

COL341: Machine Learning

Assignment 1 Report

Utkarsh N. Singh, 2015ME10686

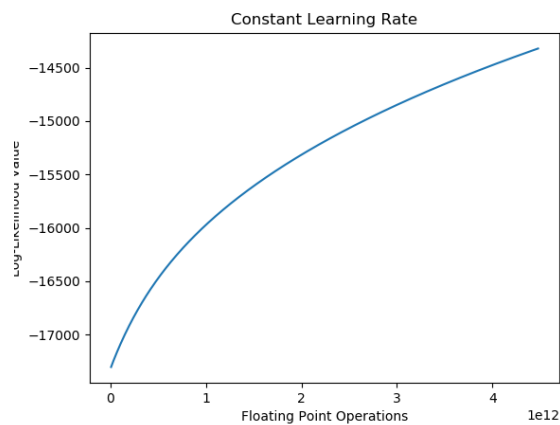
August 21, 2018

Note: For all plots, y-axis is the log-likelihood value.

Part a. - Full Batch Gradient Descent

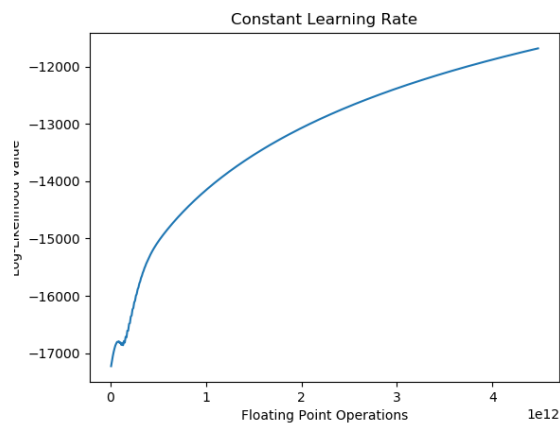
Constant Learning Rate

For learning rate = 0.0000001, 500 iterations the log-likelihood plot looks as follows:



with model accuracy of **74.18%**.

For learning rate = 0.0000005, 500 iterations the log-likelihood plot looks as follows:

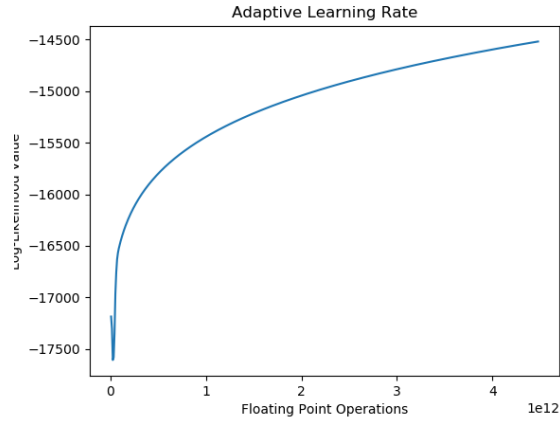


with model accuracy of **80.78%**.

Since we learn on a constant rate, it's important to set a value that isn't too large to prevent overshoot. It can be seen that values small overshooting has already began at a rate as small as 0.0000005.

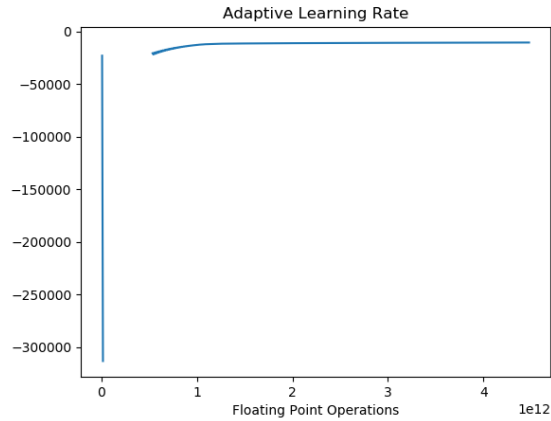
Adaptive Learning Rate

For learning rate = 0.000001, 500 iterations the log-likelihood plot looks as follows:



with model accuracy of **65.376%**.

For learning rate = 0.00001, 500 iterations the log-likelihood plot looks as follows:

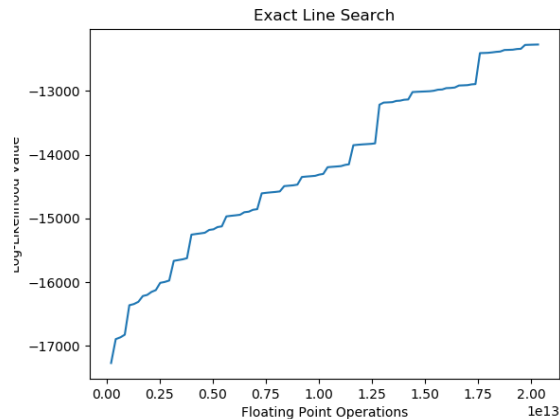


with model accuracy of **82.912%**. (Some values were becoming 'nan' because of the nature of the sigmoid function)

Here as the learning rate decreases with time, it's better to start with a much larger value, and it helps reach the optimum faster without overshooting.

