

ICC204 - Aprendizagem de Máquina e Mineração de Dados

Classificação

(parte 1/3)



Prof. Rafael Giusti
rgiusti@icomp.ufam.edu.br

Leitura recomendada

- Linguagens de hipóteses, linguagens de modelos ou linguagens de representação
 - Tom M. Mitchel. *Machine Learning*, 1ª Edição. Capítulo 1
- Aprendizado probabilístico
 - Tom M. Mitchell. *Machine Learning*, 1ª Edição. Capítulo 6
 - Seção 6.1 e 6.2: introdução e teorema de Bayes
 - Seção 6.3: aprendizado Bayesiano e MAP
 - Seção 6.9: classificador ingênuo de Bayes
- Árvores de decisão
 - Ian H. Witten. *Data Mining* (3ª Ed.)
 - Capítulo 3, Seção 3.3 – descrição e uso do modelo
 - Capítulo 4, Seção 4.3 – algoritmo de indução de árvores

Leitura recomendada (2)

- Árvores de decisão (2)
 - Russel e Norvig. "*Artificial Intelligence: A Modern Approach*", 2^a Ed.
 - Seção 18.3 - *Learning Decision Trees*

Material suplementar

- Alfabeto, linguagem e cadeias (*strings*)
 - John E. Hopcroft. "*Introduction to Automata Theory, Languages, and Computation*", Seção 1.5
- Linguagens de hipóteses, linguagens de modelos ou linguagens de representação
 - Claude Sammut. *Encyclopedia of Machine Learning*.
 - Verbetes: *Hypothesis Language*, *Hypothesis Space* e *Learning as Search*
- Teoria da Informação: entropia, informação própria e informação mútua
 - Bishop. "*Pattern Recognition and Machine Learning*"
 - Seção 1.6 – *Information Theory*
- Árvores de decisão
 - J.R. Quinlan. *Induction of Decision Trees*. *Machine Learning*, Vol. 1, Issue 1, pp. 81-106. Springer. 1986 (Disponível no ColabWeb)

Agenda

- Parte 1/3
 - Definições
 - Teoria das probabilidades
 - Aprendizado Bayesiano e modelos probabilísticos
- Parte 2/3
 - Modelos baseados em árvores
 - Modelos baseados em regras
- Parte 3/3
 - Classificação preguiçosa: k-NN
 - Máquina de vetores de suporte

Agenda

- Definições
- Teoria das probabilidades
- Aprendizado Bayesiano e modelos probabilísticos
- Modelos baseados em árvores
- Modelos baseados em regras
- Classificação preguiçosa: k-NN
- Máquina de vetores de suporte

Classificador

- Um classificador é um modelo na forma
 - $f(\mathbf{x}) = \hat{y}$
- sendo que $y \in \Omega = \{\Omega_1, \Omega_2, \dots, \Omega_c\}$ é o espaço de classes associado à função-conceito
 - Quando $|\Omega| = 2$, tem-se classificação **binária**
 - Quando $|\Omega| > 2$, tem-se classificação **multiclasse**
 - Não confundir com classificação **multirrótulo**

Linguagem de representação

- O objetivo de um modelo é **expressar conhecimento**
- Conhecimento precisa ser **aproximado** por uma linguagem
 - Linguagem: um conjunto de **sentenças**
 - Português: todas as sentenças que podem ser ditas em português
 - C++: todos os programas que podem ser escritos em C++ e que, ao menos, compilam
 - Árvores: todas as árvores que expressão decisões sobre um conjunto de atributos

Linguagem de representação

- Toda linguagem é definida por
 - Um alfabeto Σ que especifica os **símbolos** válidos da linguagem (exemplo: palavras em português)
 - Uma **gramática** que especifica quais sequências de símbolos são válidas
 - **Cadeia**: uma sequência de símbolos
 - **Sentença**: uma cadeia que faz parte de uma linguagem

Exemplo informal de uma linguagem

- Vamos descrever uma linguagem de representação
- Nossa descrição será **bastante** informal
 - Uma descrição formal foge muito ao escopo da disciplina e pode ser estudada no contexto de autômatos, linguagens formais ou teoria da computação
- Nosso exemplo será uma hipótese para **instâncias positivas** do conjunto de dados **tênis** (isto é, jogar), mas pode ser estendido para qualquer domínio

Exemplo informal de uma linguagem

- Descrição informal do alfabeto Σ :
- Fazem parte do alfabeto:
 - Os símbolos "<", ">", "?" e "Ø"
 - Os valores dos atributos (ensolarado, nublado, etc.)
 - O separador vírgula

Exemplo informal de uma linguagem

- Descrição informal da gramática
 - Seja R_i uma **decisão** sobre o atributo X_i
 - R_i pode ser um dos possíveis valores de X_i
 - Ou um dos símbolos especiais ? e \emptyset
 - Uma cadeia $\langle R_1, R_2, \dots, R_M \rangle$ é uma sentença da linguagem e uma hipótese sobre a decisão "*jogar tênis*"

Exemplo informal de uma linguagem

- Descrição informal da semântica
 - Se R_i for um valor para o atributo X_i , então a hipótese indica valores aceitáveis de X_i para exemplos da classe **sim**
 - Se $R_i = ?$, então qualquer valor para esse atributo é aceitável
 - Se $R_i = \emptyset$, então nenhum valor para esse atributo é aceitável

Exemplo informal de uma linguagem

- Então um possível modelo (ou hipótese) seria
 - $H_1 = \langle \text{ensolarado, moderado, alta, forte} \rangle$
 - Essa hipótese diz que...
 - Se aparência=ensolarado,
temperatura=moderado,
umidade=alta e
vento=forte
Então nós iremos jogar tênis

Exemplo informal de uma linguagem

- Então um possível modelo (ou hipótese) seria
 - $H_2 = \langle ?, \text{moderado}, \text{alta}, \text{forte} \rangle$
 - Essa hipótese diz que...
 - Se temperatura=moderado,
umidade=alta e
vento=forte
Então nós iremos jogar tênis

Exemplo informal de uma linguagem

- Então um possível modelo (ou hipótese) seria
 - $H_3 = \langle ?, ?, ?, ? \rangle$
 - Essa hipótese diz que...
 - **sempre** iremos jogar tênis, não importa quais são os valores dos atributos!

