

Exercícios de Fixação

1. Carregue o dataset breast cancer disponível no sklearn e obtenha a distribuição de classes deste dataset.
2. Compare as métricas accuracy e f1 em bases de dados binárias indicando as suas vantagens em relação a outra. Use o classificador Vizinho Mais Próximo e uma divisão percentual com 30% dos exemplos do dataset breast cancer usados para teste para obter o valor da acurácia e f1 deste classificador neste dataset.
3. Use o classificador Vizinho Mais Próximo e uma validação cruzada estratificada com 10 folds no dataset breast cancer para obter a matriz de confusão, a acurácia, o desvio padrão, o intervalo de confiança a 95%, a f1 macro, a precision micro e o recall micro deste classificador neste dataset.
4. Use o classificador K Vizinhos Mais Próximos com o método indicado (use 10 partições) no dataset breast cancer para realizar a busca pelo melhor valor de K. Apresente a melhor acurácia obtida, o desvio padrão, o intervalo de confiança a 95% e o valor de K correspondente.
5. Use o classificador K Vizinhos Mais Próximos com o método indicado no dataset breast cancer para realizar a busca pelo melhor valor de K sem possibilidade de superajuste. Apresente a melhor acurácia obtida, o desvio padrão e o intervalo de confiança a 95%.
6. Repita o procedimento proposto no exercício 9 somente trocando o classificador para uma Árvore de Decisão. Compare o desempenho obtido pelo K Vizinhos Mais Próximos com a Árvore de Decisão utilizando o teste T e o de Wilcoxon. Compare os dois resultados e indique em qual dos testes é mais fácil mostrar a diferença entre os métodos.