

Lista de Exercícios 07

(prazo final para entrega: 29/04/2019 - segunda)

1. Resolva o problema de classificação binária sobre o dataset "[Wisconsin Diagnostic Breast Cancer \(WDBC\)](#)".
2. Use 80% do dataset para treino e 20% apenas para teste. Use estratificação baseada no label.
 - a. Dica: usar [train_test_split](#) com stratify, shuffle e random_state.
3. Aplicar Standardization nas features.
4. Usar o seguinte algoritmo: Perceptron.
 - a. Use modelos usando diferentes regularizações: Nenhuma, L1, L2, ElasticNet (penalty : None, 'l2' or 'l1' or 'elasticnet').
 - b. Use também diferentes valores de alpha para ver como eles afetam os resultados na validação cruzada.
 - c. Usar a seguinte métrica: Accuracy.
5. Use K-Fold Cross Validation ([cross_val_score](#)) com k = 3 apenas sobre os dados de treino.
 - a. Use validação cruzada estratificada ([StratifiedKFold](#)) com randomização dos dados (usar shuffle e random_state).
 - i. Dica: instancie um objeto da classe StratifiedKFold e use-o no parâmetro cv do [cross_val_score](#).
6. Depois de obter o melhor modelo via cross-validation sobre o conjunto de treino, use os melhores hiper-parâmetros encontrados para treinar um modelo usando todo o dado de treino. Depois, avalie esse modelo usando os dados de teste.