Exemplo informal de uma linguagem

- Então um possível modelo (ou hipótese) seria
 - $H_4 = \langle \emptyset, \text{moderado}, \text{alta}, \text{forte} \rangle$
 - Essa hipótese diz que...
 - Se o valor de **aparência** for conhecido, não importa qual seja esse valor, então **não** iremos jogar tênis

Exemplo informal de uma linguagem

- Então um possível modelo (ou hipótese) seria
 - $H_5 = \langle \emptyset, \emptyset, \emptyset, \emptyset \rangle$
 - Essa hipótese diz que...
 - **Nunca** vamos jogar tênis, não importa quais sejam os valores dos atributos

Viés de representação

- Quando escolhemos uma linguagem de representação para as nossas hipóteses, introduzimos um **viés de representação**
- No exemplo anterior, todas as hipóteses que podemos representar são compatíveis com decisões sobre um valor específico dos atributos
- Não podemos representar uma hipótese do tipo
 - "Se a aparência for *nublado* **ou** *ensolarado*, então iremos jogar tênis"

Viés de representação

- O viés de representação limita o espaço de hipóteses
- Nesse caso, podemos facilmente calcular o tamanho
 - Aparência: pode ter 3 valores + 2 especiais
 - Temperatura: 3 valores + 2 especiais
 - Umidade: 2 valores + 2 especiais
 - Vento: 2 valores + 2 especiais
 - Total de possíveis hipóteses:
 $5 * 5 * 4 * 4 = 400$

Aprendizado como busca

- Aprender uma hipótese para o problema de decisão "*vou jogar tênis hoje*" significa percorrer esse espaço de 400 hipóteses e decidir qual é a mais adequada
 - E se a linguagem de descrição do modelo fosse outra?
 - Então teríamos um espaço de hipóteses diferente

Aprendizado como busca

- Exemplo de algoritmo: FIND-S
 - Comece com a hipótese mais restrita possível
 - $H_1 = \langle \emptyset, \emptyset, \emptyset, \emptyset \rangle$
 - Refine essa hipótese iterativamente até que não seja possível torná-la mais adequada para os exemplos do conjunto de treinamento
 - Tom M. Mitchell, "*Machine Learning*", Capítulo 2

Aprendizado como busca

- Exemplo de algoritmo: ID3
 - Escolha um atributo
 - Coloque esse atributo na raiz da árvore
 - Construa o restante da árvore iterativamente

Agenda

- Definições
- Teoria das probabilidades
- Aprendizado Bayesiano e modelos probabilísticos
- Modelos baseados em árvores
- Modelos baseados em regras
- Classificação preguiçosa: k-NN
- Máquina de vetores de suporte

Variável aleatória

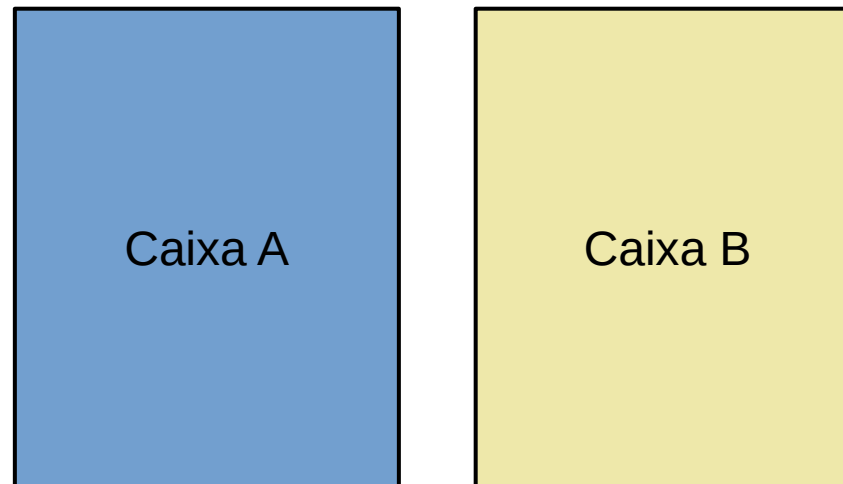
- Uma variável aleatória X é um modelo matemático para um objeto associado a **indecisão**
 - Um dado cujo lançamento produz um resultado que não pode ser **definido a priori**
 - Uma caixa contendo bolas numeradas de 1 a 50
 - Um **classificador** cujo erro no mundo real não é conhecido antes de ser utilizado
 - Uma moeda com duas caras?

Variável aleatória

- Um dado pode ser modelado como uma VA?
- O resultado do lançamento de um dado pode ser modelado como uma VA?
 - Uma VA X cujo espaço de eventos são os possíveis resultados do dado
 - $\Pr(X = 1)$: probabilidade de sair um 1
 - $\Pr(X = 2)$: ...
 - $\Pr(X = i)$: ...

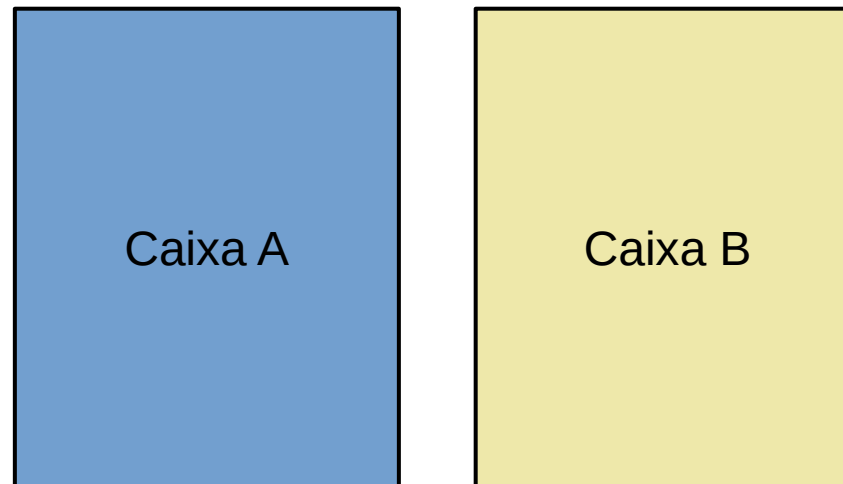
Variável aleatória

- Suponha duas caixas
 - Uma caixa A e uma caixa B
 - De cada caixa pode-se retirar bolas verdes ou laranjas, que sempre voltam para a caixa



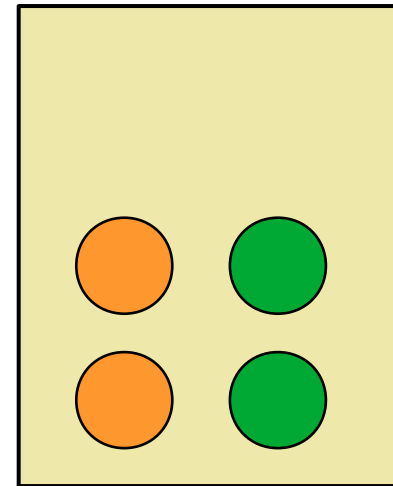
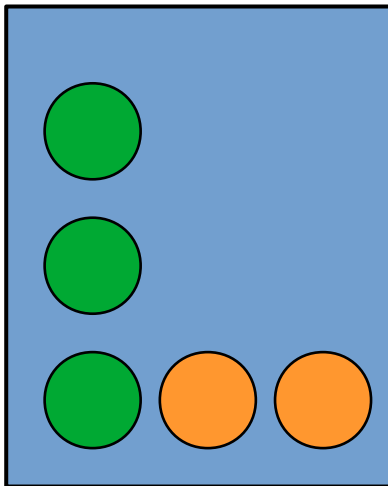
Variável aleatória

- Se escolhermos uma caixa aleatoriamente, podemos fazer perguntas como
 - Qual a probabilidade de extrair uma bola verde?
 - E se eu escolher a caixa B?



