

Inteligência Artificial

Aula 12

- Abordagens Bayesianas
- Naive Bayes

Prof. Ricardo M. Marcacini
ricardo.marcacini@ufms.br

Curso: Sistemas de Informação

1º Semestre / 2015

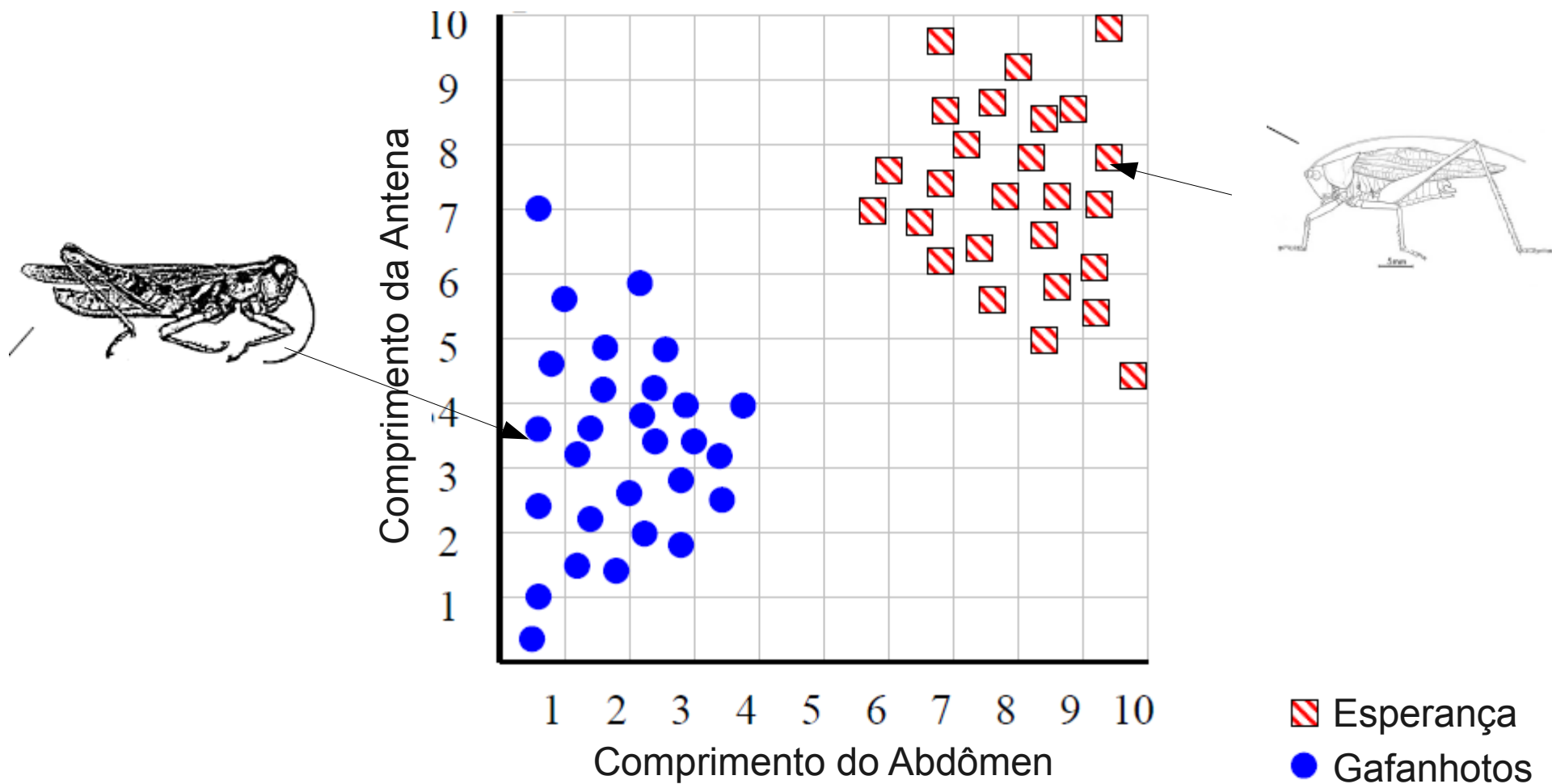
