Purevmaa Damdinsuren

CPSC 4310/Spring 2020

WR4

**Part 1: Questions from textbook chapter 4 (A.G)**

1. *Suppose the features in your training set have very different scales. Which optimization algorithms might suffer from this and how? What can you do about it?*

If the features in my training set have very different scales, Gradient Descent algorithms will take a long time to converge because the shape of the const function look like big-deep bowl, which is hard to reach its’ bottom.

1. *Which Linear Regression training algorithm can you use if you have a graining set with millions of features and why? Your answer is from one of Batch Gradient Descent, Stochastic Gradient Descent or Mini-batch Gradient.*

I can use Stochastic Gradient Descent because it has the fastest training iteration, meaning it only considers one training instance at each iteration. That is why, it is first to reach global minimum/optimum. If I use Batch Gradient Descent, it will take very, very long time to reach the global optimum even though it gives the most optimal solution. It is also great to use Mini-batch Gradient because it computes the gradients on small random sets of instances at each iteration, so it will be slower than Stochastic with little better solution.

1. *Can gradient descent get stuck in a local minimum when training Linear Regression model with cost function of MSE?*

No. The MSE cost function for a Linear Regression model is convex function, which means that if you pick any two points on the curve, the line segment joining them never crosses the curve. It means that there is just one global minimum. It guaranteed to approach arbitrarily close to the global minimum. However, you have to wait for a while.

1. *What is regularization in Machine Learning (the main purpose)?*

In Machine Learning, we use a training data to create model of purposes, such prediction and classification. If we overfit that training data, the model does not work on test data, which is not good at all. Regularization is used to reduce the overfitting, or we could say constraining the model, so it will be harder to overfit the data.

1. *Name two or regularized version of linear regression algorithms.*
2. Ridge Regression
3. Lasso Regression
4. Elastic Net (which is combination between Ridge and Lasso regression)

**Part 2: Write a reflection on your collaborative work**

1. *What challenges did you (or/and your group) have?*
2. *What was the most satisfying work you (or/and your group) did?*

As my group (Team PJ) worked with group that we have never worked on (Team with Brigit and Quinn), it was very interesting experience. They had a way of collaborating on the notebook live using Google Collab, which gave us a great tip for our DAP. For this specific regression exercise, we did not meet that much challenge because it was just loading and running the command that instructor already provided us. We talked about ROC curve and how it represents the accuracy of the classification model in TW4. Later in TW5, we ran the code, but there was not enough time to discuss about what we saw from it. That is why, we ended up deciding that we will talk about the whole thing on Thursday. At the end, it was great switching gear with different group who has completely different ways of dealing with everything.