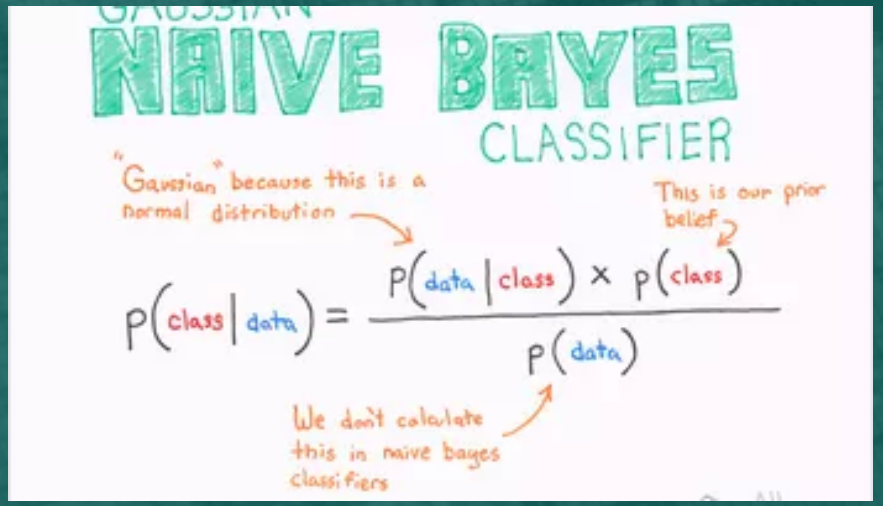
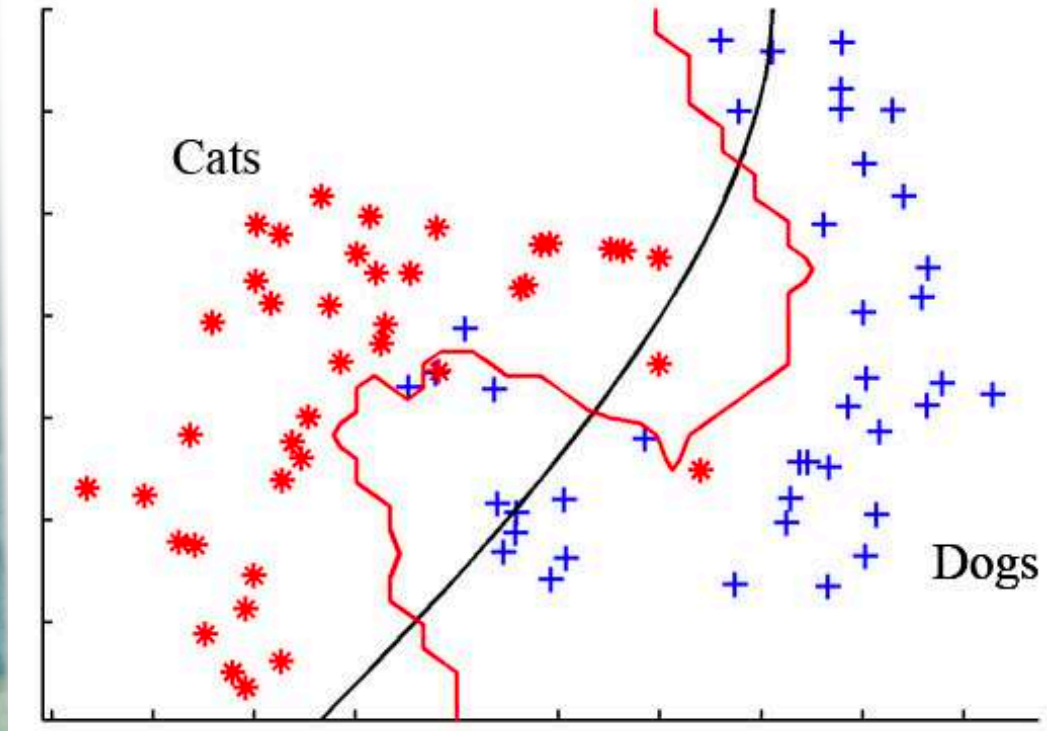


Naïve Bayes

Prof. Dr. Vladimir Costa de Alencar
UEPB
www.valencar.com

Naive Bayes

Classificação



Detecção de SPAM
Organização automática de e-mails
Identificação de páginas com conteúdo adulto
Detecção de expressões multipalavras