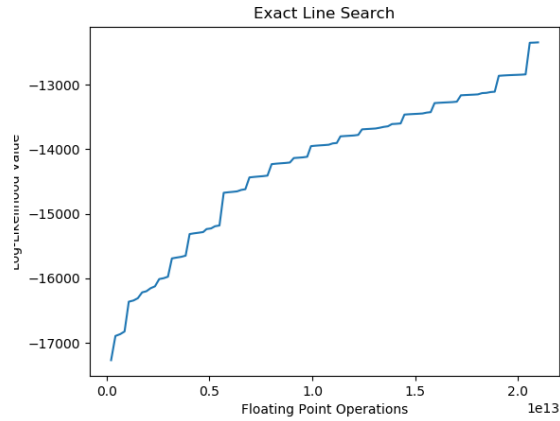
Exact Line Search

For learning rate = 0.000001, 100 iterations the log-likelihood plot looks as follows:



with model accuracy of **79.628%**.

For learning rate = 0.000005, 100 iterations the log-likelihood plot looks as follows:



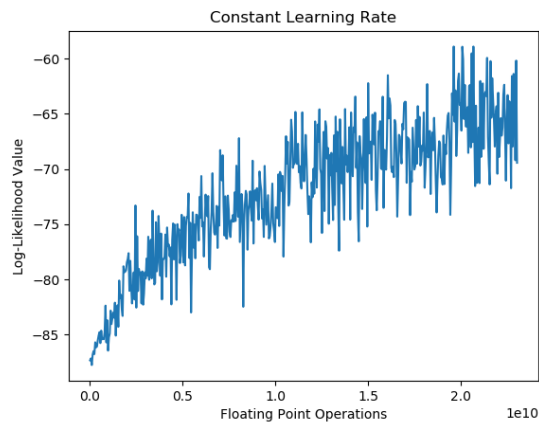
with model accuracy of **79.34%**.

Since Exact Search searches for the optimum value of learning rate before making a gradient descent step, there isn't much difference in the log-probability value, and thus the accuracy.

Part b. - Stochastic Gradient Descent

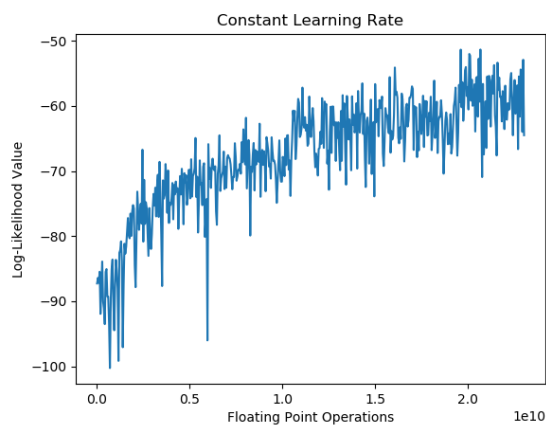
Constant Learning Rate

For learning rate = 0.00005, 500 iterations the log-likelihood plot looks as follows:



with model accuracy of **76.472%**.

For learning rate = 0.0001, 500 iterations the log-likelihood plot looks as follows:

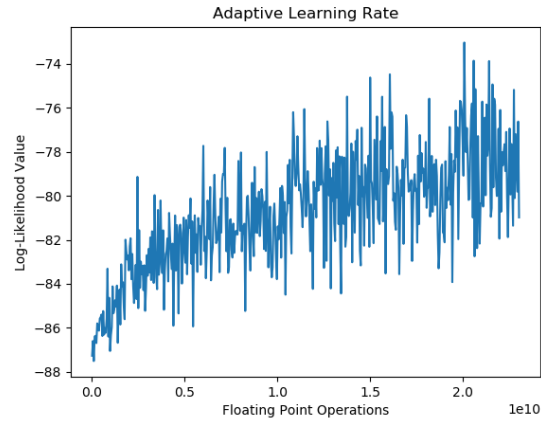


with model accuracy of **74.18%**.

Plot is noisy because at every step we look at a 'new' batch of data, one which we didn't process in the previous step. This affects the direction of gradient, and thus the steps we make towards reaching the optimum. It is relatively difficult to determine now that which learning rate value would be a good one.

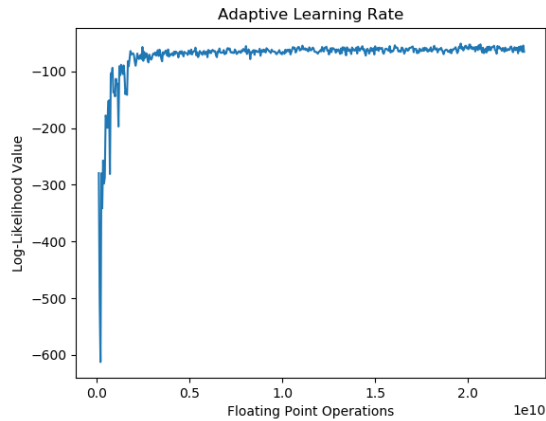
Adaptive Learning Rate

For learning rate = 0.0001, 500 iterations the log-likelihood plot looks as follows:



with model accuracy of **70.06%**.