<http://moodle.lives.net.br/>



FUNDAÇÃO
UNIVERSIDADE FEDERAL
DE MATO GROSSO DO SUL

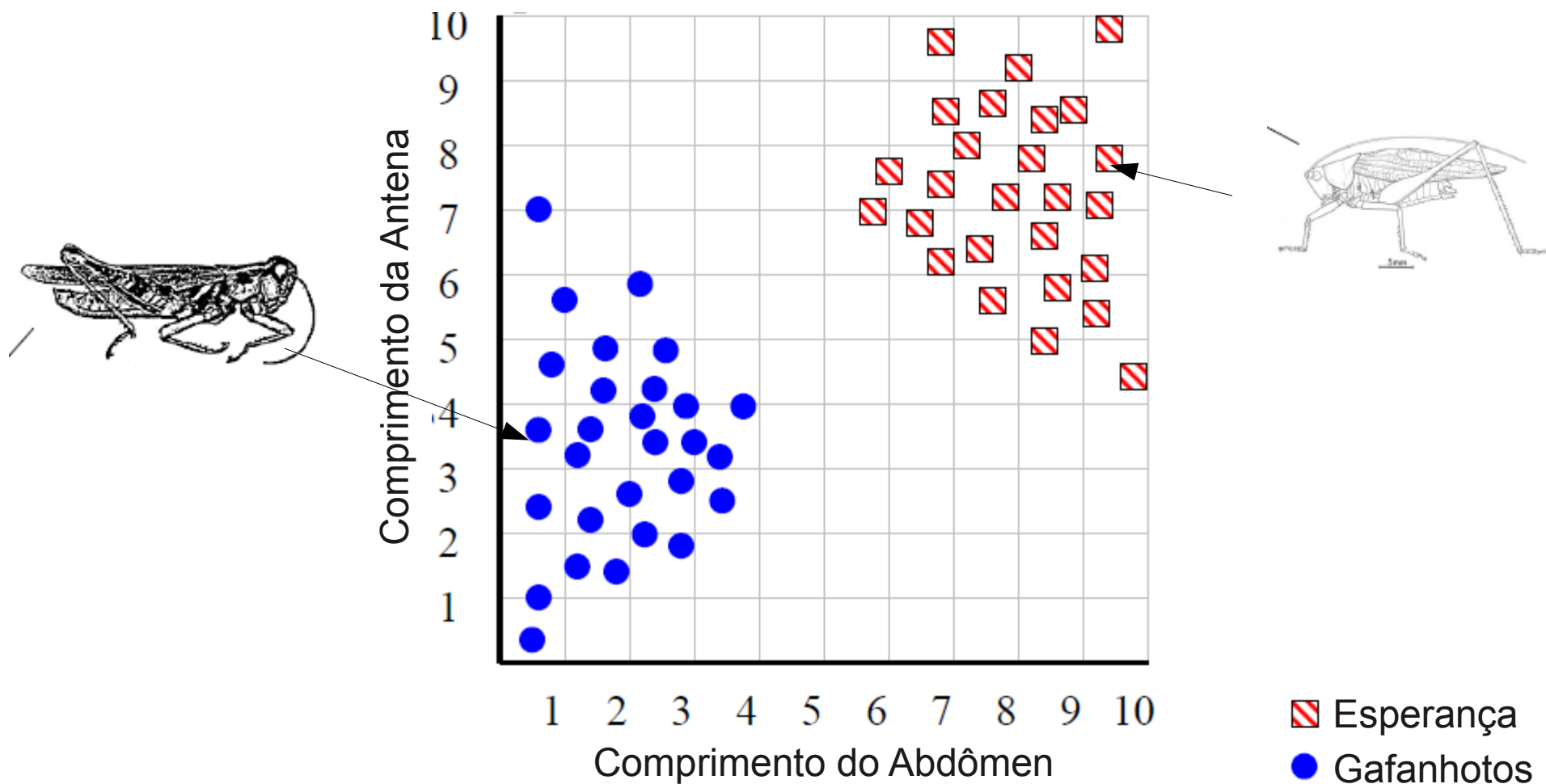
Aprendizado de Máquina: Abordagens Bayesianas

- Vamos começar com um exemplo...



