

Aprendizagem Automática

Laboratório 4: **Classificadores de *Bayes***

N.º:90007 Nome: Alice Rosa

N.º:90026 Nome: Aprígio Malveiro

Turno: 3^a feira - 14h00

1 Classificadores de *Bayes*

1.1 Descrição dos classificadores de *Bayes* e *Naive Bayes*

O classificador de *Bayes* é utilizado para determinar, a partir de uma dada observação x , a classe y a que pertence de um conjunto de K classes possíveis, $\Omega = \{\omega_0, \dots, \omega_{K-1}\}$.

A função de perda, $L(y, \hat{y})$ atribui uma penalidade quando a verdadeira classe de x é y e a classe prevista é \hat{y} . Se a função de perda for binária, onde se tem,

$$L(y, \hat{y}) = \begin{cases} 0 & \text{se } \hat{y} = y \\ 1 & \text{caso contrário} \end{cases} \quad (1)$$

o classificador ótimo é dado por,

$$\hat{y} = \arg \max_{\omega \in \Omega} P(\omega|x). \quad (2)$$

Este é conhecido como o classificador de *Bayes* e escolhe a classe com a maior probabilidade *a posteriori*, isto é, a mais provável dado as observações x .

A probabilidade de cada classe, $P(\omega_i|x)$, conseguem ser obtidas pela Lei de Bayes,

$$P(\omega_i|x) = \frac{p(x|\omega_i)P(\omega_i)}{p(x)} = \frac{p(x|\omega_i)P(\omega_i)}{\sum_{\omega \in \Omega} p(x|\omega)P(\omega)}. \quad (3)$$

As probabilidades não normalizadas $p(x, \omega_i) = p(x|\omega_i)P(\omega_i)$ são suficientes para encontrar a classe mais provável, uma vez que, a normalização do factor $p(x)$ não depende da classe.

Quando se tem um número elevado de características (*features*), $x = [x_1, \dots, x_p]^T$, a estimativa da distribuição $p(x|\omega_k)$ é difícil de obter. O classificador de *Naive Bayes* simplifica o problema assumindo que todas as *features* são condicionalmente independentes. Isto significa que a estimativa de $p(x|\omega_i)$ pode ser obtida a partir do produto da distribuição individual de cada *feature*,

$$p(x_1, \dots, x_p|\omega_k) = \prod_{i=1}^p p(x_i|\omega_k)$$

No entanto, tem-se de ter atenção que o classificador de *Naive Bayes* já não é considerado ótimo, pois a suposição de independência pode não ser verdade.

