

Exercícios 1 e 2

A.P. Braga

21 de Agosto de 2019

OBJETIVOS DO EXERCÍCIOS

O objetivo dos exercícios dessa semana é desenvolver e praticar os algoritmos *perceptron* e *adaline*, assim como ver seu funcionamento aplicado à problemas reais.

O Aluno deve utilizar as bases de dados sugeridas e resolver o problema associado a cada um utilizando o algoritmo solicitado que deve ser **implementado**, ou seja, não deverão ser utilizados pacotes prontos para a solução dos exercícios.

PROBLEMA 1 - CLASSIFICAÇÃO

- Gere uma base de dados linearmente separável como na Figura 1. Aplique o algoritmo de aprendizado do **Perceptron** para separá-la. Plote uma margem de separação como na Figura 2¹.
- Utilize o dataset *Wisconsin Breast Cancer*, disponível no pacote `mlbench`, e aplique o **Perceptron** para classificar as duas classes "M"(malignant) e "B"(benign). Apresente os resultados da classificação usando métricas como erro/acurácia, sensibilidade e especificidade. Obtenha a matriz de confusão. Discuta os resultados.

PROBLEMA 2 - REGRESSÃO

- Gere uma base de dados linear como na Figura 3, da forma $y = ax + b + \mathcal{N}(0, \sigma^2)$, onde temos um ruído Gaussiano com variância σ^2 . Usar o **Adaline** para aproximar

¹A função `contour`, *built-in* no R, plota o contorno de uma certa função em um plano.

esta função. Plotar a função obtida como na Figura 4. Obtenha o erro médio quadrático. Como ele se compara com a variância do ruído quando o número de pontos é cresce? Discuta.

- Utilize o dataset disponível junto do trabalho pelo Moodle (*BUILDING1paraR.DT*) e aplique o **Adaline** para realizar a regressão e aproximar as funções dos outputs "Energy", "Hot_Water" e "Cold_Water". Apresentar medidas de desempenho da previsão (erro médio, etc), gráficos, etc. Discuta os resultados.

APRESENTAÇÃO DOS RESULTADOS

O resultado deverá ser apresentando em **um único pdf** contendo, para cada problema, um breve resumo do método de solução proposto, desempenho de cada algoritmo e código fonte das partes importantes da implementação.

O aluno pode escolher de que forma quer avaliar o desempenho do algoritmo (métricas de desempenho, gráficos, ...). A entrega será realizada via Moodle.

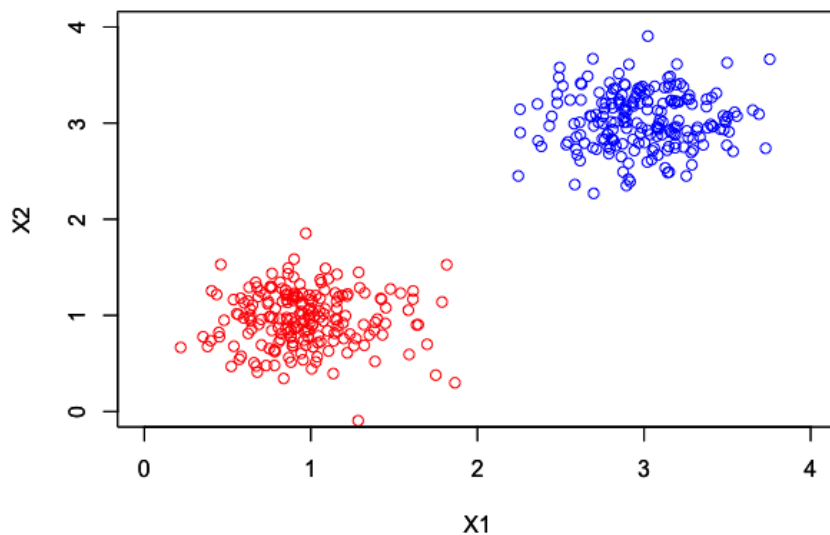


Figura 1: Problema separável constituindo de duas classes (0 e 1) com distribuição normal.

