

Roteiro Aula 5 Lab

1. Apresente as matrizes de confusão da aplicação dos classificadores ZeroR, Aleatório Uniforme e Aleatório Estratificado usando validação cruzada com 10 dobras (folds) na base de dados (dataset) wine.
2. Obtenha a acurácia média, o desvio padrão e o intervalo de confiança a 95% dos classificadores Vizinho Mais Próximo, 3 Vizinhos Mais Próximos e 5 Vizinhos Mais Próximos usando validação cruzada com 5 dobras (folds) na base de dados (dataset) wine.
3. Reexecute os classificadores K Vizinhos Mais Próximos do exercício 2 na mesma base wine com a mesma estratégia de reamostragem e usando o inverso da distância para diferenciar o peso dos vizinhos na hora de computar a classificação dos exemplos de teste. Compare os resultados com os obtidos no exercício 2.
4. Obtenha a acurácia média, o desvio padrão e o intervalo de confiança a 95% do classificador Vizinho Mais Próximo usando validação cruzada com 10 dobras (folds) na base de dados (dataset) wine padronizada e não padronizada.
5. Use os classificadores de Bagging (<https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.BaggingClassifier.html>) e AdaBoost (<https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.AdaBoostClassifier.html>) para realizar 3 repetições de validação cruzada aninhada com 10 folds no loop externo e 4 folds no loop interno. A busca em grade deve variar apenas a quantidade de classificadores do combinado considerando combinados com 10, 25 e 50 classificadores. A base de dados utilizada é a wine e deve ser padronizada. Apresente a acurácia média, o desvio padrão e o intervalo de confiança a 95% dos dois métodos e realize o teste t de Student para verificar se as diferenças são significativas.
6. Crie um classificador Tabela que classifica exemplos iguais com a mesma classe e exemplos distintos com a classe majoritária. Avalie este classificador usando na base iris a estratégia de resubstituição (na qual treino e teste são os mesmos conjuntos) e usando uma divisão percentual de 70% dos exemplos usados para treino e 30% dos exemplos usados para teste. Compare os resultados e justifique apontando qual resultado tem maior chance de ser generalizado.