

# Exercício 1

Matheus Bitarães de Novaes

30 de Maio de 2021

## 1 Introdução

Neste exercício, deve-se carregar a base de dados *Breast Cancer Wisconsin* da UCI utilizá-la para treinamento de um perceptron simples para separar as classes.

Esta base de dados possui 9 variáveis de entrada, uma variável de saída com a classificação das 699 amostras em maligno e benigno.

	id	f1	f2	f3	f4	f5	f6	f7	f8	f9	class
0	1002945	5	4	4	5	7	10	3	2	1	1
1	1015425	3	1	1	1	2	2	3	1	1	1
2	1016277	6	8	8	1	3	4	3	7	1	1
3	1017023	4	1	1	3	2	1	3	1	1	1
4	1017122	8	10	10	8	7	10	9	7	1	0

Figure 1: Amostra dos dados, contendo as *features* e a classificação. 1 para benigno e 0 para maligno

Antes de se realizar o treinamento do perceptron, foi necessário realizar a remoção de algumas linhas que não possuem todos os valores preenchidos.

Após o carregamento, limpeza e adequação dos dados, pode-se seguir para a implementação do perceptron.

## 2 O Perceptron

O Perceptron é um modelo linear de uma única camada, utilizada para classificação de dados que podem ser separados linearmente.

A saída do neurônio é calculada portanto como a soma das entradas, ponderada pelos pesos  $u = w_0 + x_1w_1 + x_2w_2 + \dots + x_nw_n$ . Este valor então é passado por uma função de ativação, que no nosso caso foi escolhida a função do degrau.

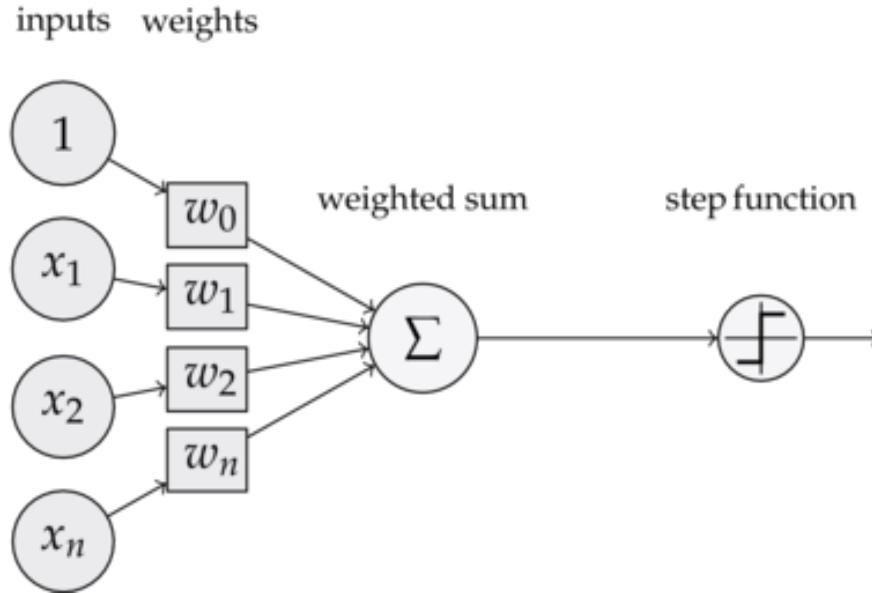


Figure 2: Arquitetura do perceptron

### 3 Treinamento do perceptron

Para o treinamento do neurônio, foi utilizado o seguinte pseudoalgoritmo:

Inicialização dos pesos;

**while** *Loop de épocas* **do**

**while** *for*  $x_i, y_i$  *in*  $(X, Y)$  **do**

        Realiza a previsão em  $x_i$ ;

        Atualiza pesos de acordo com o resultado da previsão;

**end**

**end**

**Algorithm 1:** Treinamento do Perceptron

Foi gerado um perceptron com taxa de aprendizado de 0.5 e 100 épocas. Após alguns testes, essa foi a configuração que apresentava melhores resultados de acurácia para as amostras de teste.