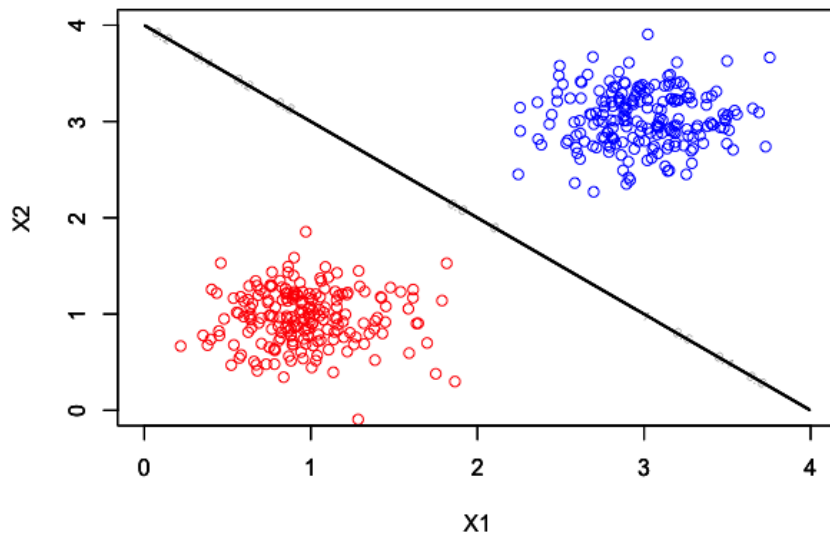


Figura 2: Superfície de separação.

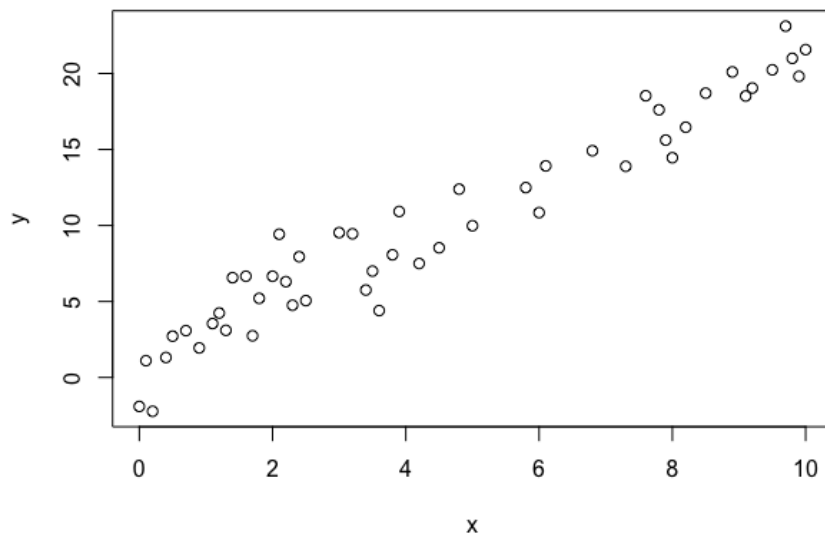


Figura 3: Problema de regressão linear.

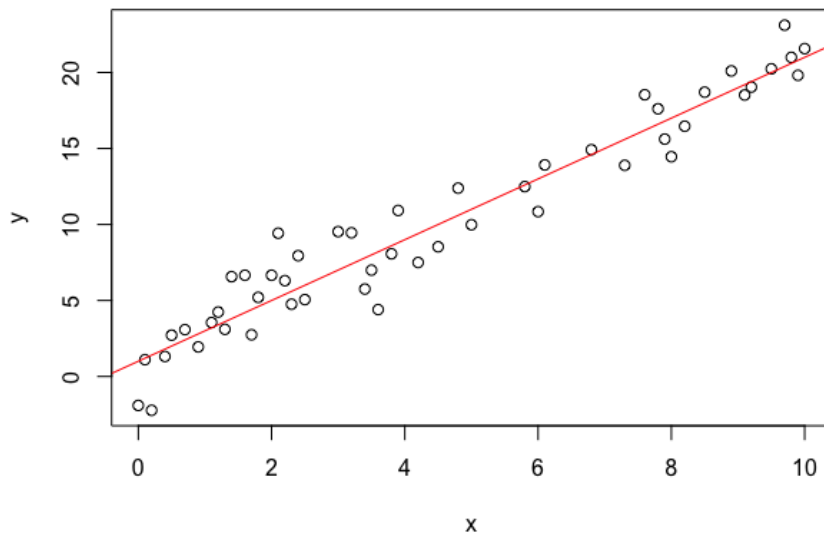


Figura 4: Função linear aprendida para o problema.

DICAS

O `r markdown` e `r sweave` apresentam integração de R com \LaTeX e o RStudio oferece modelos de exemplo do funcionamento das ferramentas. Com essas ferramentas é possível escolher exibir ou não os trechos de código com apenas um comando "T" ou "F" no chunk.

Chunk no sweave:

- início: `«echo = T»`
- fim: `@`

Chunk no markdown:

- início: `““{r, echo = TRUE}`
- fim: `““`

Os pacotes `pROC` e `cvTools` apresentam diversas implementações prontas para a avaliação do desempenho dos seus resultados. Esse item não será cobrado no trabalho.

Além da função `plot`, a biblioteca `ggplot2` contém gráficos visualmente bonitos, com a possibilidade de utilização de várias cores, além de um maior controle do que se quer

exibir por gráfico.

Para instalar novos packages pelo RStudio, use o comando `install.packages("nomeDoPacote")` na aba de console ou vá à aba de **packages** e clique em **install**.