Ensolarado|Não)\ P(Fria|Não)P(Alta|Não)P(Forte|Não)]P(Jogar=Não) = 0.0206$$

Como $P(Sim|x') < P(Não|x')$, rotulamos x' como "Não".

PROBLEMA DA FREQUÊNCIA ZERO

E QUANDO UM DETERMINADO VALOR NÃO APARECE NO TREINAMENTO, MAS APARECE NO TESTE?

EXEMPLO: TEMPO = “NUBLADO” PARA A CLASSE “NÃO”.

DIA	APARÊNCIA	TEMPERATURA	UMIDADE	VENTO	JOGAR TÊNIS
1	Ensolarado	Quente	Alta	Fraco	Não
2	Ensolarado	Quente	Alta	Forte	Não
3	Nublado	Quente	Alta	Fraco	Sim
4	Chuva	Moderada	Alta	Fraco	Sim
5	Chuva	Fria	Normal	Fraco	Sim
6	Chuva	Fria	Normal	Forte	Não
7	Nublado	Fria	Normal	Forte	Sim
8	Ensolarado	Moderada	Alta	Fraco	Não
9	Ensolarado	Fria	Normal	Fraco	Sim
10	Chuva	Moderada	Normal	Fraco	Sim
11	Ensolarado	Moderada	Normal	Forte	Sim
12	Nublado	Moderada	Alta	Forte	Sim
13	Nublado	Quente	Normal	Fraco	Sim
14	Chuva	Moderada	Alta	Forte	Não

PROBLEMA DA FREQUÊNCIA ZERO

- Nesse caso, a probabilidade a posteriori também será zero: P(“Não” | Nublado, ...) = 0
  - Multiplicação das probabilidades;
  - Não importa as probabilidades dos outros atributos.
- A base de treinamento pode não ser totalmente representativa.
  - Classes minoritárias podem ter valores raros.

PROBLEMA DA FREQUÊNCIA ZERO

- **Estimador de Laplace**: adicionar 1 unidade fictícia para cada combinação de valor-classe → valores sem exemplos de treinamento passam a conter 1 exemplo.

- Tempo = "Nublado" e Classe = "Não"
  - Somar 1 para cada combinação valor-classe
  - Somar 3 na base (3 combinações valor-classe)
    Sol: 3/5 → (3+1)/(5+3)
    Nublado: 0/5 → (0+1)/(5+3)
    Chuva: 2/5 → (2+1)/(5+3)
  - Isso deve ser feito para todas as classes: do contrário, estamos inserindo viés nas probabilidades de apenas uma classe.

PROBLEMA DA FREQUÊNCIA ZERO

- **Estimativa m**: adicionar múltiplas unidades fictícias para cada combinação de valor-classe.
  - Solução mais geral.

- Exemplo: Tempo = “Nublado” e Classe = “Não”
  - Sol: 3/5 → $\frac{3+\frac{m}{3}}{5+m}$
  - Nublado: 0/5 → $\frac{0+\frac{m}{3}}{5+m}$
  - Chuva: 2/5 → $\frac{2+\frac{m}{3}}{5+m}$

O QUE FAZER SE UMA AMOSTRA NÃO TIVER O VALOR DE UM ATRIBUTO?

TREINAMENTO

Devemos excluir a amostra do conjunto de treinamento.

TESTE

Devemos considerar apenas os demais atributos da amostra.

E SE OS ATRIBUTOS FOREM CONTÍNUOS?

ALTERNATIVA 1: DISCRETIZAR OS DADOS

- nº de intervalos fixado em k=min(10, nº de valores diferentes).
- Muita informação pode vir a ser perdida.