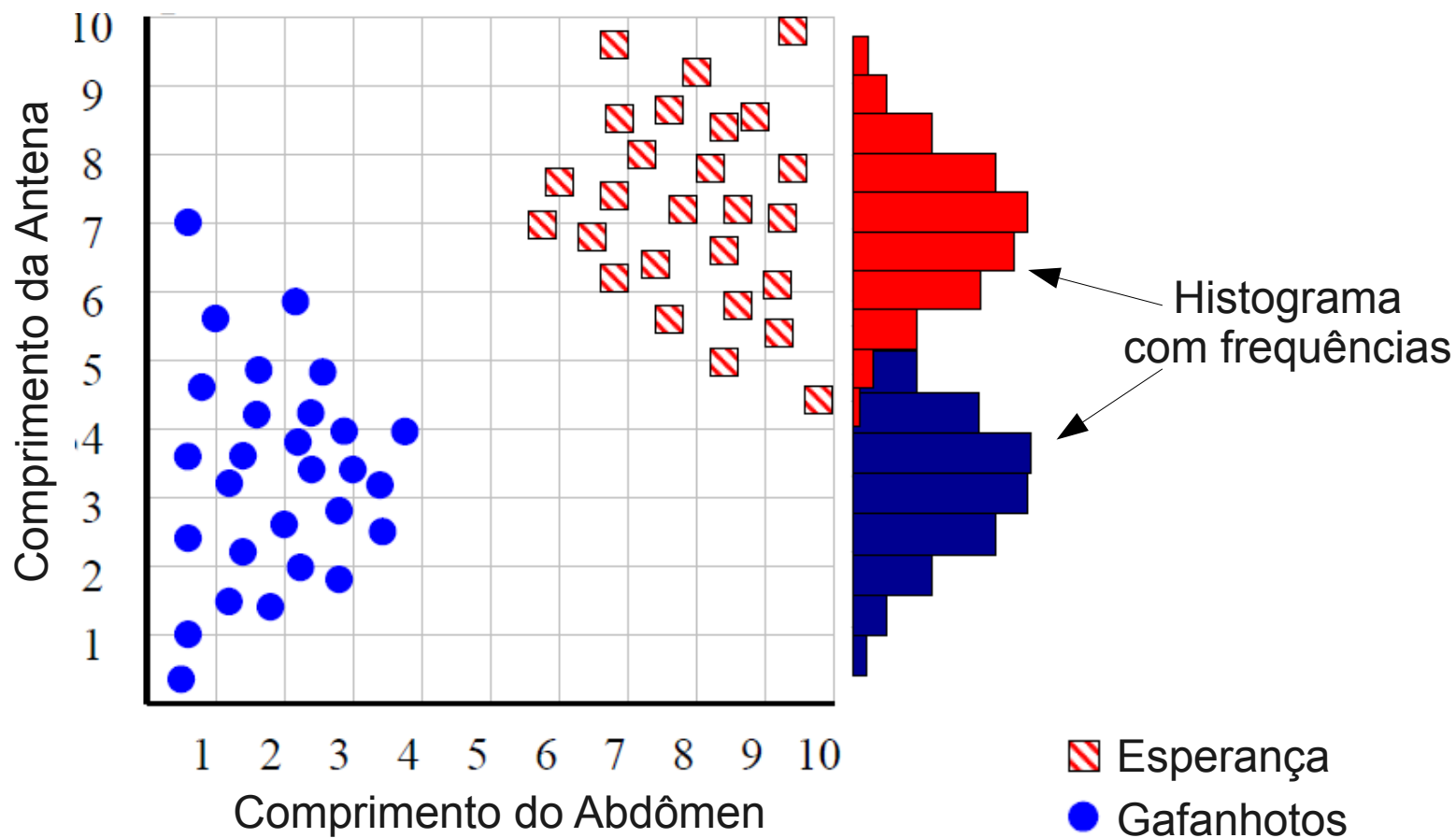
Aprendizado de Máquina: Abordagens Bayesianas

■ Diferença entre Esperanças e Gafanhotos



Aprendizado de Máquina: Abordagens Bayesianas

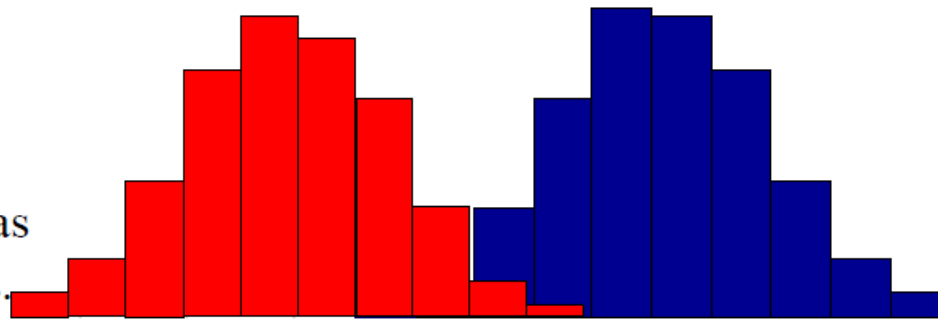
- Por enquanto, vamos olhar apenas para o comprimento da antena



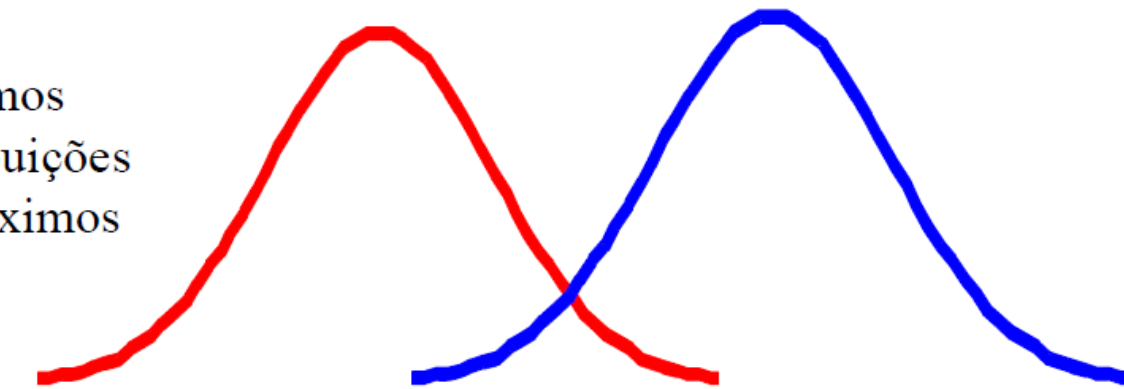
Aprendizado de Máquina: Abordagens Bayesianas