Variável aleatória

- Do ponto de vista **frequentista**, podemos encarar isso como:
 - Qual a fração dos eventos que resultam em uma bola laranja ou verde?



Regras da soma e do produto

- Seja X uma VA que indica a caixa
 - $p(X=A) = 0,5$
 - $p(X=B) = 0,5$
- Seja Y uma VA que indica a cor
 - $p(Y=\text{verde}) = p(X=A, Y=\text{verde}) + p(X=B, Y=\text{verde})$
 - $p(Y = \text{laranja}) = p(X=A, Y=\text{laranja}) + p(X= B, Y=\text{laranja})$

Regras da soma e do produto

- Regra da soma

$$p(X) = \sum_Y p(X, Y)$$

- Regra do produto

$$p(X, Y) = p(Y|X) \cdot p(X)$$

Teorema de Bayes

- Com base nas regras da soma e do produto
 - E também observando que $p(X, Y) = p(Y, X)$
- Podemos derivar o **teorema de Bayes**

$$p(Y|X) = \frac{p(X|Y) \cdot p(Y)}{p(X)}$$

- **Exercício:** derivem o teorema (depois da aula)

Teoria da decisão Bayesiana

- Seja agora X uma variável aleatória que representa um exemplo
- Seja Y uma variável aleatória que representa a classe
- Podemos utilizar o teorema de Bayes para construir um classificador?

Agenda

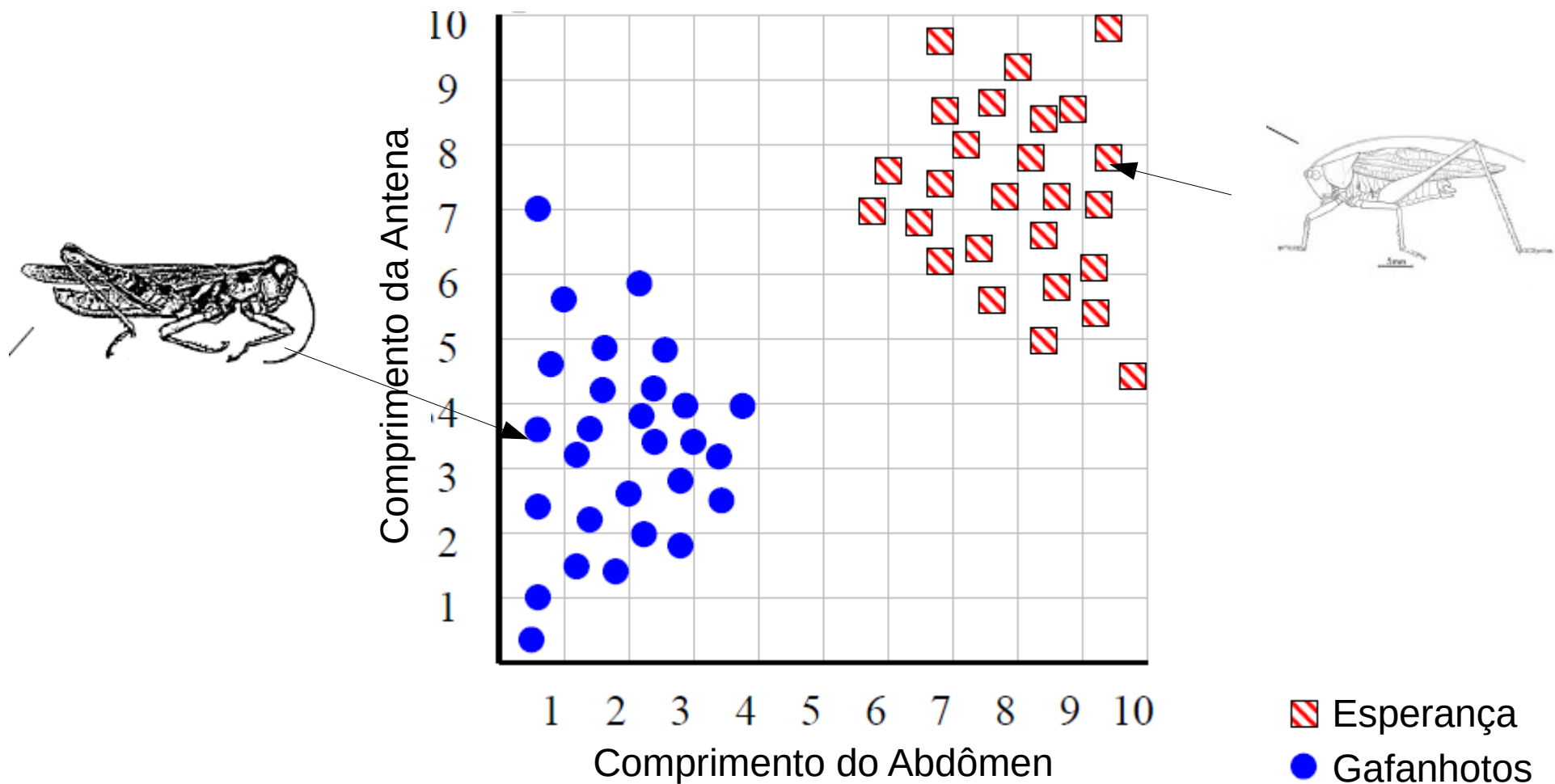
- Definições
- Teoria das probabilidades
- **Aprendizado Bayesiano e modelos probabilísticos**
- Modelos baseados em árvores
- Modelos baseados em regras
- Classificação preguiçosa: k-NN
- Máquina de vetores de suporte

Um exemplo

- Vamos começar com um exemplo
 - As figuras dos próximos *slides* foram cedidas pelo Prof. Ricardo Marcacini da Universidade Federal do Mato Grosso do Sul