Naive Bayes

Aplicações do Algoritmo Naive Bayes

- **Previsões multi-classes**
- **Classificação de textos/Filtragem de spam/Análise de sentimento**
- **Previsões em tempo real**
- **Sistema de Recomendação**

Teorema de Bayes

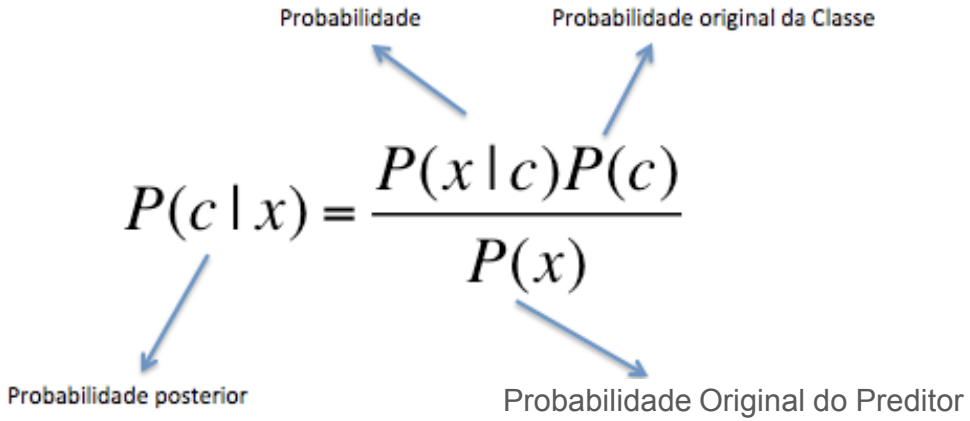
A regra de Bayes mostra como alterar as probabilidades a priori tendo em conta novas evidências de forma a obter probabilidades a posteriori.

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

Pr(A) e Pr(B) são as probabilidades a priori de A e B.

Pr(B|A) e Pr(A|B) são as probabilidades a posteriori de B condicional a A e de A condicional a B respectivamente.

Teorema de Bayes



$$P(c|X) = P(x_1|c) \times P(x_2|c) \times \ldots \times P(x_n|c) \times P(c)$$

- P (c | x) é a probabilidade posterior da classe (c, alvo) dada o preditor (x, atributos).
- P (c) é a probabilidade original da classe.
- P (x | c) é a probabilidade do preditor dada a classe.
- P (x) é a probabilidade original do preditor.

Teorema de Bayes

Em teoria da probabilidade o Teorema de Bayes mostra a relação entre uma probabilidade condicional e a sua inversa.

TEMPO	"PLAY"
Sol	Não
Nublado	Sim
Chuva	Sim
Sol	Sim
Sol	Sim
Nublado	Sim
Chuva	Não
Chuva	Não
Sol	Sim
Chuva	Sim
Sol	Não
Nublado	Sim
Nublado	Sim
Chuva	Não

Teorema de Bayes

Passo 1: Converter o conjunto de dados em uma tabela de frequência

TEMPO	"PLAY"
Sol	Não
Nublado	Sim
Chuva	Sim
Sol	Sim
Sol	Sim
Nublado	Sim
Chuva	Não
Chuva	Não
Sol	Sim
Chuva	Sim
Sol	Não
Nublado	Sim
Nublado	Sim
Chuva	Não

Tabela de Frequência		
Clima	Não	Sim
Nublado	0	4
Sol	2	3
Chuva	3	2
Total	**5**	**9**

Teorema de Bayes

Passo 2: Criar tabela de Probabilidade para encontrar as probabilidades de cada ocorrência e de cada combinação.

TEMPO	"PLAY"
Sol	Não
Nublado	Sim
Chuva	Sim
Sol	Sim
Sol	Sim
Nublado	Sim
Chuva	Não
Chuva	Não
Sol	Sim
Chuva	Sim
Sol	Não
Nublado	Sim
Nublado	Sim
Chuva	Não

Tabela de Frequência		
Clima	Não	Sim
Nublado	0	4
Sol	2	3
Chuva	3	2
Total	**5**	**9**

=5/14	=9/14
0.36	0.64

=4/14	0.29
=5/14	0.36
=5/14	0.36

Teorema de Bayes

Passo 3: Usamos a equação do Teorema de Bayes para calcular a probabilidade posterior para cada classe. A classe com maior probabilidade posterior é o resultado da previsão.