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

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

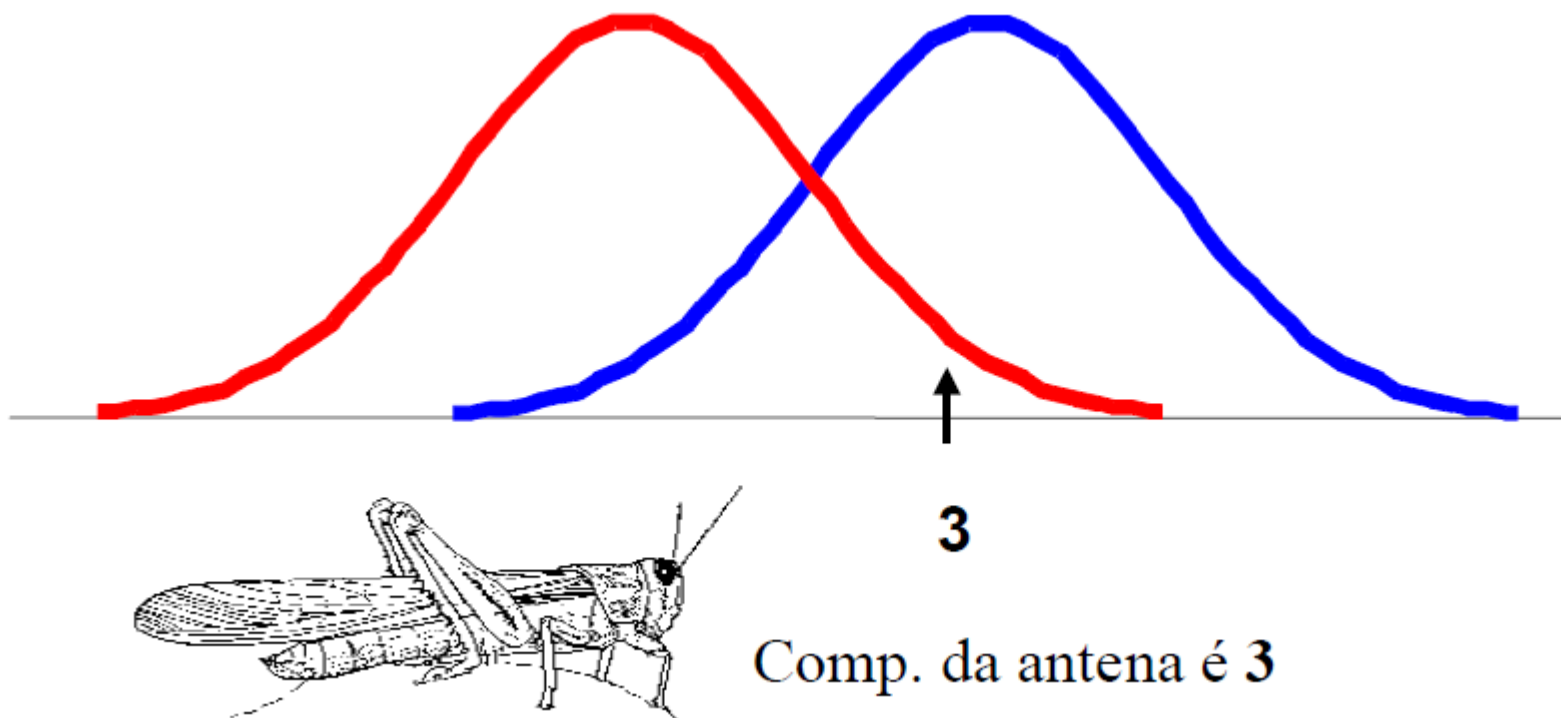


Para facilitar a visualização vamos usar duas distribuições normais nos próximos slides...



Aprendizado de Máquina: Abordagens Bayesianas

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

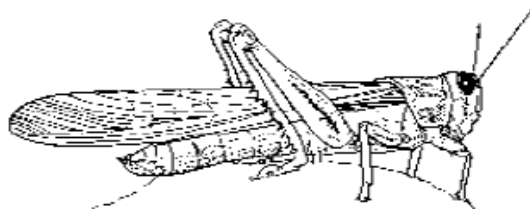
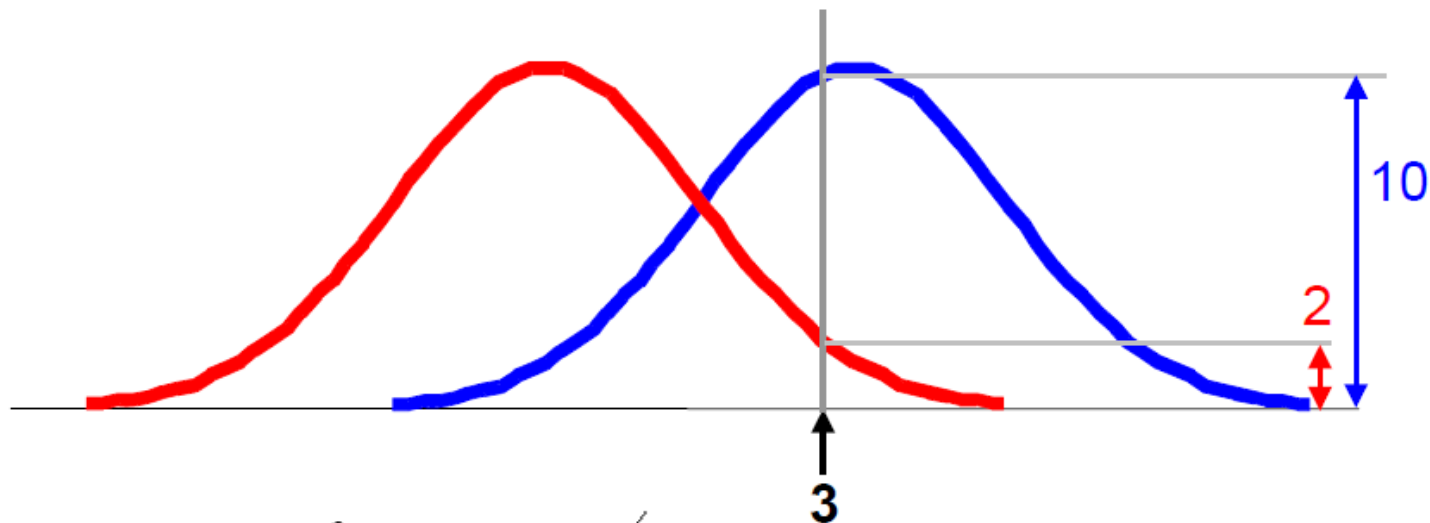