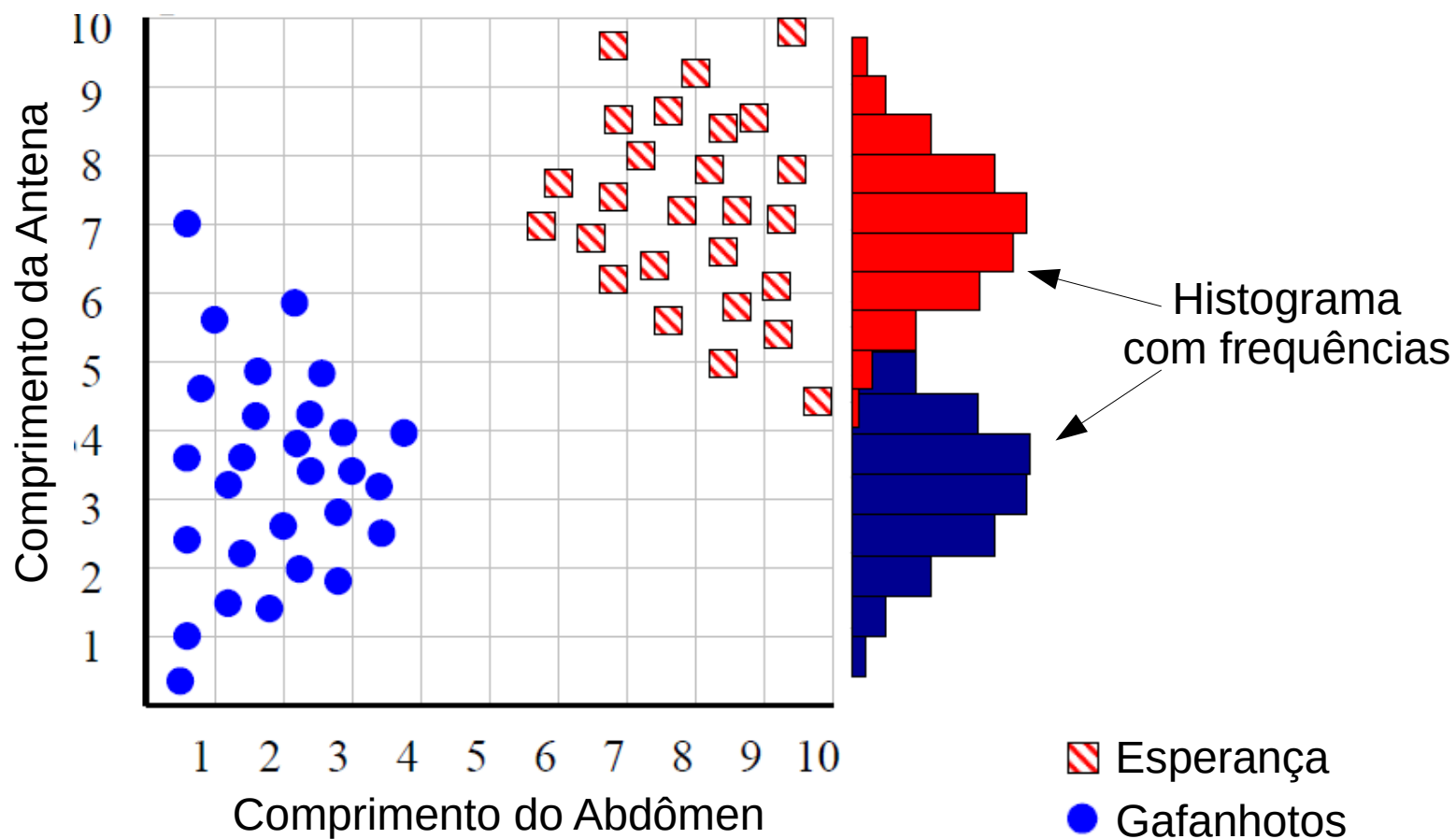
Aprendizado de Máquina: Abordagens Bayesianas

- Vamos começar com um exemplo...



Aprendizado de Máquina: Abordagens Bayesianas

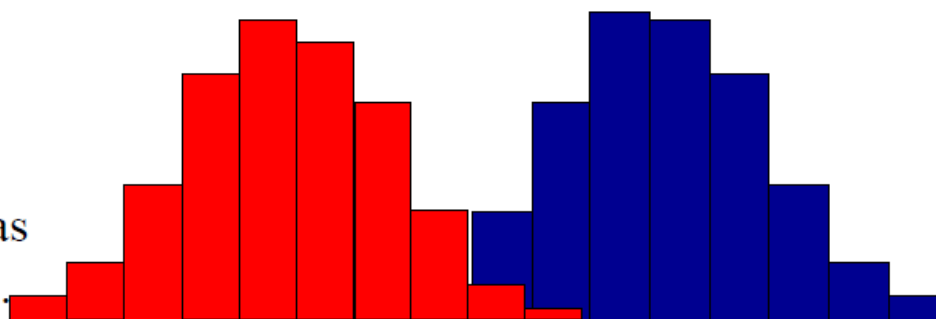
- Por enquanto, vamos olhar apenas para o comprimento da antena



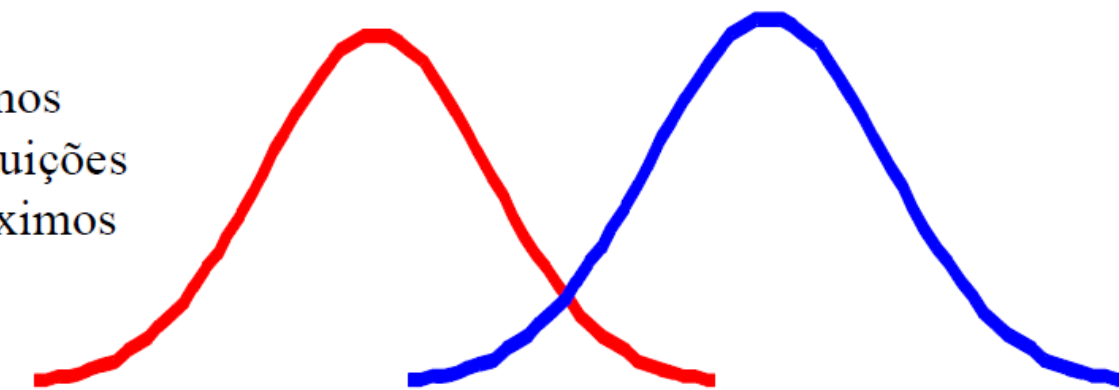
Aprendizado de Máquina: Abordagens Bayesianas

- Por enquanto, vamos olhar apenas para o comprimento da antena

Podemos deixar os histogramas como estão ou podemos sumariá-los com duas distribuições normais.

