

Redes Neurais II Primeira Lista de Exercícios

Lucas Oliveira David
ld492@drexel.edu
1188972

28/09/2016

Respostas às questões impostas

1.

(a) $J(X, \vec{y}, \theta) = \frac{1}{|X|} \sum_{i=1}^{|X|} \frac{1}{2} (y_i - \text{sign}(a \cdot \vec{x}_i + b))^2$.

(b) $a = 1.00, b = 1.00, \text{loss} = .5$.

(c)

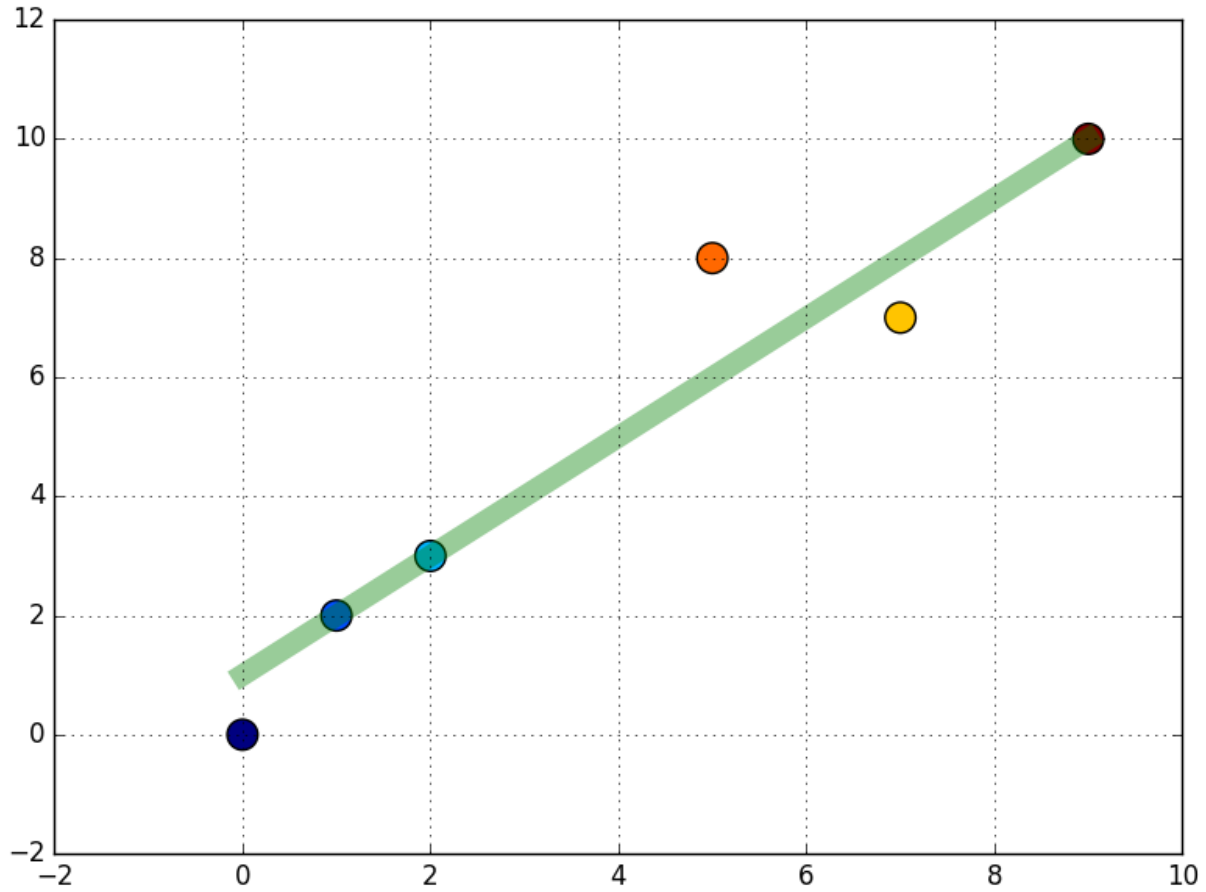


Figure 1: Desenho da linha de regressão.

