



- Observações:**
- A prova deverá ser realizada individualmente;
 - A interpretação das questões é parte integrante da avaliação;
 - Para todas as questões, considere que d é o último dígito do seu número USP.

Declaro que realizei esta prova individualmente, sem consultar outras pessoas, e que estou ciente do código de ética da USP.

Para as questões 1 e 2, considere a seguinte base de dados, na qual NUSP é um atributo categórico:

Aparência	Temperatura	Umidade	NUSP	Classe
Ensolarado	Quente	Alta	d	Não
Ensolarado	Quente	Alta	d	Não
Chuvoso	Frio	Normal	2d	Não
Ensolarado	Morna	Alta	3d	Não
Chuvoso	Morna	Alta	3d	Não
Nublado	Quente	Alta	d	Sim
Chuvoso	Morna	Alta	3d	Sim
Chuvoso	Frio	Normal	2d	Sim
Nublado	Frio	Normal	2d	Sim
Ensolarado	Frio	Normal	2d	Sim
Chuvoso	Morna	Normal	3d	Sim
Ensolarado	Morna	Normal	3d	Sim
Nublado	Morna	Alta	3d	Sim
Nublado	Quente	Normal	d	Sim

- 1) Obter o conjunto de regras de classificação pelo método 1R (*One Rule*). (2,5)
- 2) Classificar a tupla [ensolarado, quente, normal, d] pelo *Naive Bayes*, estimando as probabilidades de classes. (2,5)
- 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
1	0	1	d	A
2	1	1	d	B
3	1	0	d	A
4	5	5	d	B
5	0	0	3d	A
6	2	2	2d	B
7	4	4	d	A

- 4) Assinale Verdadeiro (V) ou Falso (F) para cada uma das afirmações abaixo (2,5):
 - () I. A validação cruzada é inapropriada para estimar a capacidade de generalização de classificadores;
 - () II. Quando possível, e na ausência de limitações computacionais, deve-se preferir a validação “*Leave One Out*”;
 - () III. Na prática, a melhor validação cruzada em k pastas é obtida com $k=1$;
 - () IV. A validação cruzada nunca deve ser usada para otimizar parâmetros de algoritmos de AM;
 - () V. Algoritmos de aprendizado mais complexos são sempre melhores porque generalizam melhor.