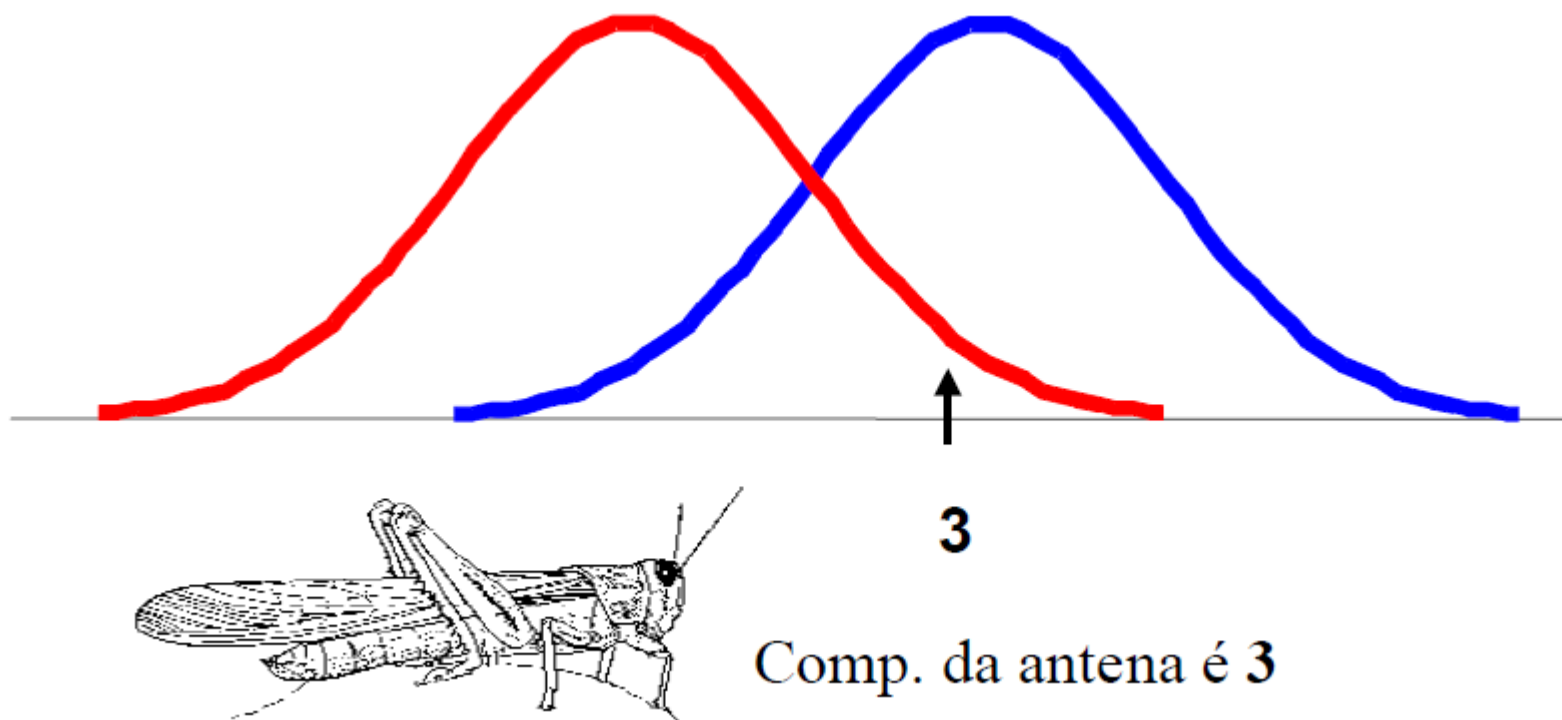


Para facilitar a visualização vamos usar duas distribuições normais nos próximos slides...



Aprendizado de Máquina: Abordagens Bayesianas

- Encontramos um novo inseto, com comprimento da antena = 3.
 - É um gafanhoto ou esperança?

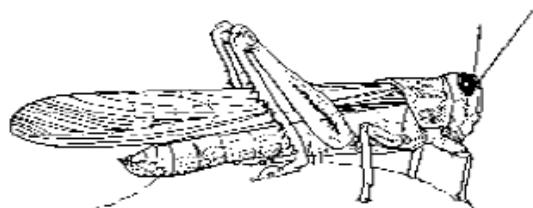
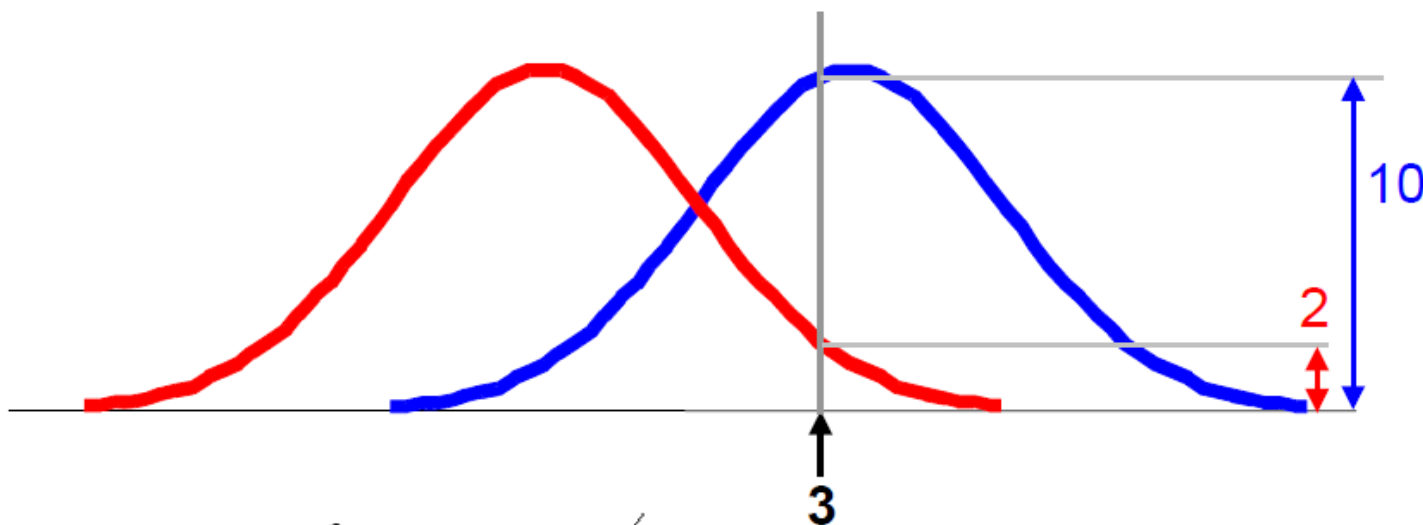


