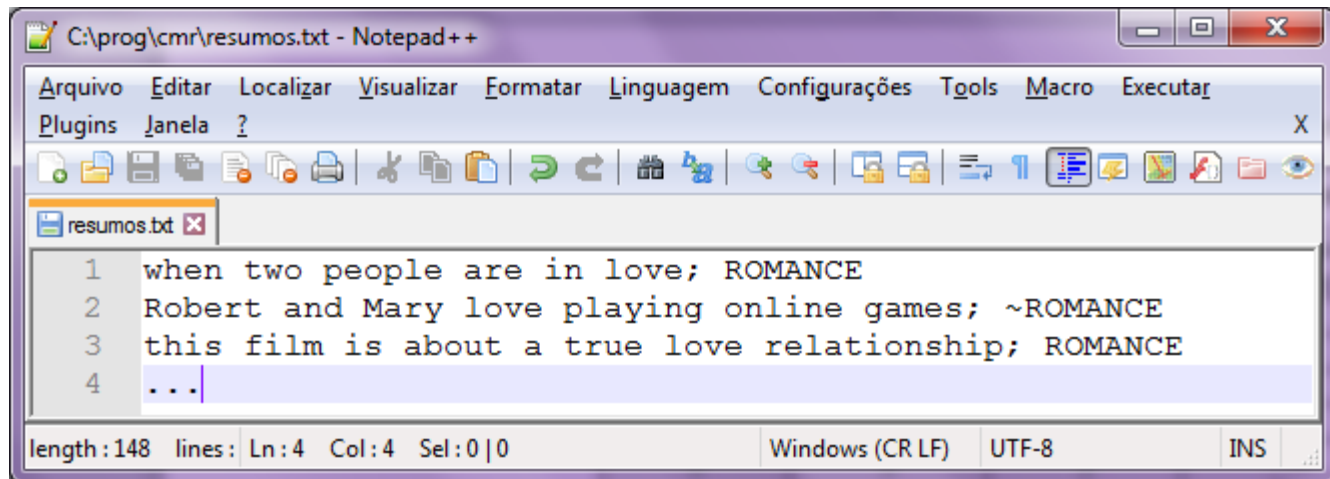


Algoritmos de Classificação

- NB (1/5) – “o classificador probabilístico”
- **EXEMPLO:** criar um classificador NB para classificar um filme como Romance (**‘SIM’** ou **‘NÃO’**) em função da ocorrência das palavras “love”, “people” e “relationship” em seu resumo.



```
C:\prog\cmr\resumos.txt - Notepad++
Arquivo  Editar  Localizar  Visualizar  Formatar  Linguagem  Configurações  Tools  Macro  Executar
Plugins  Janela  ?
resumos.txt x
1  when two people are in love; ROMANCE
2  Robert and Mary love playing online games; ~ROMANCE
3  this film is about a true love relationship; ROMANCE
4  ...
length:148 lines: Ln:4 Col:4 Sel:0|0 Windows (CR LF) UTF-8 INS
```

Algoritmos de Classificação

- NB (2/5) – “o classificador probabilístico”
- BD de treinamento (montado a partir da coleção de resumos)

love	people	relationship	Romance (classe)
1	1	0	Sim
1	0	0	Não
1	0	1	Sim
0	0	1	Sim
1	1	1	Não
1	1	0	Sim
0	0	0	Não
0	1	0	Não
1	1	1	Sim
0	1	1	Não
1	0	1	Sim
1	1	0	Não
0	0	0	Não
1	1	1	Sim
1	0	1	Sim

Algoritmos de Classificação

- **NB (3/5)** – “o classificador probabilístico”
- **Treinamento:** computar tabela de probabilidades condicionais que resume o BD de treinamento.

Romance (classe)	love		people		relationship	
	0	1	0	1	0	1
Não 7/15 (46,67%)	4/7 (57,14%)	3/7 (42,86%)	3/7 (42,86%)	4/7 (57,14%)	5/7 (71,43%)	2/7 (28,57%)
Sim 8/15 (53,33%)	1/8 (12,50%)	7/8 (87,50%)	4/8 (50,00%)	4/8 (50,00%)	2/8 (25,00%)	6/8 (75,00%)

- **Classificação:** Aplica-se o Teorema de Bayes.

$$P(Y | X) = \frac{P(X | Y) \times P(Y)}{P(X)}$$

Algoritmos de Classificação

- NB (4/5) – “o classificador probabilístico”

• **Exemplo:** classificar um novo filme $t = (\text{love}=1, \text{people}=0, \text{relationship}=1)$.

Romance (classe)	love		people		relationship	
	0	1	0	1	0	1
Não 7/15 (46,67%)	4/7 (57,14%)	3/7 (42,86%)	3/7 (42,86%)	4/7 (57,14%)	5/7 (71,43%)	2/7 (28,57%)
Sim 8/15 (53,33%)	1/8 (12,50%)	7/8 (87,50%)	4/8 (50,00%)	4/8 (50,00%)	2/8 (25,00%)	6/8 (75,00%)

Estimativa.1: $P(\text{Romance} = \text{'Não'} \mid t) = 0,4286 \times 0,4286 \times 0,2857 \times 0,4667 = \mathbf{0,0245}$

Estimativa 2: $P(\text{Romance} = \text{'Sim'} \mid t) = 0,8750 \times 0,5000 \times 0,7500 \times 0,5333 = \mathbf{0,1750}$

Algoritmos de Classificação

- NB (5/5) – “o classificador probabilístico”

• **Exemplo:** classificar um novo filme $t = (\text{love}=1, \text{people}=0, \text{relationship}=1)$.

Romance (classe)	love		people		relationship	
	0	1	0	1	0	1
Não 7/15 (46,67%)	4/7 (57,14%)	3/7 (42,86%)	3/7 (42,86%)	4/7 (57,14%)	5/7 (71,43%)	2/7 (28,57%)
Sim 8/15 (53,33%)	1/8 (12,50%)	7/8 (87,50%)	4/8 (50,00%)	4/8 (50,00%)	2/8 (25,00%)	6/8 (75,00%)

Normalizando Est. 1: $P(\text{Romance} = \text{'Não'} \mid t) = 0,0245 / (0,0245 + 0,1750) = \mathbf{12,28\%}$

Normalizando Est. 2: $P(\text{Romance} = \text{'Sim'} \mid t) = 0,1750 / (0,0245 + 0,1750) = \mathbf{87,72\%}$