TEMPO	"PLAY"
Sol	Não
Nublado	Sim
Chuva	Sim
Sol	Sim
Sol	Sim
Nublado	Sim
Chuva	Não
Chuva	Não
Sol	Sim
Chuva	Sim
Sol	Não
Nublado	Sim
Nublado	Sim
Chuva	Não

Tabela de Frequência		
Clima	Não	Sim
Nublado	0	4
Sol	2	3
Chuva	3	2
Total	**5**	**9**
	=5/14	=9/14
	0.36	0.64

=4/14	0.29
=5/14	0.36
=5/14	0.36

$$P(c \mid x) = \frac{P(x \mid c)P(c)}{P(x)}$$

Teorema de Bayes

Os jogadores irão praticar esporte se o tempo estiver ensolarado.
Esta afirmação está correta?

Tabela de Frequência		
Clima	Não	Sim
Nublado	0	4
Sol	2	3
Chuva	3	2
Total	**5**	**9**

=4/14	0.29
=5/14	0.36
=5/14	0.36

=5/14	=9/14
0.36	0.64

$$P(Sim \mid Sol) = \frac{P(Sol \mid Sim) * P(Sim)}{P(Sol)}$$

Teorema de Bayes

Os jogadores irão praticar esporte se o tempo estiver ensolarado.
Esta afirmação está correta?

Tabela de Frequência		
Clima	Não	Sim
Nublado	0	4
Sol	2	3
Chuva	3	2
Total	**5**	**9**

Aqui temos:

P (Sol | Sim) = 3/9 = 0.33
P (Sol) = 5/14 = 0.36
P (Sim) = 9/14 = 0.64

$$P(Sim \mid Sol) = \frac{P(Sol \mid Sim) * P(Sim)}{P(Sol)}$$

P (Sim | Sol) = 0.33 * 0.64 / 0.36 = 0.60

P (Sim | Sol) = 0.60 = Os Jogadores Irão praticar esporte

Teorema de Bayes

Os jogadores irão praticar esporte se o tempo estiver ensolarado.
Esta afirmação está correta? ***Sim, a afirmação está correta!***

Agora, é só colocar na fórmula:

$$P(Sim \mid Sol) = P(Sol \mid Sim) * P(Sim) / P(Sol)$$

$$P(Sim \mid Sol) = 0.33 * 0.64 / 0.36 = 0.60$$

Os jogadores irão praticar esporte se o tempo estiver ensolarado. Esta afirmação está correta?

Classe condicional (independente) Entradas Fn

- Sol
- Nuvem
- Chuva

C: Jogar

1.Comece com dados brutos

Tempo	Jogar
Sol	Não
Nublado	Sim
Chuva	Sim
Sol	Sim
Sol	Sim
Nublado	Sim
Chuva	Não
Chuva	Não
Sol	Sim
Chuva	Sim
Sol	Não
Nublado	Sim
Nublado	Sim
Chuva	Não

2. Criar a tabela de frequência a partir dos dados brutos

Tabela de Frequência		
Clima	**Não**	**Sim**
Nublado	0	4
Sol	2	3
Chuva	3	2
Total	5	9

3. Criar uma tabela de probabilidade da tabela de frequência

Tabela de Probabilidade				
Clima	**Não**	**Sim**		
Nublado	0	4	= 4/14	0,29
Sol	2	3	=5/14	0,36
Chuva	3	2	=5/14	0,36
Total	5	9		
	=5/14	9/14		
	0,36	0,64		

4. Usar a equação de Bayes para prever a probabilidade posterior

Teorema de Bayes:

$P(A \mid B) = (P(B \mid A) * P(A)) / P(B)$

Teorema de Bayes modificado para o problema:

$P(Sim \mid Sol) = (P(Sol \mid Sim) * P(Sim)) / P(Sol)$

Então temos:

$P(Sol \mid Sim) = 3/9 = 0.33$

$P(Sol) = 5/14 = 0.36$

$P(Sim) = 9/14 = 0.64$

Logo

$P(Sim|Sol) = P(Sol|Sim) * P(Sim) / P(Sol)$

$P(Sim|Sol) = (0,33*0,64)/0,36 = 0,60$

Os jogadores irão praticar esporte se o tempo estiver ensolarado. ***Sim, afirmação está correta!***