In the name of God

Assignment 2 Solution

Neural Networks: Fall 2021, Dr. Mozayani

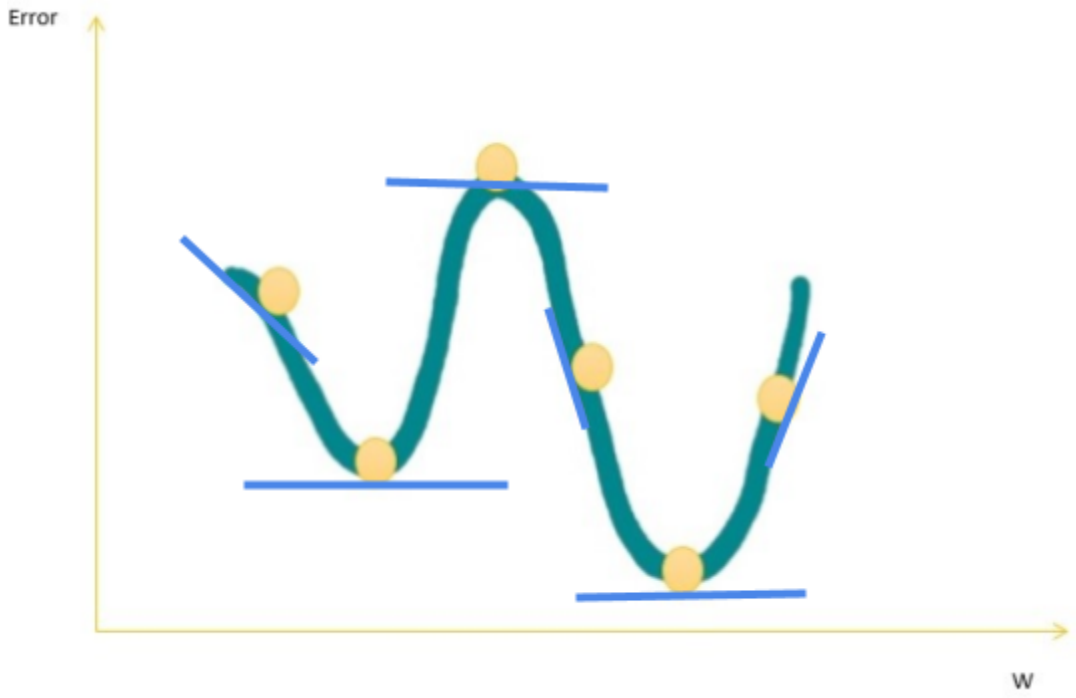
Ghazaleh Mahmoudi

400722156

Saturday, November 22, 2021

Problem 1

- (a) According to Gradient Decent Algorithm, In point number **1,4** weight should be larger to reach the global minimum (right arrow). But in point number **6** weight should be smaller to reach the global minimum. **Point 5 is the global minimum and point 3 is the global maximum**. In points numbers 2,3, 5 weights don't change since the slope is zero.



- (b) we called point number 2 local minimum and point number 3 global maximum. SGD operates by using one **randomly selected data** from the dataset at a time and **updating parameters by calculating the error and loss of this single sample**. This is often preferred because it reduces the chances of

the algorithm getting stuck in local minima instead of finding the global minimum, **Due to a large number of consecutive updates to each sample**. The stochastic nature of the points at which the function is evaluated allows the algorithm to "jump" out of any local minima that may exist. I think another answer is to consider the **high learning rate parameter at the beginning** of the training **but reduce it gradually as we progress** (learning rate decay). Also, instead of SGD, we can speed up convergence by using minibatch gradients and also avoid local minima too.

In practice, the probability of being at the global maximum is zero, because according to the SGD algorithm, we are always moving in the direction of decreasing the gradient, while reaching the glocal maximum requires increasing the gradient.

Problem 2

- ### Shallow vs deep networks

- A Shallow neural network is a kind of neural network that consists of 1 or 2 hidden layers. Although studies show that shallow neural networks can estimate any function, this is not the case in practice and more hidden layers need to explain complex functions.

- A deep neural network(DNN) is an artificial neural network(ANN) with multiple hidden layers between input and output layers. DNN can model complex nonlinear functions.
- For complex problems, adding more layers allows more easy representation and function approximation because of increased computational complexity. So for this problem, deep neural networks are more suitable. But in other problems with low complexity shallow neural networks (one or two hidden layers) is enough.

Problem 3

- 24 input features that are different.
- 24 input neurons,10 hidden layers, 3 output neurons. We can use sigmoid as an activation function and use SGD with backpropagation to update weights. If we do not remove the same input feature, the input dimension change to **64.**

References

Shallow Neural Networks. In this post, I have explained what... | by Rochak Agrawal

A Multilayer Perceptron with NumPy

What is wrong with this multi-layer perceptron backpropagation implementation?

The Multilayer Perceptron - Theory and Implementation of the Backpropagation Algorithm

A Single-Layer Artificial Neural Network in 20 Lines of Python

https://github.com/ghazaleh-mahmoodi/Neural_Networks/blob/master/MLP_with_Numpy.ipynb

https://github.com/amanchadha/coursera-deep-learning-specialization/blob/master/C1%20-%20Neural%20Networks%20and%20Deep%20Learning/Week%203/Planar%20data%20classification%20with%20one%20hidden%20layer/Planar_data_classification_with_onehidden_layer_v6c.ipynb

Explain Deep Neural network and Shallow neural networks?

Appendix

The same feature that was deleted