Aprendizado de Máquina: Abordagens Bayesianas

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

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

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



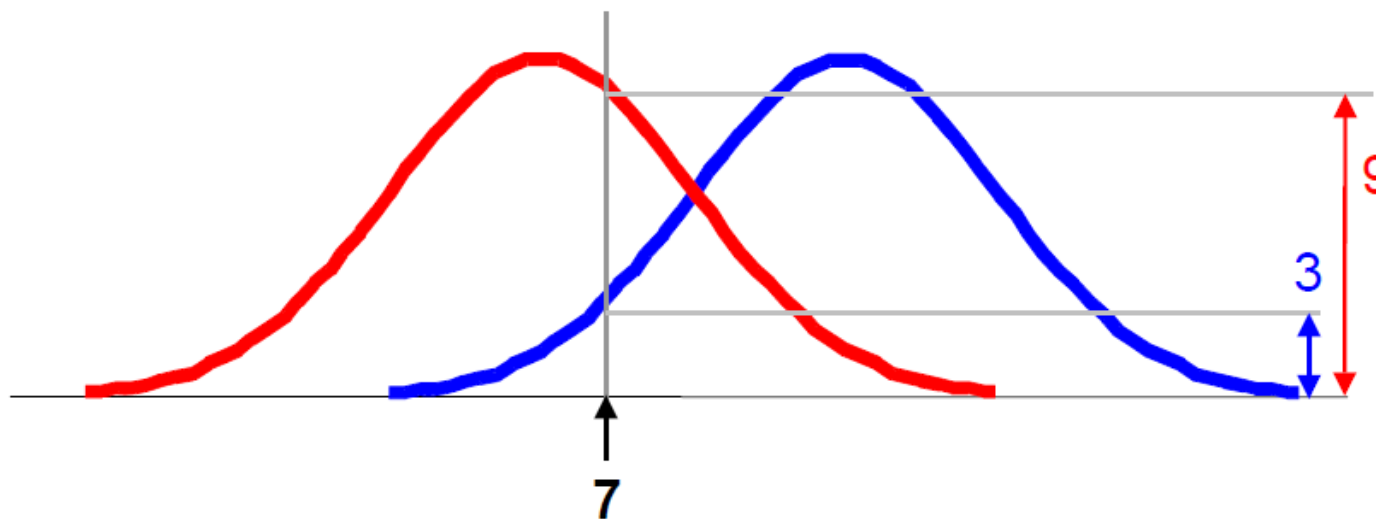
Comp. da antena é 3

Aprendizado de Máquina: Abordagens Bayesianas

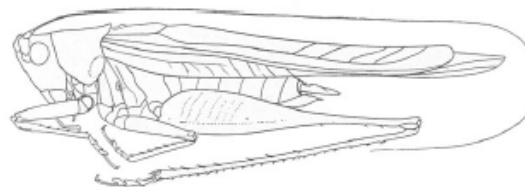
$p(c_j | d)$ = probabilidade da classe c_j , dado que observamos d