2. (a)

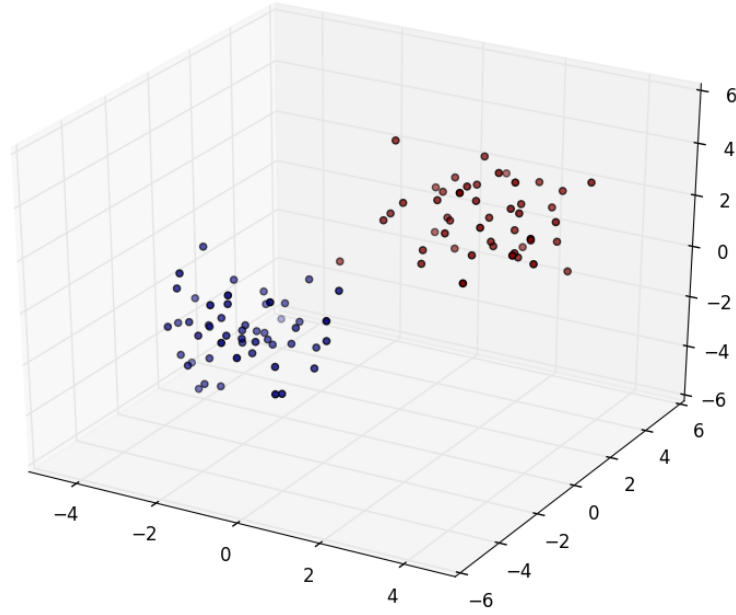


Figure 2: Conjunto de dados 1.

Aparentam ser linearmente separáveis? Sim.

(b) Parâmetros: $\alpha = .1, n_{max_epochs} = 200$.

Resultado: 100% de acerto, 7 epochs necessários para convergência.

Os padrões são linearmente separáveis: eu achei um hiperplano que separa todas as amostras no \mathbb{R}^3 .

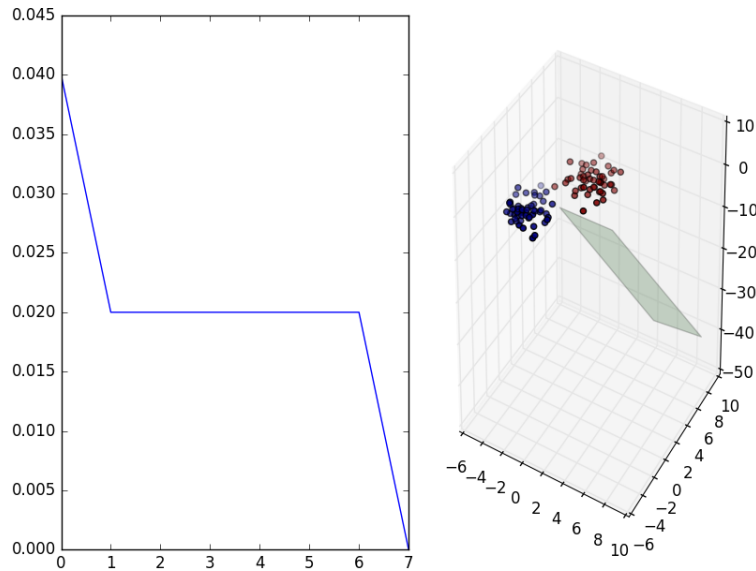


Figure 3: Conjunto 1 e o hiperplano de separação.

- (c) Parâmetros: $\alpha = .0001, n_max_epochs = 2$.
Resultado: 70% de acerto.

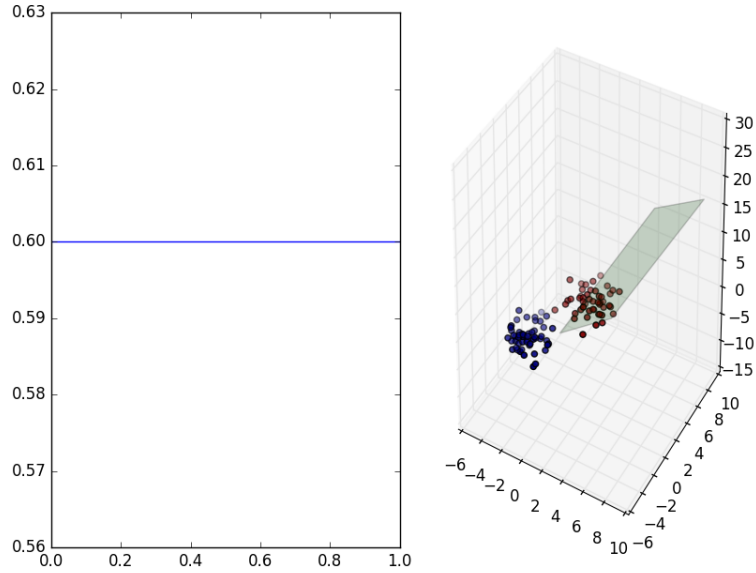


Figure 4: Conjunto de dados 1 e hiperplano de separação.

- (d) Parâmetros: $\alpha = .1, n_max_epochs = 200$.
Resultado: 100% de acerto, 13 epochs necessários para convergência.

