John Fiedler
CSCE 4201

Chapter 4

1. Which Linear Regression training algorithm can you use if you have a training set with millions of features?
    a. Stochastic Gradient Descent would probably be a good choice in this scenario.
2. Suppose the features in your training set have very different scales. Which algorithms suffer from this and how? What can you do about it?
    a. Gradient Descent algorithms struggle with this scenario because it would take a long time for them to converge. I am unsure how to solve it.
3. Can Gradient Descent get stuck in a local minimum when training a Logistic Regression model?
    a. No, it cannot.
4. Do all Gradient Descent algorithms lead to the same model, provided you let them run long enough?
    a. In general, they do approach a similar model, but there are some that deviate. For example, Stochastic Gradient Descent if the learning rate is kept the same.
5. Suppose you use Batch Gradient Descent and you plot the validation error at every epoch. If you notice that the validation error consistently goes up, what is likely going on? How can you fix this?
    a. It is possible that the learning rate is too high. If that is the case, then you can just reduce the learning rate.
6. Is it a good idea to stop Mini-batch Gradient Descent immediately when the validation error goes up?
    a. No, it is not a good idea because it is somewhat random and may not always improve during iteration.
7. Which Gradient Descent algorithm (among those we discussed) will reach the vicinity of the optimal solution the fastest? Which will actually converge? How can you make the others converge as well?
    a. Stochastic Gradient Descent is the fastest. I am unsure which will actually converge.
8. Suppose you are using Polynomial Regression. You plot the learning curves and you notice that there is a large gap between the training error and the validation error. What is happening? What are three ways to solve this?
    a. It is possible that the model is overfitting the training data. To fix this, you can increase the size of the training set, regularize the model, and reduce the polynomial degree.
9. Suppose you are using Ridge Regression and you notice that the training error and the validation error are almost equal and fairly high. Would you say that the model suffers from high bias or high variance? Should you increase the regularization hyperparameter or reduce it?
    a. The model suffers from high bias. You should reduce the regularization hyperparameter in this scenario.
10. Why would you want to use:
    a. Ridge regression instead of plain Linear Regression (i.e., without any regularization?)
        i. Regularization typically improves performance, so Ridge regression would be the better option.
    b. Lasso instead of Ridge Regression?
        i. I am not sure.

   c. Elastic Net instead of Lasso?
      i. There are situations where Lasso can be random or inaccurate, so Elastic Net generally performs better.
11. Suppose you want to classify pictures as outdoor/indoor and daytime/nighttime. Should you implement two Logistic Regression classifiers or one Softmax Regression classifier?
   a. Two Logistic Regression classifiers would be better in this scenario.