Deep Learning for NLP 2020
Home Exercise 02 Solution

May 4, 2020

1 TensorFlow Playground (3P)

TensorFlow offers a playground for experimenting with neural networks in a web browser[1].

1. Take a look at the circular dataset in figure 1. Can a perceptron architecture (no hidden layers) learn a good discriminator for this dataset? Justify your answer. (1P)

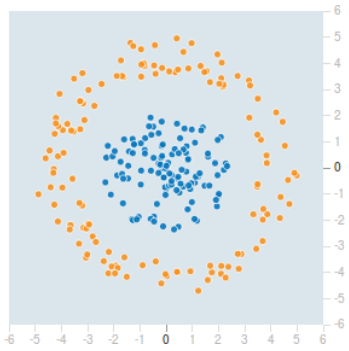

Figure 1: Circular dataset

Answer: It depends on the input features. In the classic setup with inputs are x^1 and x^2, a perceptron cannot learn a proper discriminator because the circular dataset is not linearly separable. If the perceptron is given useful input features such as x^{11} and x^{22}, discrimination is possible.

2. In a multi-layer perceptron, the number of neurons can be chosen differently for each hidden layer. Pick the spiral dataset and specify at least four hidden layers. Then, try out three different scenarios:
   - Same amount of neurons in every hidden layer.
   - More neurons towards the input, less neurons towards the output.
   - Less neurons towards the input, more neurons towards the output.

   Which scenario produces the best results on the test set? Which scenario converges the fastest? Explain in up to three sentences. (2P)

   Answer: The best results and the fastest convergence can be achieved by putting more neurons towards the input layer. Convergence when using the same amount of neurons in every layer is acceptable. Explanation: When putting more neurons towards the input layer, the early layers focus on specific areas of the input data, which are then combined into more complex shapes by the later layers. A network with the same amount of neurons can in every layer can learn the same, but has a large number of idling neurons in later layers which mostly slow

[1] `https://playground.tensorflow.org`

down training. Networks with more neurons towards the output layer lose a lot of information in the early layers, leading to bad results overall.

More information: Putting more neurons towards the input layer is standard practice for neural networks, especially for convolutional neural networks (CNNs) which we will cover in a few weeks. The concept has been published in the 1990s under the name "pyramid rule".

2 Sentiment Polarity in Movie Reviews (7P)

The movie-review dataset[2] consists of movie reviews labeled with sentiment polarity (i.e. "positive review" or "negative review"). Your task is to implement a variation of the perceptron from last week which learns to identify the sentiment in a review.

2.1 Dataset reader (1P)

See hex02_P3_perceptron.py

2.2 Numpy implementation (6P)

See hex02_P3_perceptron.py

[2] http://www.cs.cornell.edu/people/pabo/movie-review-data/