

Universidade Federal do Rio de Janeiro

Instituto Alberto Luiz Coimbra de  
Pós-Graduação e Pesquisa de Engenharia



Programa de Engenharia de Sistemas e  
Computação

CPS844 - Inteligência Computacional I

Prof. Dr. Carlos Eduardo Pedreira

*Trabalho prático*

Luiz Henrique Souza Caldas

email: lhscaldas@cos.ufrj.br

18 de maio de 2024

# 1 Perceptron

Neste problema, você criará a sua própria função target  $f$  e uma base de dados  $D$  para que possa ver como o Algoritmo de Aprendizagem Perceptron funciona. Escolha  $d = 2$  pra que você possa visualizar o problema, e assuma  $\chi = [-1, 1] \times [-1, 1]$  com probabilidade uniforme de escolher cada  $x \in \mathcal{X}$ .

Em cada execução, escolha uma reta aleatória no plano como sua função target  $f$  (faça isso selecionando dois pontos aleatórios, uniformemente distribuídos em  $\chi = [-1, 1] \times [-1, 1]$ , e pegando a reta que passa entre eles), de modo que um lado da reta mapeia pra +1 e o outro pra -1. Escolha os inputs  $x_n$  da base de dados como um conjunto de pontos aleatórios (uniformemente em  $\mathcal{X}$ ), e avalie a função target em cada  $x_n$  para pegar o output correspondente  $y_n$ .

Agora, pra cada execução, use o Algoritmo de Aprendizagem Perceptron (PLA) para encontrar  $g$ . Inicie o PLA com um vetor de pesos  $w$  zerado (considere que  $\text{sign}(0) = 0$ , de modo que todos os pontos estejam classificados erroneamente ao início), e a cada iteração faça com que o algoritmo escolha um ponto aleatório dentre os classificados erroneamente. Estamos interessados em duas quantidades: o número de iterações que o PLA demora para convergir pra  $g$ , e a divergência entre  $f$  e  $g$  que é  $\mathbb{P}[f(x) \neq g(x)]$  (a probabilidade de que  $f$  e  $g$  vão divergir na classificação de um ponto aleatório). Você pode calcular essa probabilidade de maneira exata, ou então aproximá-la ao gerar uma quantidade suficientemente grande de novos pontos para estimá-la (por exemplo, 10.000).

A fim de obter uma estimativa confiável para essas duas quantias, você deverá realizar 1000 execuções do experimento (cada execução do jeito descrito acima), tomando a média destas execuções como seu resultado final.

Para ilustrar os resultados obtidos nos seus experimentos, acrescente ao seu relatório gráficos scatterplot com os pontos utilizados para calcular  $E_{out}$ , assim como as retas correspondentes à função target e à hipótese  $g$  encontrada.

1. Considere  $N = 10$ . Quantas iterações demora, em média, para que o PLA convirja com  $N = 10$  pontos de treinamento? Escolha o valor mais próximo do seu resultado.
2. Qual das alternativas seguintes é mais próxima de  $\mathbb{P}[f(x) \neq g(x)]$  para  $N = 10$ ?
3. Agora considere  $N = 100$ . Quantas iterações demora, em média, para que o PLA convirja com  $N = 100$  pontos de treinamento? Escolha o valor mais próximo do seu resultado/.
4. Qual das alternativas seguintes é mais próxima de  $\mathbb{P}[f(x) \neq g(x)]$  para  $N = 100$ ?
5. possível estabelecer alguma regra para a relação entre  $N$ , o número de iterações até a convergência, e  $\mathbb{P}[f(x) \neq g(x)]$ ?

## 2 Regressão Linear

Nestes problemas, nós vamos explorar como Regressão Linear pode ser usada em tarefas de classificação. Você usará o mesmo esquema de produção de pontos visto na parte acima do Perceptron, com  $d = 2$ ,  $\mathcal{X} = [-1, 1] \times [-1, 1]$ , e assim por diante.

1. Considere  $N = 100$ . Use Regressão Linear para encontrar  $g$  e calcule  $E_{in}$ , a fração de pontos dentro da amostra que foram classificados incorretamente (armazene os  $g$ 's pois eles serão usados no item seguinte). Repita o experimento 1000 vezes. Qual dos valores abaixo é mais próximo do  $E_{in}$  médio?
2. Agora, gere 1000 pontos novos e use eles para estimar o  $E_{out}$  dos  $g$ 's que você encontrou no item anterior. Novamente, realize 1000 execuções. Qual dos valores abaixo é mais próximo do  $E_{out}$  médio?
3. Agora, considere  $N = 10$ . Depois de encontrar os pesos usando Regressão Linear, use-os como um vetor de pesos iniciais para o Algoritmo de Aprendizagem Perceptron (PLA). Execute o PLA até que ele convirja num vetor final de pesos que separa perfeitamente os pontos dentro-de-amostra. Dentre as opções abaixo, qual é mais próxima do número médio de iterações (sobre 1000 execuções) que o PLA demora para convergir?
4. Vamos agora avaliar o desempenho da versão pocket do PLA em um conjunto de dados que não é linearmente separável. Para criar este conjunto, gere uma base de treinamento com  $N_2$  pontos como foi feito até agora, mas selecione aleatoriamente 10% dos pontos e inverta seus rótulos. Em seguida, implemente a versão pocket do PLA, treine-a neste conjunto não-linearmente separável, e avalie seu  $E_{out}$  numa nova base de  $N_2$  pontos na qual você não aplicará nenhuma inversão de rótulos. Repita para 1000 execuções, e mostre o  $E_{in}$  e  $E_{out}$  médios para as seguintes configurações (não esqueça dos gráficos scatterplot, como anteriormente):

### 3 Regressão Não-Linear

Nestes problemas, nós vamos novamente aplicar Regressão Linear para classificação. Considere a função target

$$f(x_1, x_2) = \text{sign}(x_1^2 + x_2^2 - 0.6)$$

Gere um conjunto de treinamento de  $N = 1000$  pontos em  $\mathcal{X} = [-1, 1] \times [-1, 1]$  com probabilidade uniforme escolhendo cada  $x \in \mathcal{X}$ . Gere um ruído simulado selecionando aleatoriamente 10% do conjunto de treinamento e invertendo o rótulo dos pontos selecionados.

1. Execute a Regressão Linear sem nenhuma transformação, usando o vetor de atributos  $(1, x_1, x_2)$  para encontrar o peso  $w$ . Qual é o valor aproximado de classificação do erro médio dentro da amostra  $E_{in}$  (medido ao longo de 1000 execuções)?
2. agora, transforme os  $N = 1000$  dados de treinamento seguindo o vetor de atributos não-linear  $(1, x_1, x_2, x_1x_2, x_1^2, x_2^2)$ . Encontre o vetor  $\tilde{w}$  que corresponda à solução da Regressão Linear. Quais das hipóteses a seguir é a mais próxima à que você encontrou? Avalie o resultado médio obtido após 1000 execuções.
3. Qual o valor mais próximo do erro de classificação fora da amostra  $E_{out}$  de sua hipótese na questão anterior? (Estime-o gerando um novo conjunto de 1000 pontos e usando 1000 execuções diferentes, como antes).