

UNIVERSIDADE FEDERAL DO ESTADO DE SANTA CATARINA
PROGRAMA DE PÓS-GRADUAÇÃO EM SISTEMAS ELETRÔNICOS
APRENDIZADO DE MÁQUINA
WUERIKE HENRIQUE DA SILVA CAVALHEIRO

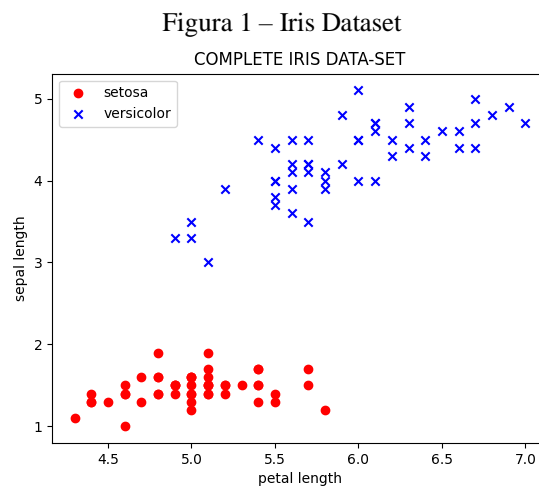
PERCEPTRON

Relatório da implementação de uma rede neural perceptron de um layer e um neurônio, sendo capaz de separar duas classes entre si através da análise de dois atributos.

Para validação do algoritmo implementado utilizou-se o conjunto de dados conhecido como *Iris Dataset* [1], este conjunto contém 150 amostras de quatro atributos sobre três espécies da flor íris, sendo 50 amostras para cada. As espécies são Setosa, Virginica e Versicolor enquanto os atributos envolvidos são comprimento e largura das sépalas e pétalas.

Para adequar os dados à rede neural implementada, considerou-se apenas amostras das espécies Setosa e Versicolor, bem como os atributos para a classificação foram limitados aos comprimentos da sépala e pétala. Obtendo assim um subconjunto de dois atributos e duas classes, passível de ser submetido ao perceptron de um layer e um neurônio.

A Figura 1 abaixo demonstra graficamente a distribuição do conjunto de dados utilizados nesta implementação.



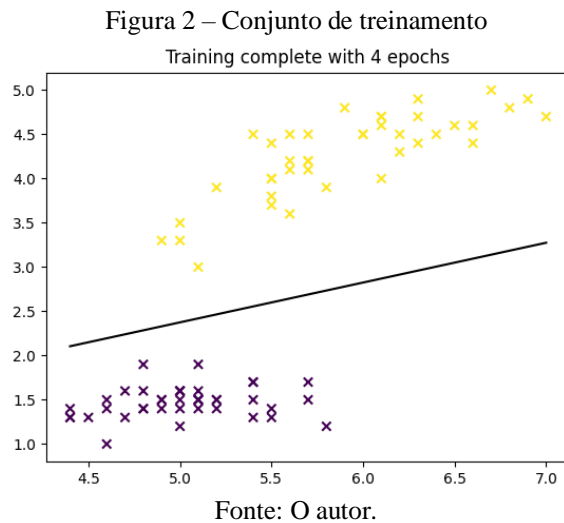
Fonte: O autor.

Como visível na Figura 1, a intersecção entre as classes é vazia, sendo um problema linearmente separável de baixa complexidade. Esta figura representa todos os pontos do conjunto, entretanto, na execução da rede neural utilizou-se 80% destes pontos para treinamento o restante para testar a generalização obtida. Além dos conjuntos de treino e teste, o algoritmo também espera pela definição da quantidade máxima de épocas de treinamento e da taxa de aprendizagem a ser utilizada.

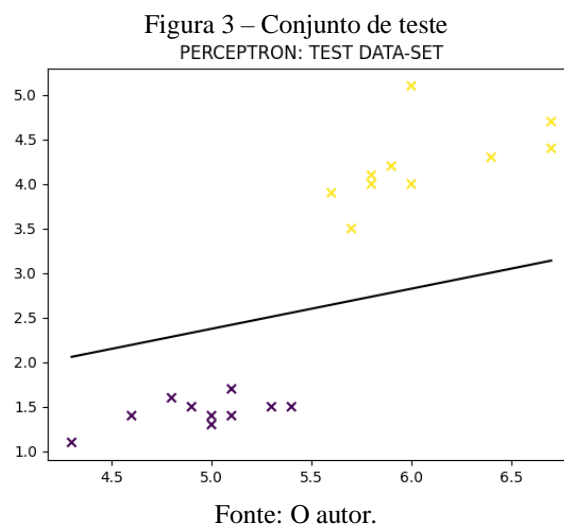
Nesta implementação, criou-se um método que verifica a convergência da fronteira de decisão, buscando encerrar o treinamento de forma antecipada em relação ao número máximo de épocas definido. Essa verificação ocorre ao comparar o valor do bias entre épocas, se não houver mudança do bias significa que o erro entre valor real e valor previsto não está variando, indicando assim a convergência ou um mínimo local.

