Exemplo

Name	Debt	Salary	Married	Risk
JOSÉ	High	High	yes	High
ANA	Low	High	yes	Low
JOÃO	High	High	No	Low
MARIA	High	Low	yes	High
RUI	Low	Low	yes	High

→ seleciona-se os vizinho e desses escolhe-se os K mais próximos

a) K-Nearest-Neighbors Classifier, with K=3

1º converter para as variáveis categóricas para numéricas
Name não conta para variável preditiva, pois é uma chave

Debt	Salary	Married
1	1	1
∅	1	1
1	1	∅
1	∅	1
∅	∅	1

→ distância euclidiana

Distância	
$\sqrt{2}$	← $*_1$ (*√2)
$\sqrt{1} = 1$	← $*_1$ / $*_2$
$\sqrt{1} = 1$	← $*_1$ / $*_2$
$\sqrt{3}$	
$\sqrt{2}$	← $*_2$ (*√2)

Menor dista. logo temos $*_1$ ou $*_2$ → Moda = low, pois é o que aparece mais

Risk
High
Low
Low
High
High

→ atributo objetivo

$\frac{2}{5}$ - Low (*b)

como atributo categórico vai-se pela MODA

Debt = ∅ ; Salary = 1 ; Married = ∅ ⇒ Risk = Low

$$\sqrt{(x_1-y_1)^2+(x_2-y_2)^2+\dots} = \sqrt{(1-0)^2+(1-1)^2+(1-0)^2} = \sqrt{1^2+0+1^2} = \sqrt{2} \quad (*\sqrt{2})$$

Naive Bayes... ↳ não é necessário usar variáveis numéricas
↳ Admite que as variáveis são independentes e calcula as frequências

TABELA DE FREQUÊCIAS

Debt \ Risk	Low	High
Low	1+1=2	1+1=2
High	1+1=2	2+1=3

soma 1 a tudo. Temos de usar a correção da estimativa laplace=1 pois estamos perante valores nulos

Salary	Low	High
Low	∅+1=1	2+1=3
High	2+1=3	1+1=2

Married	Low	High
yes	1+1=2	3+1=4
No	1+1=2	∅+1=1

TABELA DE PROBABILIDADES

	Low	High
Low	1/2	2/5
High	1/2	3/5

	Low	High
Low	1/4	3/5
High	3/4	2/5

	Low	High
yes	1/2	4/5
No	1/2	1/5

b)

$P(Risk = Low \mid Debt = Low, Salary = High, Married = No) = \frac{2}{5} \times \frac{1}{2} \times \frac{3}{4} \times \frac{1}{2} =$

$= \frac{6}{80}$ //

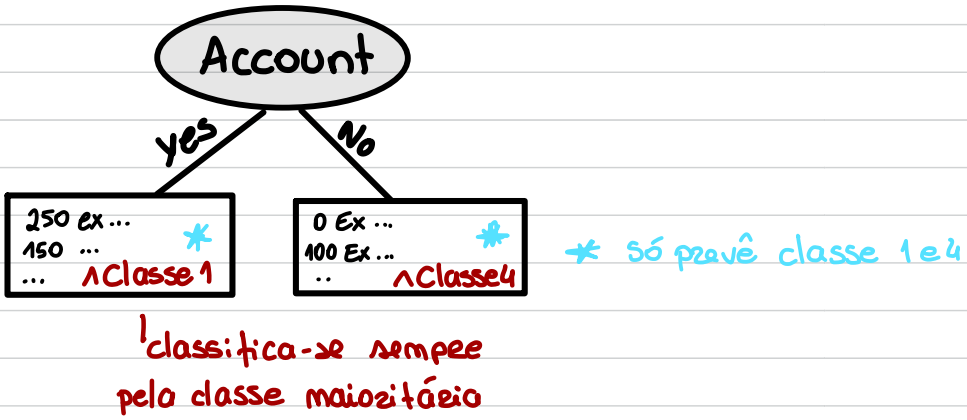
$P(Risk = High \mid Debt = Low\ Salary = High, Married = No) = \frac{3}{5} \times \frac{2}{5} \times \frac{2}{5} \times \frac{1}{5} =$

$= \frac{12}{375}$ //

$P(Risk = Low \mid \ldots) > P(Risk = High \mid \ldots) \Rightarrow Risk = Low$

---

4 classes (1, 2, 3, 4)
250 instâncias para cada classe



a)

Reais \ Prevista	^classe 1	^classe 2	^classe 3	^classe 4
classe 1	250	∅	∅	∅
classe 2	150 (prevê classe 1 e é 2)	∅	∅ c)	100 d)
classe 3	150 d)	∅	∅	100
classe 4	50	∅	∅	200

b) $\frac{\text{total de n bem classificados}}{\text{total}}$ = taxa de erro = 1 - taxa de aceito
= 1 - 0.45 = 0.55 // muito alta

taxa de aceito = $\frac{\text{linha}\searrow}{\text{total}} = \frac{450}{1000} = 0.45$

c) Modelo não consegue prever a classe, pois valores nulos *

d)

	$\wedge$ classe 1	$\wedge \overrightarrow{\neg}$ classe 1
classe 1	250 TP	$\emptyset$
$\neg$ classe 1	350 FP *	400 *

$$\text{Precision} = \frac{250}{600} = 41.7\%.$$

Recall = 100%.

aceita todas as instâncias do classe 1 pelo que se escolhesse este

	$\wedge$ classe 4	$\wedge \overrightarrow{\neg}$ classe 4
classe 4	200 TP	50
$\neg$ classe 4	200 FP	550

↓ precisão de cerca de 50%.

$$\text{Precision} = \frac{200}{400} = 50\%$$

$$\text{Recall} = \frac{200}{250} = 80\%$$

↓

acaba por errar menos do que aceita

200 → 200