



- 1) Obter o conjunto de regras de classificação pelo método 1R (*One Rule*). (2,5)

Atributo	Regras	Erros	Total erros
Outlook	Sunny → No	2/5	4/14
	Overcast → Yes	0/4	
	Rainy → Yes	2/5	
Temp	Hot → No*	2/4	5/14
	Mild → Yes	2/6	
	Cool → Yes	1/4	
Humidity	High → No	3/7	4/14
	Normal → Yes	1/7	

- As regras para Outlook ou Humidity são obtidas pelo 1R.

- Note que NUSP é perfeitamente correlacionado com Temperatura para $d=1,2,\dots,9$ ou irrelevante para $d=0$, não influenciando na resposta obtida em nenhum dos casos.

- 2) Classificar a tupla [ensolarado, quente, normal, d] pelo *Naive Bayes*. (2,5)

Aparência	Sim	Não		Temperatura	Sim	Não		Umidade	Sim	Não
Ensolarado	2/9	3/5		Quente	2/9	2/5		Alta	3/9	4/5
Chuvoso	3/9	2/5		Fria	3/9	1/5		Normal	6/9	1/5
Nublado	4/9	0		Morna	4/9	2/5				

Para $d=0$:

Sim: $2/9 \times 2/9 \times 6/9 \times 9/9 \times 9/14 = 216/10.206 = 0,02 = 55\%$

Não: $3/5 \times 2/5 \times 1/5 \times 5/5 \times 5/14 = 30/1.750 = 0,01 = 45\%$

Para $d \neq 0$ (atributo NUSP fornece a mesma informação que temperatura, sendo $d=\text{quente}$):

Sim: $216/10206 \times 2/9 = 41\%$

Não: $30/1750 \times 2/5 = 59\%$

- 3) Considere a base de dados abaixo como base de treinamento e classifique a tupla [1, 1, d] de acordo com o método dos vizinhos mais próximos (k -NN, com $k=3$) utilizando a distância Euclidiana elevada ao quadrado. Apresente os cálculos realizados. (2,5)

Exemplo	A ₁	A ₂	NUSP	Classe	Distância	Vizinho
1	0	1	d	A	1	2,5
2	1	1	d	B	0	1
3	1	0	d	A	1	2,5
4	5	5	d	B	32	$f(d), k \geq 4$
5	0	0	3d	A	$2+4d^2$	$f(d), k \geq 4$
6	2	2	2d	B	$2+d^2$	$f(d), k \geq 4$
7	4	4	d	A	18	$f(d), k \geq 4$

Note que, independentemente do valor de “d”, os 3 exemplos mais próximos são sempre {1,2,3}.

Resposta: As distâncias são $d(1)=1$ (A), $d(2)=0$ (B), $d(3)=1$ (A) e a classe majoritária é portanto A.

- 4) Assinale Verdadeiro (V) ou Falso (F) para cada uma das afirmações abaixo (2,5):

(F) I. A validação cruzada é inapropriada para estimar a capacidade de generalização de classificadores;

(V) II. Quando possível, e na ausência de limitações computacionais, deve-se preferir a validação “*Leave One Out*”;

(F) III. Na prática, a melhor validação cruzada em k pastas é obtida com $k=1$;

(F) IV. A validação cruzada nunca deve ser usada para otimizar parâmetros de algoritmos de AM;

(F) V. Algoritmos de aprendizado mais complexos são sempre melhores porque generalizam melhor.