

TP2: IMPLEMENTAÇÃO DO ALGORITMO DE BOOSTING

Descrição

Sabemos que erros de um modelo de previsão podem ser decompostos em dois fatores: viés e variância. Um modelo complexo apresenta pouco viés, mas grande variância, enquanto um modelo simples apresenta grande viés e pouca variância.

Boosting é o processo de reduzir o viés de um grande conjunto de modelos simples (i.e., com baixa variância). Esses modelos são chamados de modelos fracos, e são levemente correlacionados com a classificação correta. No processo de Boosting, os modelos fracos formam um modelo mais forte de maneira iterativa. Os modelos fracos são escolhidos iterativamente, de forma que cada modelo é escolhido levando-se em conta vieses independentes. Ou seja, cada modelo componente realiza erros diferentes de outros modelos componentes.

Objetivo

Nosso foco neste trabalho prático é ganhar experiência com o processo de Boosting. Você deverá implementar o processo de Boosting (visto em sala de aula) assumindo um problema de classificação binária com atributos categóricos. Em particular, você deverá realizar os experimentos utilizando o dataset tic-tac-toe, disponível em <https://archive.ics.uci.edu/ml/datasets/Tic-Tac-Toe+Endgame>. Sua avaliação deverá seguir a metodologia de validação cruzada com 5 partições. A medida de eficácia a ser considerada é a taxa de erro simples. Você deverá implementar todo o processo de Boosting, e portanto não é permitido utilizar soluções prontas (embora não seja necessário implementar funções básicas). A escolha da linguagem de programação é uma decisão do aluno.

Entregáveis

Você deverá entregar, via Moodle da disciplina e dentro do prazo, arquivo contendo o código fonte do programa desenvolvido, bem como a documentação. A documentação deve conter toda a informação referente a suas escolhas de implementação. Além disso, você deverá fazer uma série de análises de erro, e reportar os resultados obtidos. Os gráficos a serem apresentados devem refletir experimentos que avaliem aspectos relacionados ao processo de Boosting, como por exemplo o número de iterações.