

Trabalho Prático II

Igor Lacerda Faria da Silva

igorlfs@ufmg.br

Introdução

O objetivo deste trabalho é aprofundar o conhecimento do algoritmo de *boosting*, implementando-o “do zero”. Mais especificamente, foi implementado o *AdaBoost* com *stumps* (árvores de decisão com apenas um nó). Como exemplo, foi analisado o *dataset Tic-Tac-Toe Endgame*, que contém todas as instâncias de “Jogo da Velha” em que o jogador inicial é x , além do resultado do jogo. No entanto, o programa é robusto: é possível usar outros bancos de dados apenas trocando alguns parâmetros.

Foi usada validação cruzada (*5-fold*) para avaliar o desempenho do modelo. Este relatório apresenta a evolução do erro variando-se a quantidade de *stumps*.

Desenvolvimento



Figura 1: Erro pelo número de *stumps*, na validação cruzada.

O desempenho do modelo, de forma geral, não foi satisfatório. É incomum que alguma divisão da validação cruzada atinja a “barreira” de 20% de erro, e a média raramente fica abaixo de 25%. É possível perceber uma melhora no

desempenho, conforme são adicionados mais *stumps*, mas a partir de cerca de 20 *stumps*, o desempenho piora drasticamente. Presumivelmente isso acontece porque são adicionados vários *stumps* da mesma classe, o que tende a degradar o modelo.