Aprendizado de Máquina: Abordagens Bayesianas

$p(c_j | d)$ = probabilidade da classe c_j , dado que observamos d

$$P(\text{Gafanhoto} | 3) = 10 / (10 + 2) = 0.833$$

$$P(\text{Esperança} | 3) = 2 / (10 + 2) = 0.166$$



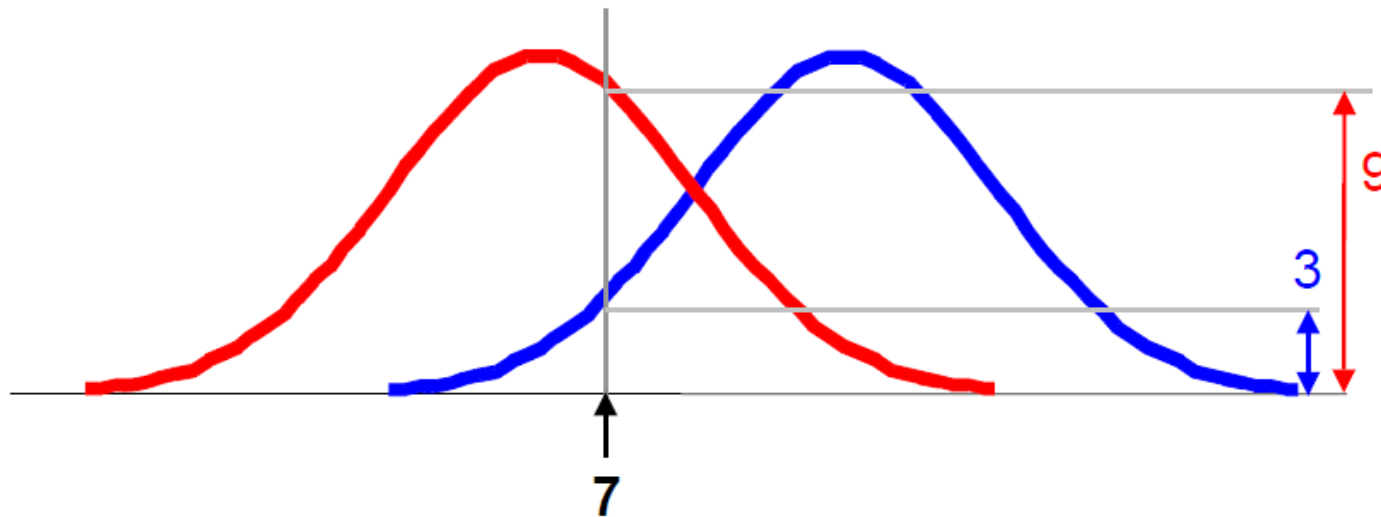
Comp. da antena é 3

Aprendizado de Máquina: Abordagens Bayesianas

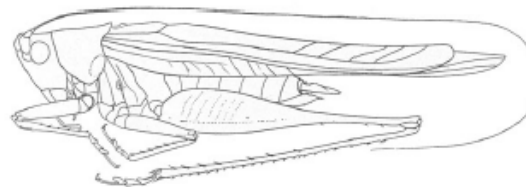
$p(c_j | d)$ = probabilidade da classe c_j , dado que observamos d

$$P(\text{Gafanhoto} | 7) = 3 / (3 + 9) = 0.250$$

$$P(\text{Esperança} | 7) = 9 / (3 + 9) = 0.750$$



Comp. da antena é 7

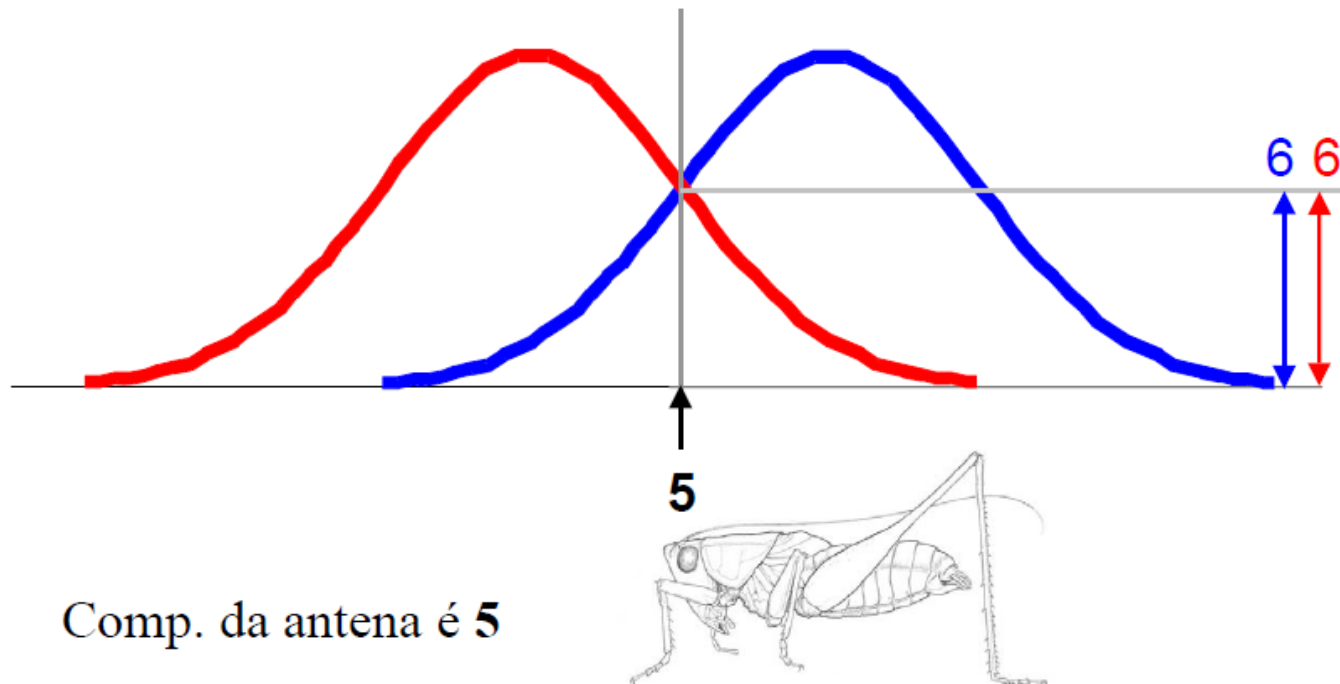


Aprendizado de Máquina: Abordagens Bayesianas

$p(c_j | d)$ = probabilidade da classe c_j , dado que observamos d

$$P(\text{Gafanhoto} | 5) = 6 / (6 + 6) = 0.500$$

$$P(\text{Esperança} | 5) = 6 / (6 + 6) = 0.500$$



Classificador Bayesiano

- Para uma única variável x , podemos considerar

$$p(c_j|x) = \frac{p(x|c_j) \cdot p(c_j)}{p(x)}$$

- $p(c_j|x)$: probabilidade de o exemplo ser da classe c_j dado que seu valor é x
- $p(x|c_j)$: prob. de um exemplo da classe c_j ter o valor x
- $p(c_j)$: prob. de um exemplo pertencer à classe c_j
- $p(x)$: prob de um exemplo ter o valor x

Classificador Bayesiano

- O classificador de Bayes ou **MAP** "aposta" na classe mais provável para o exemplo observado
 - MAP: *maximum a posteriori probability* ou probabilidade posterior/*a posteriori* máxima

$$\begin{aligned} f_{\text{MAP}}(x) &= \arg \max_{c_j \in \Omega} \{p(c_j|x)\} \\ &= \arg \max_{c_j \in \Omega} \left\{ \frac{p(x|c_j) \cdot p(c_j)}{p(x)} \right\} \end{aligned}$$

Podemos ignorar $p(x)$
porque é idêntico
para todas as classes

$$= \arg \max_{c_j \in \Omega} \{p(x|c_j) \cdot p(c_j)\}$$

Classificador Bayesiano

- Considerando que o exemplo é representado por um vetor de características
- Considerando cada característica como uma variável aleatória X_1, X_2, \dots, X_M
- Considerando um exemplo $\mathbf{x} = (x_1, x_2, \dots, x_M)$

Classificador Bayesiano

- Apresentando o **classificador ótimo de Bayes!**

$$f_{\text{MAP}}(\mathbf{x}) = \arg \max_{c_j \in \Omega} \{p(c_j | \mathbf{x})\}$$

$$= \arg \max_{c_j \in \Omega} \{p(\mathbf{x} | c_j) \cdot p(c_j)\}$$

$$= \arg \max_{c_j \in \Omega} \{p(x_1, x_2, \dots, x_M | c_j) \cdot p(c_j)\}$$

Classificador ótimo de Bayes

- O classificador ótimo de Bayes é o melhor modelo possível, no caso médio, para qualquer situação
- Ele viola o teorema do *no free lunch*?
 - Não, pois o classificador ótimo de Bayes considera a distribuição dos dados
 - Ele não é independente de contexto dos dados, nem do usuário
- Então por que não se utiliza o classificador ótimo de Bayes em aplicações práticas?