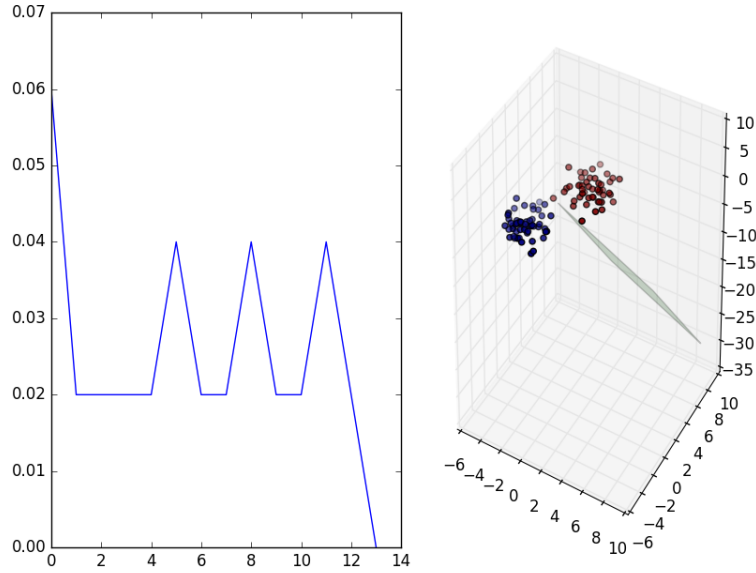


Figure 5: Conjunto de dados 1 e hiperplano de separação.

3. “Grid search” foi empregado para encontrar os melhores parâmetros da rede neural. Neste método, cada combinação de parâmetros é utilizada para instanciar uma rede

diferente. O conjunto de treinamento é separado em 3 “folds”. Duas são utilizadas para treinar o modelo e a última para validá-lo. Finalmente, o modelo com maior acurácia é salvo em memória.

- (a) Parâmetros do *grid*: {'hidden_layer_sizes': [(100,), (200,), (1024,)], 'max_iter': [100, 200, 300]}.

Melhores parâmetros: {'max_iter': 100, 'hidden_layer_sizes': (100,)}.

Best estimator's score in testing fold: 77%.

Score on entire dataset: 89%.

Observa-se que o número de unidades e *epochs* que produziu melhor resultado sobre o conjunto de validação foram ambos 100, os mais baixos. Isso exemplifica o processo de *over-fitting*.

- (b) Parâmetros do *grid*: {'ex_n_features': [32, 64, 128, 256, 512, 1024], 'rg_max_iter': [200], 'rg_hidden_layer_sizes': [(100,), (200,), (1024,), (2048,)]}

Best parameters: 'ex_n_features': 32, 'rg_random_state': 1, 'rg_hidden_layer_sizes': (100,), 'ex_random_state': 0, 'rg_max_iter': 200.

Score on entire dataset: 84%. Score on entire dataset: 95%.

O melhor resultado obtido foi utilizando uma *extreme machine* que projetava as amostras para um espaço de características 32-dimensional, o mais baixo de todos. Ainda sim, houve um aumento na acurácia, em relação à rede treinada em (a).

- (c) Best estimator found in (a)'s score on dataset 3: 90%.

Best estimator found in (b)'s score on dataset 3: 95%.

A pontuação sobre o conjunto de teste (dataset 3) se manteve aproximadamente constante em relação às pontuações obtidas sobre os conjuntos de treinamento e validação. Isso indica que o modelo generalizou adequadamente os padrões contidos no conjunto de treinamento OU que houve *overfitting*, mas que este conjunto de dados contém amostras com baixa variação posicional (pouco ruído).

4. O treinamento do modelo de aprendizagem nesta etapa seguiu os seguintes passos:

- Para cada $c \in \{20, 30, 40, 50, 100\}$:
 - O conjunto de dados foi separado em subconjuntos de treinamento (80%) e validação (20%).
 - *K-means* foi executado com os parâmetros $\{n_clusters : c\}$.
 - Os c centroides encontrados pelo algoritmo *K-means* foram utilizados para iniciar a rede RBF.

- A acurácia do modelo treinado sobre o conjunto de treinamento e validação foi computada.
- o parâmetro c que induz uma acurácia satisfatória foi escolhido. Embora $c = 100$ resulte em uma acurácia de 100%, $c = 50$ produz uma de 99% de maneira mais eficiente.

Parâmetros de treino escolhidos: $\{n_clusters : 50\}$.

Alguns dos 50 centros: [0.26808911 -0.28454831], [1.0557571 -0.92876046], [1.16678465 0.95468916], [-0.08038568 -1.10660401], ...

Acurácia no conjunto de treinamento: 99%

Acurácia no conjunto de validação: 99%