2 Um exemplo simples

2.7 Resultados e comentários sobre os classificadores de *Bayes* e *Naive Bayes*

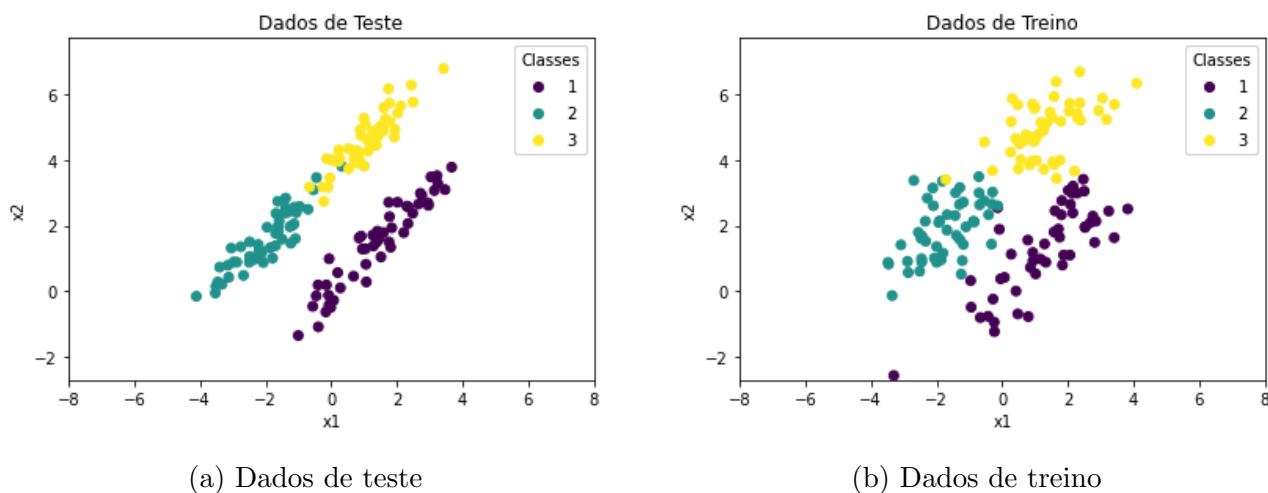


Figura 1: Dados de teste e treino fornecidos

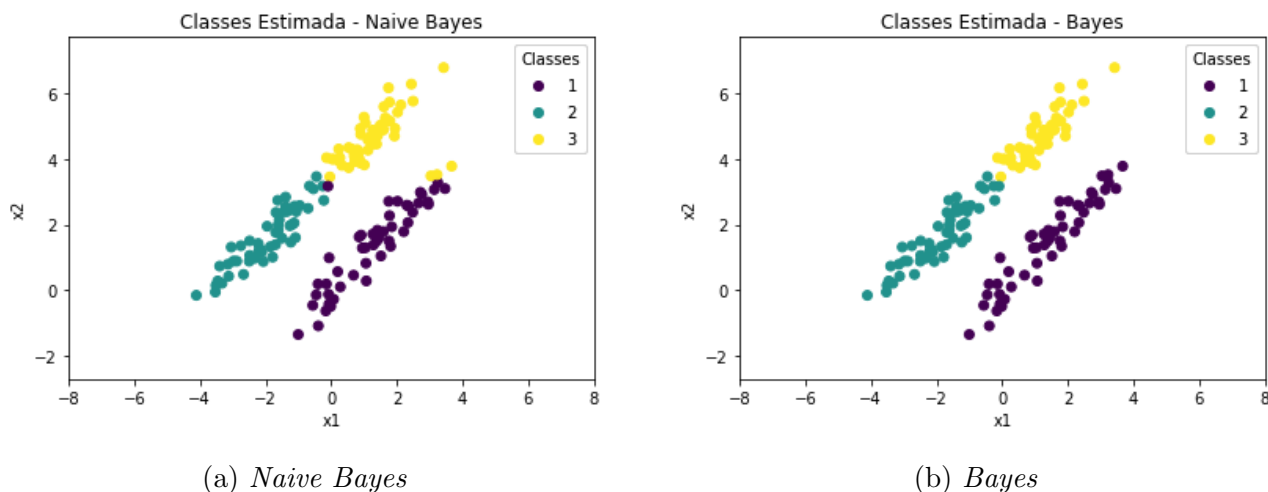


Figura 2: Resultados obtidos para os classificadores

Tabela 1

Classificador	<i>Naive Bayes</i>	<i>Bayes</i>
Erro (%)	5,(3)	3,(3)

A partir da tabela 1, verifica-se que os valores de erro percentual para ambas as estimativas são baixos, sendo o *Naive Bayes* ligeiramente maior.

A partir da figura 1(a) pode-se observar que para ambos os dados de treino e teste, tem-se pontos em zonas onde existem dados de maioritariamente de uma classe, classificados com outra classe. Logo é de esperar que estes pontos não sejam classificados incorretamente.

Quando analisamos a figura 2(a) observa-se que, apesar de poucos, alguns dos erros cometidos pelo classificador de *Naive Bayes* são muito discrepantes da realidade. A principal causa destes erros é devido à dependência de x_1 e x_2 não considerada por este modelo. A partir das figuras 1(a), constata-se que x_1 pode ser sensivelmente relacionado com x_2 , pela equação $x_2 = ax_1 + b$ para cada classe, que verifica a sua dependência. O que leva a classificar, por exemplo, pontos como classe 1 numa zona onde existem em abundância pontos da classe 2 e 3.

Na figura 2(b), verifica-se que no classificador de *Bayes* os limites das regiões de decisão dividem as três classes em 3 zonas próximas às que seriam criadas intuitivamente.