ALTERNATIVA 2: CONSIDERAR UMA FUNÇÃO DE DENSIDADE DE PROBABILIDADE NO CÁLCULO DA PROBABILIDADE CONDICIONAL

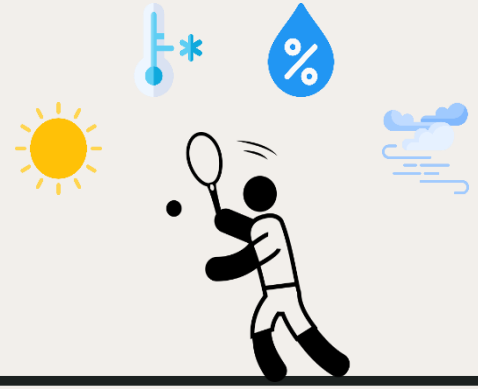
- Geralmente, usa-se a distribuição Normal.
- Pode-se considerar outras distribuições que melhor caracterizam os dados.

ALTERNATIVA 2:
CONSIDERAR UMA FUNÇÃO DE DENSIDADE DE PROBABILIDADE NO CÁLCULO DA PROBABILIDADE CONDICIONAL

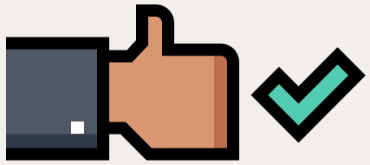
Ex: temperatura, $\mu = 73, \sigma = 6{,}2$

$$f(temperatura = 66, jogo = sim) = \frac{1}{\sqrt{2\pi}6{,}2} \exp\left[-\frac{1}{2}\left(\frac{66-73}{6{,}2}\right)^2\right] = 0{,}034$$

Esse seria o valor da probabilidade condicional



VANTAGENS



BOM DESEMPENHO NA PREDIÇÃO DE MULTICLASSSES

FÁCIL DE IMPLEMENTAR DE FORMA INCREMENTAL

FÁCIL INTERPRETAÇÃO

MELHOR DESEMPENHO COM ENTRADA CATEGÓRICA

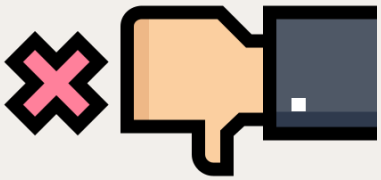
ROBUSTO A OUTLIERS E ATRIBUTOS IRRELEVANTES

CAPAZ DE CLASSIFICAR AMOSTRAS COM VALORES AUSENTES

PARALELIZAÇÃO

BOA SOLUÇÃO QUANDO SE TEM POUCOS DADOS

DESVANTAGENS



SUPOSIÇÃO DE PREDITORES INDEPENDENTES

DESEMPENHO PODE SER AFETADO PELA PRESENÇA DE ATRIBUTOS CORRELACIONADOS

PROBLEMA DA FREQUÊNCIA ZERO

APLICAÇÕES

- **PREVISÕES EM TEMPO REAL**: ALGORITMO RÁPIDO, PODE SER USADO PARA FAZER PREVISÕES EM TEMPO REAL.
- **PREVISÕES MULTICLASSES**: CAPAZ DE PREDIZER A PROBABILIDADE DE MÚLTIPLAS CLASSES DAS VARIÁVEIS-ALVO.
- **CLASSIFICAÇÃO DE TEXTOS/FILTRAGEM DE SPAM/ANÁLISE DE SENTIMENTO**
- **SISTEMA DE RECOMENDAÇÃO**: UTILIZADO NA PREDIÇÃO DE SERVIÇOS QUE UM USUÁRIO PODERIA GOSTAR.

RESUMO

- TÉCNICA DE CLASSIFICAÇÃO BASEADA NO TEOREMA DE BAYES.
- SUPOSIÇÃO DE INDEPENDÊNCIA ENTRE OS PREDITORES.
- FAZ PREDIÇÕES DE MÚLTIPLAS CLASSES.
- DESEMPENHO MELHOR PARA VARIÁVEIS DE ENTRADA CATEGÓRICAS.
- USO DO ESTIMADOR DE LAPLACE EM PROBLEMAS DE FREQUÊNCIA ZERO.