

Exercícios de Fixação

1. O que é uma base de dados desbalanceada?
2. Compare as métricas accuracy e f1 em bases de dados binárias indicando as suas vantagens em relação a outra.
3. Explique porque há necessidade de usar os conceitos micro, macro e weighted em conjunto com as métricas precision, recall e f1? Porque esses conceitos não se aplicam a accuracy?
4. Qual a relação entre micro precision, micro recall, micro f1 e accuracy quando o conjunto de teste é perfeitamente balanceado? Essa relação se aplica a macro f1?
5. Porque o método de amostragem por ressubstituição não é indicado para avaliação de métodos de classificação? Apresente uma utilidade para este método de amostragem.
6. Apresente as vantagens do método de validação cruzada estratificada em comparação ao método de divisão percentual.
7. Quando o método de exclusão de 1 deve ser preferido em relação ao método de validação cruzada? Qual a maior dificuldade para aplicação desse método?
8. Os métodos mais robustos de classificação supervisionada contém vários parâmetros. De que forma pode ser determinado os valores desses hiperparâmetros? Qual é o método mais usado e porque ele é o preferido?
9. Qual a forma mais indicada de se evitar superajuste dos métodos. Explique como ela funciona. Considere o uso do classificador K Vizinhos Mais Próximos com o método indicado e explique porque nesse caso não é possível indicar o valor de K encontrado.
10. Porque é importante usar testes estatísticos na comparação de desempenho entre métodos de classificação? Qual a diferença entre usar um método paramétrico e não paramétrico de testes de hipóteses? Explique porque é necessário ter 30 valores para aplicar o teste T. Indique em qual dos testes é mais fácil mostrar a diferença entre os métodos. Explique porque isso ocorre.
11. Explique o que é o problema do bias de similaridade apresentado no artigo "An Experimental Methodology to Evaluate Machine Learning Methods for Fault Diagnosis based on Vibration Signals" e apresente o

contexto em que ele pode ocorrer.

12. Considere a seguinte matriz de confusão:

Classe	A (Preditas)	B (Preditas)	C (Preditas)
A (Verdadeiras)	6	2	0
B (Verdadeiras)	1	4	1
C (Verdadeiras)	1	0	5

Calcule as seguintes métricas:

- a) acurácia (accuracy)
- b) precisão (precision) da classe B
- c) revocação macro (macro recall)

13. Explique como funciona o método de reamostragem de validação cruzada aninhada. Apresente a vantagem desse método em relação ao método de validação cruzada simples.