

Homework 9

Rodrigo Valle

December 8, 2017

1. The accuracy of the model on the training set for me was about 91.9%, as opposed to an accuracy of 91.7% on the test set. This difference in accuracy is because our model has already been shown the data in the training set and has already had the opportunity to minimize cost in these examples. Also, the accuracy is different with each run because of the stochastic gradient descent algorithm used to train our neural net. The fact that the accuracy between the two sets is so similar shows that our model generalizes well.
2. With only 10 iterations of stochastic gradient descent over the training set, our model achieves a significantly decreased accuracy of about 78.1%. Increasing the number of iterations to 10000 yields a slightly higher (compared to 1000 iterations) accuracy of 92.3%. It seems that after 1000 iterations we experience diminishing returns on the accuracy of our model because we've already found the local minimum and further iterations cannot improve the quality of this optimization.

If the model had started overfitting our training data we would expect to see a decrease in the accuracy of our model when run on the test data, but this is not the case here.

3. Initializing W and b with ones rather than zeros doesn't have much of an effect on test accuracy, keeping it at or around 91.9%. In theory, changing these values could allow gradient descent to settle on a different local minimum that is more or less accurate than the one discovered using `tf.zeros`, but in this case it seems to have discovered the same local minimum.