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

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



Comp. da antena é 7

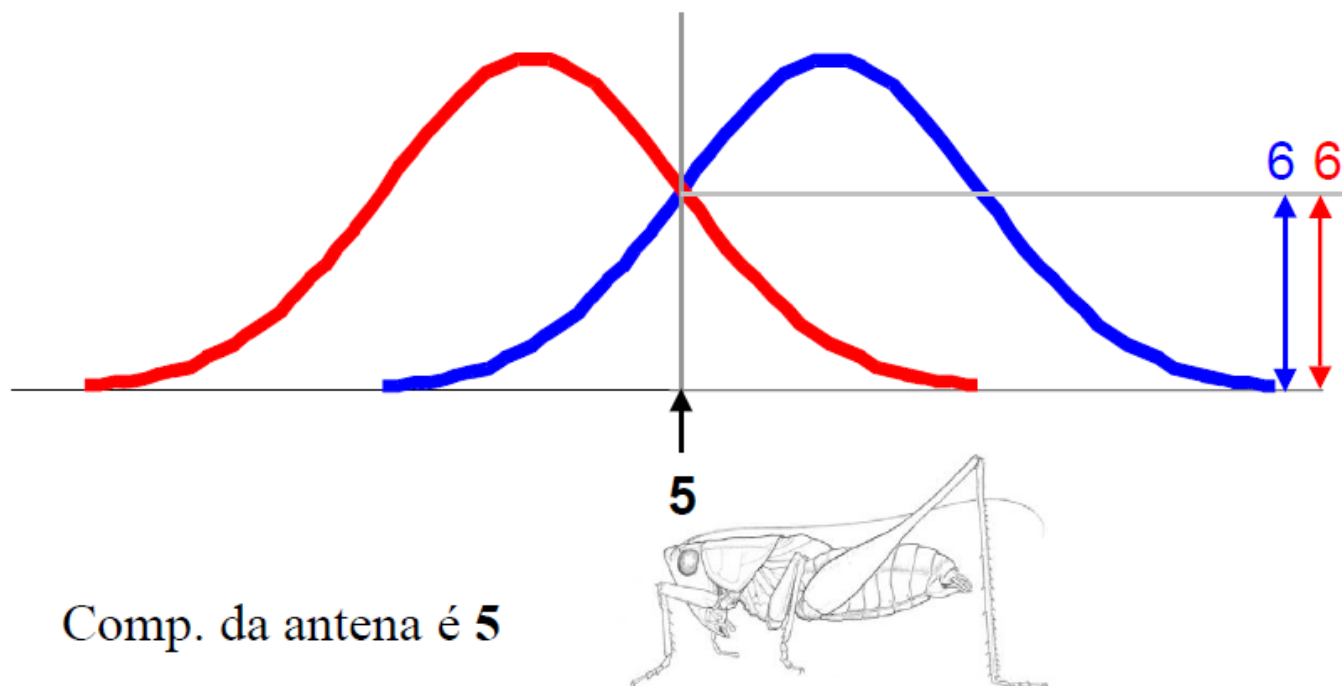


Aprendizado de Máquina: Abordagens Bayesianas

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

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

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



Naive Bayes

- Acamos de realizar uma classificação “intuitiva” com uso do Naive Bayes
 - Bayes Ingênuo
 - Bayes Simples
- Baseado no Teorema de Bayes

$$p(c_j | d) = \frac{p(d | c_j) p(c_j)}{p(d)}$$



Thomas Bayes
1702 - 1761

Naive Bayes

■ Teorema de Bayes

$$p(c_j | d) = \frac{p(d | c_j) p(c_j)}{p(d)}$$

$p(c_j | d)$ = probabilidade do exemplo d ser da classe c_j

Esse é o objetivo da tarefa de aprendizado (o que desejamos computar)

$p(d | c_j)$ = probabilidade de gerar exemplo d dada a classe c_j .

Ser da classe c_j leva a ter a característica d com alguma probabilidade

$p(c_j)$ = probabilidade de ocorrência da classe c_j .

Representa a frequência da classe c_j em nossa base de dados

$p(d)$ = probabilidade de exemplo d ocorrer

Pode ser ignorado, pois é uma constante para todas as classes

Naive Bayes

- Considerando o Teorema de Bayes, um novo exemplo será atribuído à sua classe mais provável
 - $\arg \max$ retorna a classe v_j com maior probabilidade
 - a_1, a_2, \dots, a_n são as características

$$v_{map} = \arg \max_{v_j \in V} P(v_j | a_1, a_2, \dots, a_n)$$

$$v_{map} = \arg \max_{v_j \in V} \frac{P(a_1, a_2, \dots, a_n | v_j) P(v_j)}{P(a_1, a_2, \dots, a_n)}$$

$$v_{map} = \arg \max_{v_j \in V} P(a_1, a_2, \dots, a_n | v_j) P(v_j)$$

Naive Bayes

- Assim, para determinar a classe de um novo exemplo basta computar:

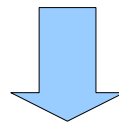
$$v_{map} = \arg \max_{v_j \in V} P(a_1, a_2, \dots, a_n | v_j) P(v_j)$$

- v_j é a frequência da classe j
- No entanto, computar $P(a_1, a_2, \dots, a_n | v_j)$ é inviável na prática
- O que fazer?

Naive Bayes

- Naive Bayes considera que todas as características são independentes
 - Suposição ingênua
 - Não é realidade na prática
 - O que motivou o seu nome (naive)
- O problema original é transformado em:

$$v_{map} = \arg \max_{v_j \in V} P(a_1, a_2, \dots, a_n | v_j) P(v_j)$$

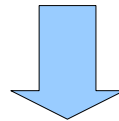


$$v = \arg \max_{v_j \in V} P(v_j) \prod_i P(a_i | v_j)$$

Naive Bayes

- Naive Bayes considera que todas as características são independentes
 - Suposição ingênua
 - Esse é o motivo do seu nome (“naive”)
- O problema original é transformado em:

$$v_{map} = \arg \max_{v_j \in V} P(a_1, a_2, \dots, a_n | v_j) P(v_j)$$



$$v = \arg \max_{v_j \in V} P(v_j) \prod_i P(a_i | v_j)$$

Nós sabemos computar essa probabilidade.

