

Aprendizado de Máquina

Trabalho Prático 2

Luís Felipe Ramos Ferreira
2019022553
lframos_ferreira@outlook.com

June 18, 2023

1 Introdução

O Trabalho Prático 2 da disciplina de Aprendizado de Máquina teve como objetivo o desenvolvimento de um algoritmo de *boosting* para classificação binária. Em particular, o algoritmo a ser desenvolvido é o *Adaboost* e a base de dados a ser utilizada nos testes é o conjunto *Tic-Tac-Toe*. Além disso, os modelos criados deveriam ser analisados por meio da metodologia de validação cruzada com 5 partições para avaliação do modelo.

2 Implementação

A linguagem escolhida para o desenvolvimento do trabalho foi *Python* (versão 3.10), devida a sua grande variedade de bibliotecas úteis para ciência de dados e aprendizado de máquina. A modelagem do algoritmo *AdaBoost* foi feita com o uso de bibliotecas de análise numérica como *NumPy* e *Pandas*, uma vez que se tratam de ferramentas extremamente completas que facilitaram o desenvolvimento do algoritmo.

Para organizar o ambiente de desenvolvimento, que englobava vários pacotes diferentes, foi utilizado o gerenciador de pacotes *Anaconda*, o que facilitou o trabalho com os pacotes de ciência de dados citados. O projeto final foi salvo em um *repositório* no GitHub para fácil versionamento e organização de código.

2.1 Classificador

A implementação do classificador *AdaBoost* seguiu o padrão utilizado pela biblioteca *NumPy*, de modo que armazenar o classificador e todas as suas funcionalidades em um objeto permitia uma maior abstração do código e facilitou seu uso. A classe em questão possui um construtor e os métodos de treinamento e predição.

A implementação foi pensada especificamente para a classificação binária, de modo que bases de dados com mais de duas classes não irão funcionar. Como já citado, as funcionalidades das bibliotecas *NumPy* e *Pandas* foram extensamente utilizadas durante o desenvolvimento.

2.2 Validação cruzada

3 Análise dos resultados de teste

De maneira ger

4 Convergência do erro empírico

Durante o treinamento das redes neurais propostas, o histórico do erro empírico pode ser armazenado para análise de sua convergência, considerando cada configuração de rede proposta. Para fins de simplificação, serão mostrados aqui

5 Conclusão

Em suma, após as análises e discussões apresentadas neste relatório, fica claro que os parâmetros da rede neuronal, como o número de neurônios na camada oculta,