

Exercício 6

A.P. Braga

Janeiro de 2021

O objetivo dos exercícios desta semana é utilizar as ELMs para resolver problemas multidimensionais, a partir de bases de dados reais.

As bases de dados devem ser baixadas do repositório *UCI Machine Learning Repository* (<https://archive.ics.uci.edu/ml/datasets.php>).

A primeira base de dados a ser estudada é a base *Breast Cancer (diagnostic)*, disponível no link <https://archive.ics.uci.edu/ml/datasets/Breast+Cancer+Wisconsin+%28Diagnostic%29>

Para esta base, os alunos deverão dividir de forma aleatória os dados entre treinamento e teste e comparar as acurácias de treinamento e teste para diferentes valores do hiperparâmetro que controla o número de neurônios. Os valores de acurácia devem ser apresentados na forma de *media ± desvio_padrao* para, pelo menos, cinco execuções diferentes.

Algumas perguntas que devem ser respondidas são:

- Com quantos neurônios (aproximadamente) a acurácia de teste aparenta ser máxima?
- O que acontece com os valores de acurácia de treinamento e teste conforme aumentamos progressivamente o número de neurônios (por exemplo, para 5, 10, 30, 50, 100, 300 neurônios)?

O mesmo deve ser feito para a base *Statlog (Heart)*, disponível no link:

<https://archive.ics.uci.edu/ml/datasets/Statlog+%28Heart%29>

Além das Extreme Learning Machines, os alunos deverão, também, treinar, utilizando a rotina desenvolvida para as atividades anteriores, um perceptron, e avaliar seu desempenho na solução dos dois problemas, comparado às ELMs.

Por questões de convergência, pode ser necessário escalonar os valores dos atributos para que fiquem restritos entre 0 e 1. Para tanto, uma possibilidade é utilizar a forma abaixo:

$$z_i = \frac{x_i - \min(x)}{\max(x) - \min(x)}$$