Para evitar parar o treinamento num mínimo local, existe então o atributo *convergence factor* que diz quantas vezes o valor de bias deve repetir antes de parar o treinamento. Este valor deve ser ajustado empiricamente, sendo que nesta aplicação foi definido como sendo 3, isto pois a aplicação é simples e não tende a cair em mínimos locais.

A figura abaixo exibe o conjunto de treinamento bem como a reta que determina a fronteira de decisão encontrada. Como visto na figura, o treinamento para estes pontos durou 4 épocas para que a reta fosse capaz de separar as duas classes e parar de variar sua posição, sendo que a rede foi configurada para ter no máximo 100 épocas e taxa de aprendizado de 0,001.



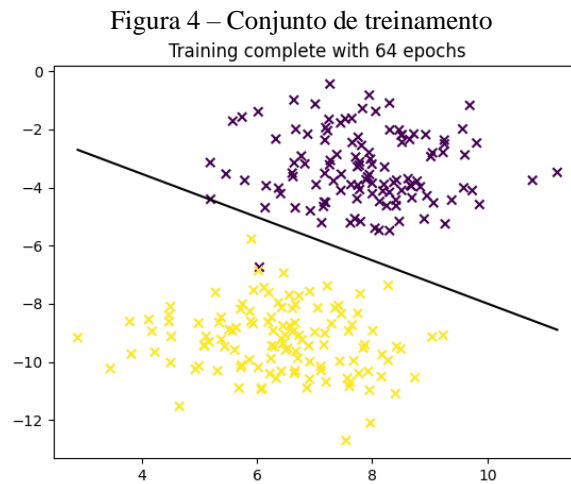
A Figura 3 representa os 20% dos pontos da Figura 1 que foram utilizados para o teste da rede treinada. Para estes dados não foram entregues ao perceptron as classes que cada ponto representa, sendo a divisão feita através da fronteira de decisão encontrada.



Neste caso, a acurácia do teste foi de 100%, de forma que todos os pontos do teste foram corretamente classificados entre as classes Setosa e Versicolor.

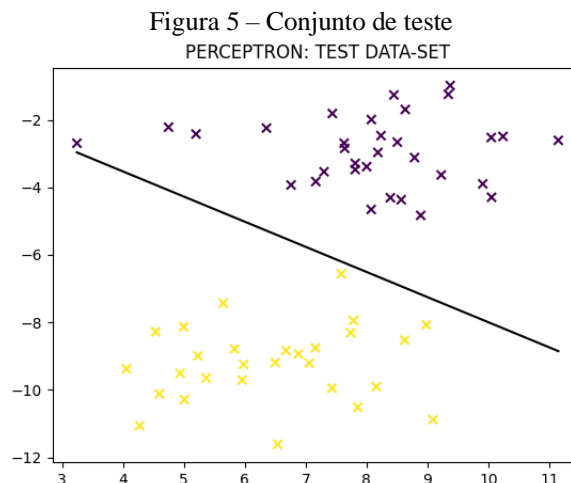
Para demonstrar de forma breve o comportamento da implementação em outro grupo de dados, utilizou-se o método *make_blobs*, da biblioteca de aprendizagem de máquina *scikit-learn* para gerar 300 pontos representados num plano bidimensional divididos em duas classes. Como no exemplo anterior, o conjunto foi dividido em 80% para teste e 20% para avaliação do desempenho do classificador.

A Figura 4 apresenta o *plot* realizado ao fim do treinamento, onde foram necessárias 64 épocas para se obter a convergência da fronteira de decisão. Ressalta-se que para esse conjunto de dados existe uma certa sobreposição das classes e que a separação linear realizada apesar de não ser perfeita buscou representar as classes da melhor forma possível.



Fonte: O autor.

Embora seja impossível a perfeita separação das classes, nota-se na Figura 5 que o conjunto utilizado no teste ficou completamente separado pela fronteira de decisão, ou seja, a acurácia obtida foi de 100% ao comparar valores previstos com os valores reais.



Fonte: O autor.

Assim como a execução do *Iris Dataset*, essa simulação também utilizou como parâmetros o número máximo de épocas em 100 e a taxa de aprendizagem em 0,001.

Assume-se que a diferença de 60 épocas entre os dois casos se deve a dois fatores, o primeiro é que o segundo caso não possui a intersecção entre as classes vazia e por isso o algoritmo passa mais tempo buscando a melhor separação. E o segundo fator que se percebeu é que a reta da fronteira de decisão tende a iniciar com um perfil horizontal, de forma que quanto mais vertical for a separação entre as classes, mais épocas serão necessárias para a fronteira ir girando e se posicionando de forma vertical.

Apesar de ser uma rede neural que aplica o modelo mais simples de um neurônio, os resultados obtidos demonstram a capacidade do perceptron em separar duas classes, inclusive obtendo boa generalização dos dados mesmo em cenários de sobreposição.

REFERÊNCIA

[1] <https://archive.ics.uci.edu/ml/datasets/iris>