Lembra do Gafanhoto e Esperança?

Naive Bayes

- O processo de aprendizado bayesiano consiste em estimar $P(v_j)$ e $P(a_i | v_j)$
 - A partir de todo o conjunto de dados (treinamento)

Naive Bayes

- O processo de aprendizado bayesiano consiste em estimar $P(v_j)$ e $P(a_i | v_j)$
 - A partir de todo o conjunto de dados (treinamento)
- As probabilidades aprendidas são utilizadas para classificar novos exemplos
- Quando Naive Bayes apresenta um resultado ótimo?

Naive Bayes

- O processo de aprendizado bayesiano consiste em estimar $P(v_j)$ e $P(a_i | v_j)$
 - A partir de todo o conjunto de dados (treinamento)
- As probabilidades aprendidas são utilizadas para classificar novos exemplos
- Quando Naive Bayes apresenta um resultado ótimo?
 - Se as características forem independentes
 - Muito difícil na prática!

Naive Bayes - Exemplo

■ Problema: Jogar Tênis?

dia	aparência	temperatura	umidade	vento	jogar_tênis
D1	ensolarado	quente	alta	fraco	não
D2	ensolarado	quente	alta	forte	não
D3	nublado	quente	alta	fraco	sim
D4	chuva	moderada	alta	fraco	sim
D5	chuva	fria	normal	fraco	sim
D6	chuva	fria	normal	forte	não
D7	nublado	fria	normal	forte	sim
D8	ensolarado	moderada	alta	fraco	não
D9	ensolarado	fria	normal	fraco	sim
D10	chuva	moderada	normal	fraco	sim
D11	ensolarado	moderada	normal	forte	sim
D12	nublado	moderada	alta	forte	sim
D13	nublado	quente	normal	fraco	sim
D14	chuva	moderada	alta	forte	não

Naive Bayes - Exemplo

■ Problema: Jogar Tênis?

■ Conjunto de Treinamento

dia	aparência	temperatura	umidade	vento	jogar_tênis
D1	ensolarado	quente	alta	fraco	não
D2	ensolarado	quente	alta	forte	não
D3	nublado	quente	alta	fraco	sim
D4	chuva	moderada	alta	fraco	sim
D5	chuva	fria	normal	fraco	sim
D6	chuva	fria	normal	forte	não
D7	nublado	fria	normal	forte	sim
D8	ensolarado	moderada	alta	fraco	não
D9	ensolarado	fria	normal	fraco	sim
D10	chuva	moderada	normal	fraco	sim
D11	ensolarado	moderada	normal	forte	sim
D12	nublado	moderada	alta	forte	sim
D13	nublado	quente	normal	fraco	sim
D14	chuva	moderada	alta	forte	não

■ Novo exemplo (classificar como “sim” ou “não”)

<aparencia=ensolarado, temperatura=moderada, umidade=alta, vento=forte>

Naive Bayes - Exemplo

■ Problema: Jogar Tênis?

■ Instanciando Bayes...

$$v = \operatorname{argmax}_{v_j \in V} P(v_j) \prod_i P(a_i | v_j)$$

$$v = \operatorname{argmax}_{v_j \in \{\text{sim}, \text{não}\}} P(v_j) * P(\text{aparência} = \text{ensolarado} | v_j) *$$

$$P(\text{temperatura} = \text{moderada} | v_j) * P(\text{umidade} = \text{alta} | v_j) *$$

$$P(\text{vento} = \text{forte} | v_j)$$

■ Quais probabilidades precisamos computar?

Naive Bayes - Exemplo

- Problema: Jogar Tênis?
 - Probabilidades que precisamos computar
 - Probabilidade $P(v_j)$ das classes:
 $P(\text{sim})$ e $P(\text{não})$
 - Cada probabilidade $P(a_i | v_j)$ para as duas possíveis classes
 $P(\text{aparência}=\text{ensolarado} | \text{sim})$ e $P(\text{aparência}=\text{ensolarado} | \text{não})$
 $P(\text{temperatura}=\text{moderada} | \text{sim})$ e $P(\text{temperatura}=\text{moderada} | \text{não})$
 $P(\text{umidade}=\text{alta} | \text{sim})$ e $P(\text{umidade}=\text{alta} | \text{não})$
 $P(\text{vento}=\text{forte} | \text{sim})$ e $P(\text{vento}=\text{forte} | \text{não})$
- Exemplos...

