

Exercício 3 - Perceptron Simples

Matheus Bitarães de Novaes

12 de Novembro de 2021

1 Introdução

Esta atividade abordará o treinamento do *Perceptron* simples em 4 exercícios propostos.

2 Exercício 1

Neste exercício, devem-se amostrar duas distribuições Normais no espaço R^2 , ou seja, duas distribuições com duas variáveis cada. As distribuições são caracterizadas como $N(2, 2, \sigma^2)$ e $N(4, 4, \sigma^2)$, conforme figura 1.

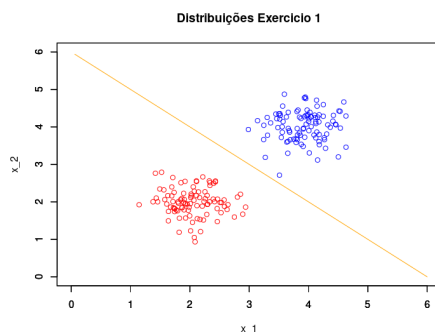


Figure 1: Dados amostrados de duas distribuições Normais, com médias em (2,2) e (4,4)

O Perceptron foi então treinado e a separação das regiões pode ser vista na figura 2 e 3.

3 Exercício 2

Neste exercício, mais amostras das distribuições do Exercício 1 serão geradas e será feita a divisão dos dados em treinamento e teste. O conjunto de

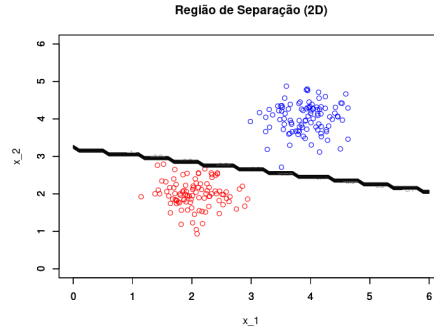


Figure 2: Região de separação definida após treinamento do *Perceptron* simples (visão 2D)

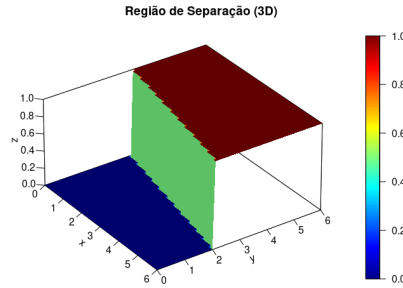


Figure 3: Região de separação definida após treinamento do *Perceptron* simples (visão 3D)

treinamento irá conter 70% das amostras e o de teste 30%. Após isto, os dados de treinamento serão utilizados para o cálculo dos vetores de pesos w enquanto os dados de teste serão utilizados para a medida de acurácia do modelo. A acurácia realizando este procedimento foi de 100%, conforme podemos ver na matriz de confusão, figura 4.

4 Exercício 3

Agora será utilizada a base de dados Iris, que possui 150 amostras e 4 características, sendo 50 para cada uma das três espécies de plantas que constitui a base. Para transformar o problema em binário, o objetivo será a separação da espécie 1 das outras duas espécies. 70% dos dados serão utilizados para treinamento e 30% dos dados serão utilizados para teste, assim como no Exercício 2. O treinamento foi executado 100 vezes e o erro percentual pode ser visto nas

```

Confusion Matrix and Statistics

      truth
pred  0   1
  0  59   0
  1   0  61

      Accuracy : 1
      95% CI : (0.9697, 1)
      No Information Rate : 0.5083
      P-Value [Acc > NIR] : < 2.2e-16

      Kappa : 1

      Mcnemar's Test P-Value : NA

      Sensitivity : 1.0000
      Specificity : 1.0000
      Pos Pred Value : 1.0000
      Neg Pred Value : 1.0000
      Prevalence : 0.4917
      Detection Rate : 0.4917
      Detection Prevalence : 0.4917
      Balanced Accuracy : 1.0000

      'Positive' Class : 0

```

Figure 4: Matriz de confusão do Exercício 2

figuras 5 e 6. É possível ver a distribuição das acurácias na figura 7.

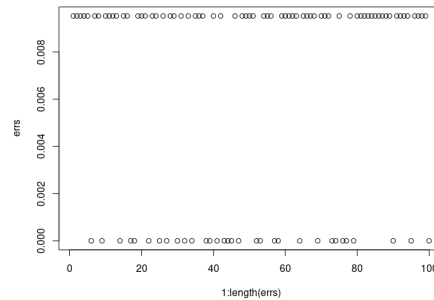


Figure 5: Erro percentual por iteração do treinamento do *Perceptron* para o Iris Dataset (comparando amostra da classe 1 com as demais)

Pode-se notar portanto que a acurácia média foi de 98%.

5 Exercício 4

Neste exercício, foi utilizada a base de dados Breast Cancer para o treinamento do Perceptron. Inicialmente foi necessária a limpeza e normalização dos dados. Após isto, foi feita a divisão entre treinamento e teste. Os dados de treinamento foram utilizados para ajustar os pesos do *Perceptron* e os dados de teste foram utilizados para a avaliação do modelo. Este procedimento foi repetido 30 vezes e a distribuição da acurácia pode ser vista na figura 8. A acurácia média foi de 95,2%

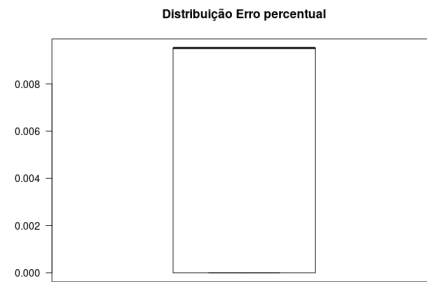


Figure 6: Distribuição do erro percentual do treinamento do Perceptron para o Iris Dataset (comparando amostra da classe 1 com as demais)

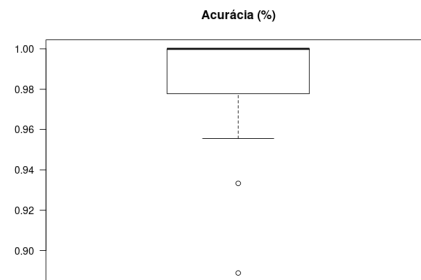


Figure 7: Distribuição da avaliação de acurácia do Perceptron para o Iris Dataset (comparando amostra da classe 1 com as demais)

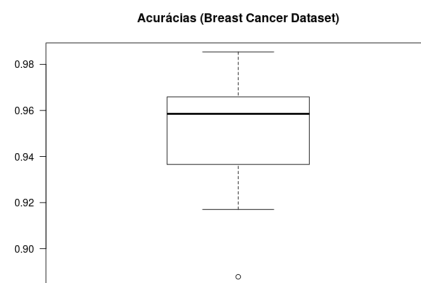


Figure 8: Distribuição das acurácias de 30 execuções