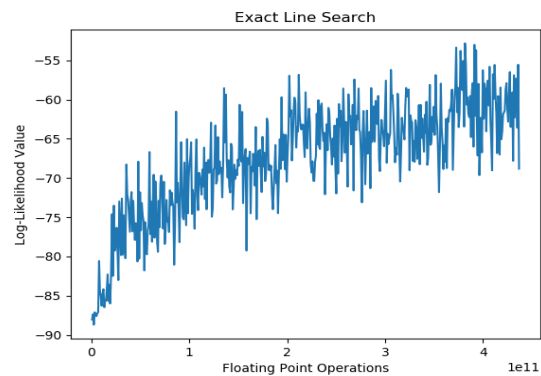
For learning rate = 0.001, 500 iterations the log-likelihood plot looks as follows:



with model accuracy of **79.024%**.

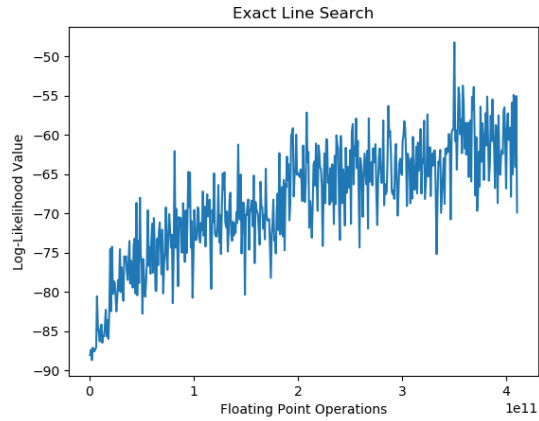
Exact Line Search

For learning rate = 0.000001, 500 iterations the log-likelihood plot looks as follows:



with model accuracy of **79.94%**.

For learning rate = 0.000005, 500 iterations the log-likelihood plot looks as follows:



with model accuracy of **79.464%**.

Part c. - Stopping Criterion

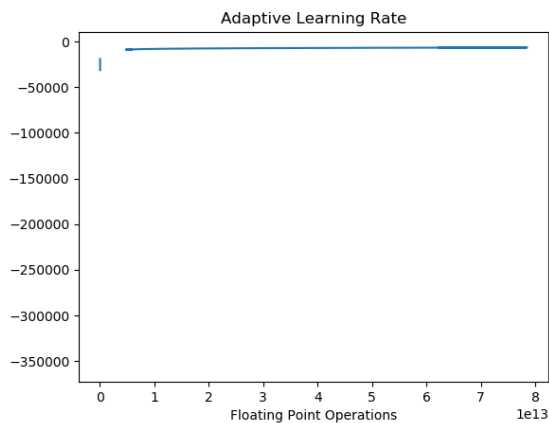
We will experiment with two stopping criteria, on adaptive learning rate.

1. Change in norm of gradient of log-probability
2. Change in log-probability value

We will fix an initial learning rate, and do a full batch gradient descent.

Change in norm of gradient of log-probability

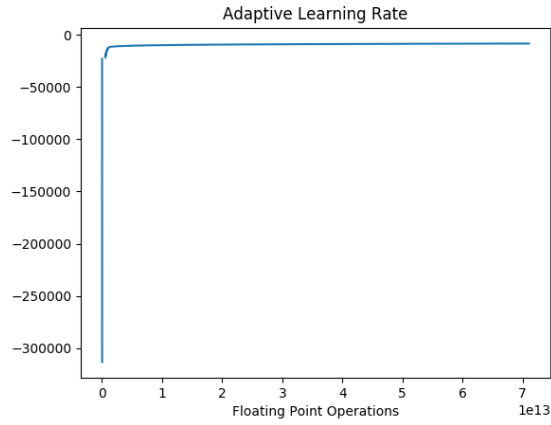
For the above stopping criteria with initial learning rate of 0.00003, the gradient descent ran for 11493 iterations until finally converging on the predefined condition that the change in norm of gradient isn't more than 0.01. The log-likelihood plot looks as follows:



with model accuracy of **87.636%**. (Some values were becoming 'nan' because of the nature of the sigmoid function)

Change in log-probability value

For the above stopping criteria with initial learning rate of 0.00001, the gradient descent ran for 7939 iterations until finally converging on the predefined condition that the change in norm of gradient isn't more than 0.1. The log-likelihood plot looks as follows:



with model accuracy of **85.844%**. (Some values were becoming 'nan' because of the nature of the sigmoid function)

Conclusions

It can be seen clearly that among the three strategies of learning, **Adaptive Learning Rate** seem to perform the best in terms of accuracy and log-probability likelihood value, however there may be few overshoots observed that might sometimes hamper the learning. In that regards, **Exact Line Search** seems to be more robust, because before every step, we ensure to take one that gives us the best result without overshooting. **Constant Learning Rate** requires us to be very careful with the learning rate value, since too small a value will result in slow convergence, but a high value will never take us the optimum.