Classificador ótimo de Bayes

- Vamos fazer um experimento...
 - Conjunto de dados playtennis
 - Características: descrição do dia
 - Classe: decisão sobre ir ou não jogar tênis

Classificador ótimo de Bayes

dia	aparência	temperatura	umidade	vento	jogar
D1	ensolarado	quente	alta	fraco	não
D2	ensolarado	quente	alta	forte	não
D3	nublado	quente	alta	fraco	sim
D4	chuvoso	moderado	alta	fraco	sim
D5	chuvoso	frio	baixa	fraco	sim
D6	chuvoso	frio	baixa	forte	não
D7	nublado	frio	baixa	forte	sim
D8	ensolarado	moderado	alta	fraco	não
D9	ensolarado	frio	baixa	fraco	sim
D10	chuvoso	moderado	baixa	fraco	sim
D11	ensolarado	moderado	baixa	forte	sim
D12	nublado	moderado	alta	forte	sim
D13	nublado	quente	baixa	fraco	sim
D14	chuvoso	moderado	alta	forte	não

Dia 15: <ensolarado, moderado, alta, forte, ?>

Classificador ótimo de Bayes

- Para decidir se vamos ou não jogar tênis no dia D15, precisamos calcular as probabilidades condicionais:
 - $p(\text{sim} \mid \text{ensolarado, moderado, alta, forte})$
 - $p(\text{não} \mid \text{ensolarado, moderado, alta, forte})$
- Infelizmente não temos nenhum registro que nos permita calcular a probabilidade conjunta de todos esses eventos...

Classificador ingênuo de Bayes

- O classificador **ingênuo** de Bayes assume que todos os atributos são **independentes** entre si
 - Se duas variáveis X e Y são independentes, então
 - $p(X, Y) = p(X) \cdot P(Y)$
 - $p(Z | X, Y) = p(Z|X) \cdot P(Z|Y)$

Classificador ingênuo de Bayes

- Classificador ingênuo de Bayes
 - *Naive Bayes*
 - *Naïve Bayes*

$$f_{\text{NB}}(\mathbf{x}) = \arg \max_{c_j \in \Omega} \{p(x_1, x_2, \dots, x_M | c_j) \cdot p(c_j)\}$$

$$= \arg \max_{c_j \in \Omega} \left\{ p(c_j) \prod_{i=1}^M p(x_i | c_j) \right\}$$

Classificador ingênuo de Bayes

- Para decidir se vamos ou não jogar tênis no dia D15, precisamos calcular as probabilidades condicionais:
 - $p(\text{ensolarado}|\text{sim})$, $p(\text{moderado}|\text{sim})$, $p(\text{alta}|\text{sim})$, $p(\text{forte}|\text{sim})$
 - $p(\text{ensolarado}|\text{não})$, $p(\text{moderado}|\text{não})$, $p(\text{alta}|\text{não})$, $p(\text{forte}|\text{não})$

Classificador ingênuo de Bayes

dia	aparência	temperatura	umidade	vento	jogar
D1	ensolarado	quente	alta	fraco	não
D2	ensolarado	quente	alta	forte	não
D3	nublado	quente	alta	fraco	sim
D4	chuvoso	moderado	alta	fraco	sim
D5	chuvoso	frio	baixa	fraco	sim
D6	chuvoso	frio	baixa	forte	não
D7	nublado	frio	baixa	forte	sim
D8	ensolarado	moderado	alta	fraco	não
D9	ensolarado	frio	baixa	fraco	sim
D10	chuvoso	moderado	baixa	fraco	sim
D11	ensolarado	moderado	baixa	forte	sim
D12	nublado	moderado	alta	forte	sim
D13	nublado	quente	baixa	fraco	sim
D14	chuvoso	moderado	alta	forte	não

Dia 15: <ensolarado, moderado, alta, forte, ?>

Primeiro encontramos $p(\text{sim}) = 9/14$

Classificador ingênuo de Bayes

dia	aparência	temperatura	umidade	vento	jogar
D1	ensolarado	quente	alta	fraco	não
D2	ensolarado	quente	alta	forte	não
D3	nublado	quente	alta	fraco	sim
D4	chuvoso	moderado	alta	fraco	sim
D5	chuvoso	frio	baixa	fraco	sim
D6	chuvoso	frio	baixa	forte	não
D7	nublado	frio	baixa	forte	sim
D8	ensolarado	moderado	alta	fraco	não
D9	ensolarado	frio	baixa	fraco	sim
D10	chuvoso	moderado	baixa	fraco	sim
D11	ensolarado	moderado	baixa	forte	sim
D12	nublado	moderado	alta	forte	sim
D13	nublado	quente	baixa	fraco	sim
D14	chuvoso	moderado	alta	forte	não

Dia 15: <ensolarado, moderado, alta, forte, ?>

Depois $p(\text{ensolarado}|\text{sim}) = 2/9$

Classificador ingênuo de Bayes

dia	aparência	temperatura	umidade	vento	jogar
D1	ensolarado	quente	alta	fraco	não
D2	ensolarado	quente	alta	forte	não
D3	nublado	quente	alta	fraco	sim
D4	chuvoso	moderado	alta	fraco	sim
D5	chuvoso	frio	baixa	fraco	sim
D6	chuvoso	frio	baixa	forte	não
D7	nublado	frio	baixa	forte	sim
D8	ensolarado	moderado	alta	fraco	não
D9	ensolarado	frio	baixa	fraco	sim
D10	chuvoso	moderado	baixa	fraco	sim
D11	ensolarado	moderado	baixa	forte	sim
D12	nublado	moderado	alta	forte	sim
D13	nublado	quente	baixa	fraco	sim
D14	chuvoso	moderado	alta	forte	não

Dia 15: <ensolarado, moderado, alta, forte, ?>

$$p(\text{moderado} \mid \text{sim}) = 3/9$$

Classificador ingênuo de Bayes

dia	aparência	temperatura	umidade	vento	jogar
D1	ensolarado	quente	alta	fraco	não
D2	ensolarado	quente	alta	forte	não
D3	nublado	quente	alta	fraco	sim
D4	chuvoso	moderado	alta	fraco	sim
D5	chuvoso	frio	baixa	fraco	sim
D6	chuvoso	frio	baixa	forte	não
D7	nublado	frio	baixa	forte	sim
D8	ensolarado	moderado	alta	fraco	não
D9	ensolarado	frio	baixa	fraco	sim
D10	chuvoso	moderado	baixa	fraco	sim
D11	ensolarado	moderado	baixa	forte	sim
D12	nublado	moderado	alta	forte	sim
D13	nublado	quente	baixa	fraco	sim
D14	chuvoso	moderado	alta	forte	não