Naive Bayes - Exemplo

dia	aparência	temperatura	umidade	vento	jogar_tênis
D1	ensolarado	quente	alta	fraco	não
D2	ensolarado	quente	alta	forte	não
D3	nublado	quente	alta	fraco	sim
D4	chuva	moderada	alta	fraco	sim
D5	chuva	fria	normal	fraco	sim
D6	chuva	fria	normal	forte	não
D7	nublado	fria	normal	forte	sim
D8	ensolarado	moderada	alta	fraco	não
D9	ensolarado	fria	normal	fraco	sim
D10	chuva	moderada	normal	fraco	sim
D11	ensolarado	moderada	normal	forte	sim
D12	nublado	moderada	alta	forte	sim
D13	nublado	quente	normal	fraco	sim
D14	chuva	moderada	alta	forte	não

$$P(\text{sim}) = 9 / 14$$

Naive Bayes - Exemplo

dia	aparência	temperatura	umidade	vento	jogar_tênis
D1	ensolarado	quente	alta	fraco	não
D2	ensolarado	quente	alta	forte	não
D3	nublado	quente	alta	fraco	sim
D4	chuva	moderada	alta	fraco	sim
D5	chuva	fria	normal	fraco	sim
D6	chuva	fria	normal	forte	não
D7	nublado	fria	normal	forte	sim
D8	ensolarado	moderada	alta	fraco	não
D9	ensolarado	fria	normal	fraco	sim
D10	chuva	moderada	normal	fraco	sim
D11	ensolarado	moderada	normal	forte	sim
D12	nublado	moderada	alta	forte	sim
D13	nublado	quente	normal	fraco	sim
D14	chuva	moderada	alta	forte	não

$$P(\text{não}) = 5 / 14$$

Naive Bayes - Exemplo

dia	aparência	temperatura	umidade	vento	jogar_tênis
D1	ensolarado	quente	alta	fraco	não
D2	ensolarado	quente	alta	forte	não
D3	nublado	quente	alta	fraco	sim
D4	chuva	moderada	alta	fraco	sim
D5	chuva	fria	normal	fraco	sim
D6	chuva	fria	normal	forte	não
D7	nublado	fria	normal	forte	sim
D8	ensolarado	moderada	alta	fraco	não
D9	ensolarado	fria	normal	fraco	sim
D10	chuva	moderada	normal	fraco	sim
D11	ensolarado	moderada	normal	forte	sim
D12	nublado	moderada	alta	forte	sim
D13	nublado	quente	normal	fraco	sim
D14	chuva	moderada	alta	forte	não

$$P(\text{vento=forte} \mid \text{sim}) = 3/9$$

Naive Bayes - Exemplo

dia	aparência	temperatura	umidade	vento	jogar_tênis
D1	ensolarado	quente	alta	fraco	não
D2	ensolarado	quente	alta	forte	não
D3	nublado	quente	alta	fraco	sim
D4	chuva	moderada	alta	fraco	sim
D5	chuva	fria	normal	fraco	sim
D6	chuva	fria	normal	forte	não
D7	nublado	fria	normal	forte	sim
D8	ensolarado	moderada	alta	fraco	não
D9	ensolarado	fria	normal	fraco	sim
D10	chuva	moderada	normal	fraco	sim
D11	ensolarado	moderada	normal	forte	sim
D12	nublado	moderada	alta	forte	sim
D13	nublado	quente	normal	fraco	sim
D14	chuva	moderada	alta	forte	não

$$P(\text{vento=forte} \mid \text{não}) = 3/5$$

Naive Bayes - Exemplo

■ Problema: Jogar Tênis?

- Após computar todas as probabilidades, como podemos classificar o novo exemplo?

<aparência=ensolarado, temperatura=moderada, umidade=alta, vento=forte>

- $P(\text{sim}) * P(\text{aparência=ensolarado} \mid \text{sim}) * P(\text{temperatura=moderada} \mid \text{sim}) * P(\text{umidade=alta} \mid \text{sim}) * P(\text{vento=forte} \mid \text{sim}) = 0.0070$
- $P(\text{não}) * P(\text{aparência=ensolarado} \mid \text{não}) * P(\text{temperatura=moderada} \mid \text{não}) * P(\text{umidade=alta} \mid \text{não}) * P(\text{vento=forte} \mid \text{não}) = 0.0411$

Naive Bayes - Exemplo

■ Problema: Jogar Tênis?

- Após computar todas as probabilidades, como podemos classificar o novo exemplo?

<aparência=ensolarado, temperatura=moderada, umidade=alta, vento=forte>

■ Normalizando os resultados:

- $SIM = 0.0070 / (0.0070 + 0.0411) = 0.145$ (14.5%)
- $NÃO = 0.0411 / (0.0070 + 0.0411) = 0.855$ (85.5%)

- A resposta é NÃO jogar Tênis

Exercício

■ Classificar o exemplo:

<aparência=chuva, temperatura=fria, umidade=alta, vento=forte>

dia	aparência	temperatura	umidade	vento	jogar_tênis
D1	ensolarado	quente	alta	fraco	não
D2	ensolarado	quente	alta	forte	não
D3	nublado	quente	alta	fraco	sim
D4	chuva	moderada	alta	fraco	sim
D5	chuva	fria	normal	fraco	sim
D6	chuva	fria	normal	forte	não
D7	nublado	fria	normal	forte	sim
D8	ensolarado	moderada	alta	fraco	não
D9	ensolarado	fria	normal	fraco	sim
D10	chuva	moderada	normal	fraco	sim
D11	ensolarado	moderada	normal	forte	sim
D12	nublado	moderada	alta	forte	sim
D13	nublado	quente	normal	fraco	sim
D14	chuva	moderada	alta	forte	não