Assim, tendo em conta a simplicidade do classificador de *Naive Bayes* em relação ao de *Bayes*, os resultados obtidos são satisfatórios, mas vão ser tanto piores quanto maior for a dependência entre as *features*.

3 Reconhecedor de idiomas

3.7 Resultados obtidos para cada frase e comentários

Tabela 2

Texto	Idioma real	Idioma reconhecido	Probabilidade	Margem de classificação
Que fácil es comer peras.	es	es	$6,703 \times 10^{-1}$	$3,407 \times 10^{-1}$
Que fácil é comer peras.	pt	pt	1	1
Today is a great day for sightseeing.	en	en	1	1
Je vais au cinéma demain soir.	fr	fr	1	1
Ana es inteligente y simpática.	es	es	1	1
Tu vais à escola hoje.	pt	fr	$7,931 \times 10^{-1}$	$5,861 \times 10^{-1}$
Tu vais para a escola hoje.	pt	pt	$9,958 \times 10^{-1}$	$9,915 \times 10^{-1}$

Analisando a primeira e a segunda frase é fácil perceber que são iguais a exceção de "es" que passando do espanhol para o português fica "é". Sendo que os trigramas em que "es" tem influencia estão bastante presentes nos dois idiomas, apesar de ser maior em espanhol, faz com que a probabilidade e margem de classificação da primeira frase sejam mais reduzidas. Por sua vez os trigramas em que "é" tem influencia estão muito mais presentes no idioma português, o que resulta numa probabilidade e margem de classificação de aproximadamente 1, para a segunda frase.

Relativamente à terceira frase, sendo o idioma inglês, como este tem uma origem diferente dos outros três idiomas vai ter muitos trigramas próprios e quase únicos. Logo a sua classificação, probabilidade e margem de probabilidade são o que se esperava.

Na quarta frase está presente o trigrama "ain" que permite distinguir entre o espanhol e o português, sendo que a maioria dos restantes são bastante comuns nas 3, além disso estes últimos não estão tão presentes em inglês logo é possível ter uma classificação correta e com uma margem tão boa.

Na quinta frase está presente um trigrama característico do espanhol " y ", o que leva a que as semelhanças com o português, na grande maioria dos trigramas, perca a sua importância, permitindo uma classificação correta com bastante confiança.

Analisando a sexta frase e os respetivos resultados, observamos que a frase é mal classificada. Isto é explicado pela falta de trigramas próprios do português e a existência de vários trigramas comuns a ambos. Para tentar resolver este problema alterou-se a frase seis, substituindo "à" por "para a". As alterações nos trigramas e as respetivas probabilidades para cada idioma estão refletidos nas tabelas seguintes.

Tabela 3: Trigramas Originais

	Probabilidade fr	Probabilidade pt
"s à"	$3,276 \times 10^{-4}$	$4,342 \times 10^{-5}$
" à "	$2,595 \times 10^{-3}$	$2,940 \times 10^{-4}$
"à e"	$3,583 \times 10^{-5}$	$1,813 \times 10^{-5}$
Total	$3,047 \times 10^{-11}$	$2,315 \times 10^{-13}$

Tabela 4: Trigramas Alteração

	Probabilidade fr	Probabilidade pt
"s p"	$1,550 \times 10^{-3}$	$1,361 \times 10^{-3}$
" pa"	$3,207 \times 10^{-3}$	$2,783 \times 10^{-3}$
"par"	$2,454 \times 10^{-3}$	$2,171 \times 10^{-3}$
"ara"	$4,368 \times 10^{-4}$	$1,857 \times 10^{-3}$
"ra "	$3,342 \times 10^{-4}$	$2,928 \times 10^{-3}$
"a h"	$6,415 \times 10^{-5}$	$2,141 \times 10^{-4}$
Total	$1,142 \times 10^{-19}$	$9,574 \times 10^{-18}$

Através destas tabelas conclui-se que os trigramas originais tinham um peso maior em francês enquanto que os introduzidos pela alteração têm maior peso em português, levando assim a uma classificação correta com uma boa probabilidade e margem.