Dia 15: <ensolarado, moderado, alta, forte, ?>

$$p(\text{alta}|\text{sim}) = 3/9$$

Classificador ingênuo de Bayes

dia	aparência	temperatura	umidade	vento	jogar
D1	ensolarado	quente	alta	fraco	não
D2	ensolarado	quente	alta	forte	não
D3	nublado	quente	alta	fraco	sim
D4	chuvoso	moderado	alta	fraco	sim
D5	chuvoso	frio	baixa	fraco	sim
D6	chuvoso	frio	baixa	forte	não
D7	nublado	frio	baixa	forte	sim
D8	ensolarado	moderado	alta	fraco	não
D9	ensolarado	frio	baixa	fraco	sim
D10	chuvoso	moderado	baixa	fraco	sim
D11	ensolarado	moderado	baixa	forte	sim
D12	nublado	moderado	alta	forte	sim
D13	nublado	quente	baixa	fraco	sim
D14	chuvoso	moderado	alta	forte	não

Dia 15: <ensolarado, moderado, alta, forte, ?>

$$p(\text{forte}|\text{sim}) = 3/9$$

Classificador ingênuo de Bayes

dia	aparência	temperatura	umidade	vento	jogar
D1	ensolarado	quente	alta	fraco	não
D2	ensolarado	quente	alta	forte	não
D3	nublado	quente	alta	fraco	sim
D4	chuvoso	moderado	alta	fraco	sim
D5	chuvoso	frio	baixa	fraco	sim
D6	chuvoso	frio	baixa	forte	não
D7	nublado	frio	baixa	forte	sim
D8	ensolarado	moderado	alta	fraco	não
D9	ensolarado	frio	baixa	fraco	sim
D10	chuvoso	moderado	baixa	fraco	sim
D11	ensolarado	moderado	baixa	forte	sim
D12	nublado	moderado	alta	forte	sim
D13	nublado	quente	baixa	fraco	sim
D14	chuvoso	moderado	alta	forte	não

Dia 15: <ensolarado, moderado, alta, forte, ?>

Em seguida encontramos $p(\text{não}) = 5/14$

Classificador ingênuo de Bayes

dia	aparência	temperatura	umidade	vento	jogar
D1	ensolarado	quente	alta	fraco	não
D2	ensolarado	quente	alta	forte	não
D3	nublado	quente	alta	fraco	sim
D4	chuvoso	moderado	alta	fraco	sim
D5	chuvoso	frio	baixa	fraco	sim
D6	chuvoso	frio	baixa	forte	não
D7	nublado	frio	baixa	forte	sim
D8	ensolarado	moderado	alta	fraco	não
D9	ensolarado	frio	baixa	fraco	sim
D10	chuvoso	moderado	baixa	fraco	sim
D11	ensolarado	moderado	baixa	forte	sim
D12	nublado	moderado	alta	forte	sim
D13	nublado	quente	baixa	fraco	sim
D14	chuvoso	moderado	alta	forte	não

Dia 15: <ensolarado, moderado, alta, forte, ?>

$$p(\text{ensolarado}|\text{não}) = 3/5$$

Classificador ingênuo de Bayes

dia	aparência	temperatura	umidade	vento	jogar
D1	ensolarado	quente	alta	fraco	não
D2	ensolarado	quente	alta	forte	não
D3	nublado	quente	alta	fraco	sim
D4	chuvoso	moderado	alta	fraco	sim
D5	chuvoso	frio	baixa	fraco	sim
D6	chuvoso	frio	baixa	forte	não
D7	nublado	frio	baixa	forte	sim
D8	ensolarado	moderado	alta	fraco	não
D9	ensolarado	frio	baixa	fraco	sim
D10	chuvoso	moderado	baixa	fraco	sim
D11	ensolarado	moderado	baixa	forte	sim
D12	nublado	moderado	alta	forte	sim
D13	nublado	quente	baixa	fraco	sim
D14	chuvoso	moderado	alta	forte	não

Dia 15: <ensolarado, moderado, alta, forte, ?>

$$p(\text{moderado}|\text{não}) = 2/5$$

Classificador ingênuo de Bayes

dia	aparência	temperatura	umidade	vento	jogar
D1	ensolarado	quente	alta	fraco	não
D2	ensolarado	quente	alta	forte	não
D3	nublado	quente	alta	fraco	sim
D4	chuvoso	moderado	alta	fraco	sim
D5	chuvoso	frio	baixa	fraco	sim
D6	chuvoso	frio	baixa	forte	não
D7	nublado	frio	baixa	forte	sim
D8	ensolarado	moderado	alta	fraco	não
D9	ensolarado	frio	baixa	fraco	sim
D10	chuvoso	moderado	baixa	fraco	sim
D11	ensolarado	moderado	baixa	forte	sim
D12	nublado	moderado	alta	forte	sim
D13	nublado	quente	baixa	fraco	sim
D14	chuvoso	moderado	alta	forte	não

Dia 15: <ensolarado, moderado, alta, forte, ?>

$$p(\text{alta}|\text{não}) = 4/5$$

Classificador ingênuo de Bayes

dia	aparência	temperatura	umidade	vento	jogar
D1	ensolarado	quente	alta	fraco	não
D2	ensolarado	quente	alta	forte	não
D3	nublado	quente	alta	fraco	sim
D4	chuvoso	moderado	alta	fraco	sim
D5	chuvoso	frio	baixa	fraco	sim
D6	chuvoso	frio	baixa	forte	não
D7	nublado	frio	baixa	forte	sim
D8	ensolarado	moderado	alta	fraco	não
D9	ensolarado	frio	baixa	fraco	sim
D10	chuvoso	moderado	baixa	fraco	sim
D11	ensolarado	moderado	baixa	forte	sim
D12	nublado	moderado	alta	forte	sim
D13	nublado	quente	baixa	fraco	sim
D14	chuvoso	moderado	alta	forte	não

Dia 15: <ensolarado, moderado, alta, forte, ?>

$$p(\text{forte}|\text{não}) = 3/5$$

Classificador ingênuo de Bayes

- Vamos jogar nesse dia?

$$p(\text{sim}) \cdot \prod_{i=1}^4 p(x_i|\text{sim}) = \frac{9}{14} \cdot \frac{2}{9} \cdot \frac{3}{9} \cdot \frac{3}{9} \cdot \frac{3}{9} = 0,005$$

$$p(\text{não}) \cdot \prod_{i=1}^4 p(x_i|\text{não}) = \frac{5}{14} \cdot \frac{3}{5} \cdot \frac{2}{5} \cdot \frac{4}{5} \cdot \frac{3}{5} = 0,042$$

- Não!

Indução e uso do modelo

- Os parâmetros do modelo são simplesmente as probabilidades posteriores $p(x_i | c_j)$ e as probabilidades *a priori* $p(c_j)$
 - Treinamento: calcular e armazenar os valores de probabilidades
 - Classificação de um novo exemplo: calcular as probabilidades para os atributos e classes relevante e escolher o argumento (classe) que maximiza a probabilidade posterior

Situações particulares do Naive Bayes

- O que fazer em caso de atributos ausentes?
- E em caso de valores numéricos?