### 4 Resultados do teste e cálculo de acurácia para um conjunto

Após o treinamento, podemos realizar o calculo da acurácia do algoritmo, tanto para as amostras de treinamento quando para a amostras de teste. A acurácia pode ser estimada pela equação  $1 - (y_{teste} - y_t)^T * (y_{teste} - y_t) / n_{amostras}$

- Acurácia das amostras de treinamento: 97.3%
- Acurácia das amostras de teste: 98.9%

## 5 Análise de acurácia para 20 conjuntos diferentes de teste/treinamento

Foi realizado o treinamento de 20 conjuntos distintos de teste e treinamento, para verificar a acurácia média da abordagem. Nos histogramas abaixo é possível ver a distribuição das acurácias de teste e treinamento.

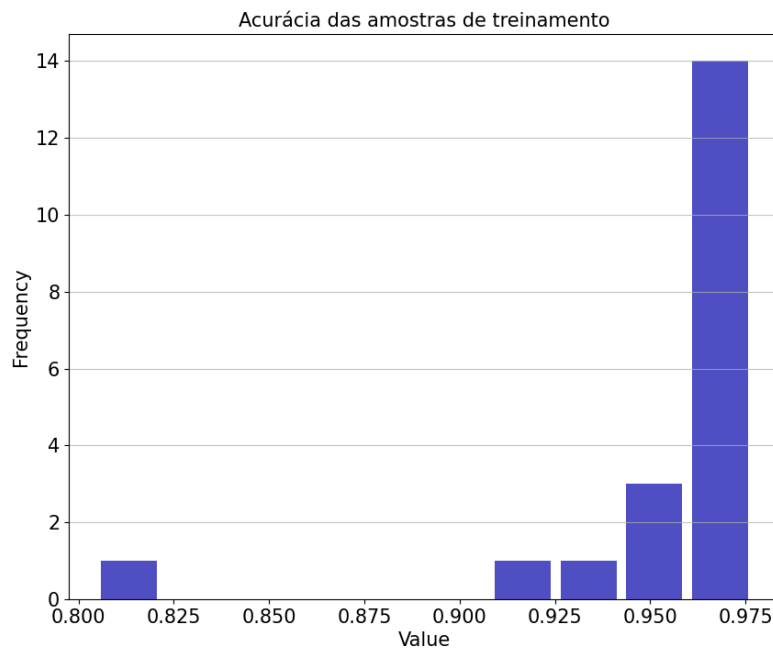


Figure 3: Histograma de acurácias para as amostras de treinamento

- Média de acurácia para amostras de treinamento: 95.5%
- Desvio padrão para amostras de treinamento: 0.0373
- Média de acurácia para amostras de treinamento: 97.93%
- Desvio padrão para amostras de treinamento: 0.0144

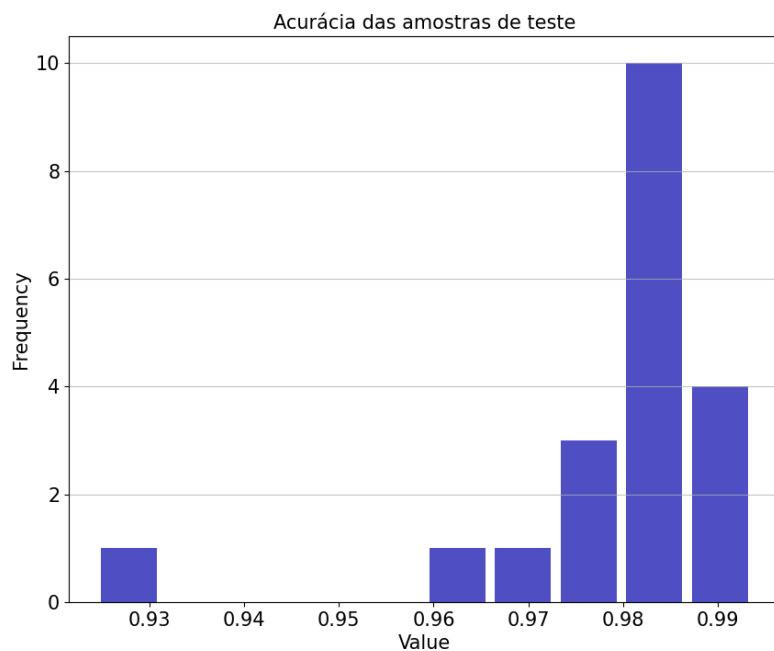


Figure 4: Histograma de acurácias para as amostras de teste

Pelas evidências acima, pode-se perceber que este modelo apresenta uma taxa de acerto média de 97% para as amostras de teste, possuindo uma generalização satisfatória e não realizando em *overfitting* nas amostras de teste.