CMPE462 - Machine Learning - Assignment 2

Berk Atıl - 2016400102

1 Introduction

In the first part of this assignment, we are asked to implement Logistic Regression with 2 different gradient descent approach(Full batch and mini-batch) and apply it on the given data. Additionally we should try 3 different learning rates. In the second part, we should solve a problem related to Naive Bayes Classification approach.

2 Part 1

In this part, we have results for different combinations of different parameters so each combination will be reported in a different sub-section to increase the clarity. 0.00001 is chosen as a threshold to terminate logistic regression algorithm. Also, for each configuration train loss plot for the first cross validation is shown. I applied min-max normalization to reduce the convergence time and also increase the performance. Furthermore, without applying it, the gradient exploded so I got overflow error.

2.1 Step 1

Here, full batch gradient descent is used.

2.1.1 Small Step Size

0.01 is chosen as the small step size. Followings are the number of iterations and total time needed to converge.

- Cross Validation 1: 8793 iterations and 27.378 seconds.
- Cross Validation 2: 9480 iterations and 31.527 seconds.
- Cross Validation 3: 8913 iterations and 28.969 seconds.
- Cross Validation 4: 8930 iterations and 29.004 seconds.
- Cross Validation 5: 9199 iterations and 30.001 seconds.

On Average, 9063 iterations and 29.376 seconds are needed to converge. Additionally average accuracy on cross validation sets is 0.884.

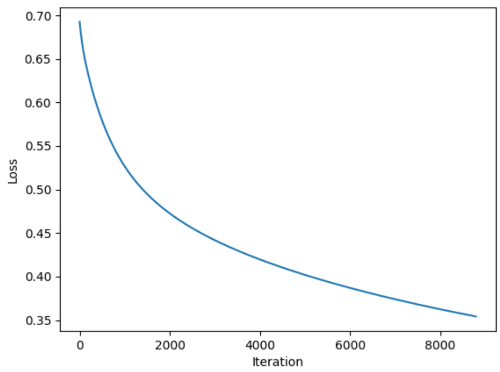

Figure 1: Train loss for step size 0.01 and full batch gradient descent

2.1.2 Medium Step Size

0.1 is chosen as the medium step size. Followings are the number of iterations and total time needed to converge.

- Cross Validation 1: 6789 iterations and 21.589 seconds.
- Cross Validation 2: 6853 iterations and 22.021 seconds.
- Cross Validation 3: 6497 iterations and 21.887 seconds.
- Cross Validation 4: 6450 iterations and 20.738 seconds.
- Cross Validation 5: 6320 iterations and 20.323 seconds.

On Average, 6582 iterations and 21.312 seconds are needed to converge. Additionally average accuracy on cross validation sets is 0.942.

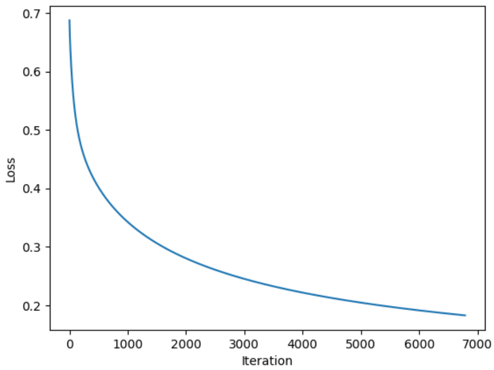

Figure 2: Train loss for step size 0.1 and full batch gradient descent

2.1.3 Large Step Size

1 is chosen as the large step size. Followings are the number of iterations and total time needed to converge.

- Cross Validation 1: 3665 iterations and 12.136 seconds.
- Cross Validation 2: 3874 iterations and 12.358 seconds.
- Cross Validation 3: 3617 iterations and 10.736 seconds.
- Cross Validation 4: 3454 iterations and 10.349 seconds.
- Cross Validation 5: 3611 iterations and 10.657 seconds.

On Average, 3644 iterations and 11.247 seconds are needed to converge. Additionally average accuracy on cross validation sets is 0.966.

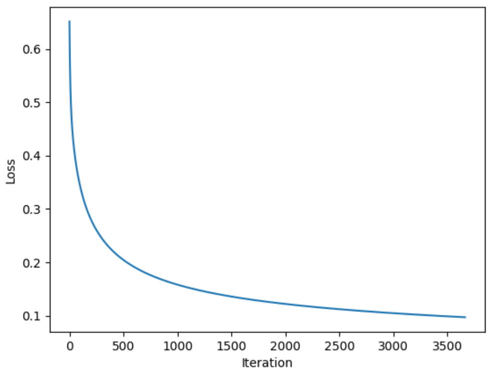

Figure 3: Train loss for step size 1 and full batch gradient descent

2.2 Step 2

Mini batch size is chosen as the train size/5 which is 66 with our data

2.2.1 Small Step Size

0.01 is chosen as the small step size. Followings are the number of iterations and total time needed to converge.

- Cross Validation 1: 7935 iterations and 25.245 seconds.
- Cross Validation 2: 7949 iterations and 26.149 seconds.
- Cross Validation 3: 7577 iterations and 24.582 seconds.
- Cross Validation 4: 7569 iterations and 24.648 seconds.
- Cross Validation 5: 7414 iterations and 24.083 seconds.

On Average, 7689 iterations and 24.941 seconds are needed to converge. Additionally average accuracy on cross validation sets is 0.93.

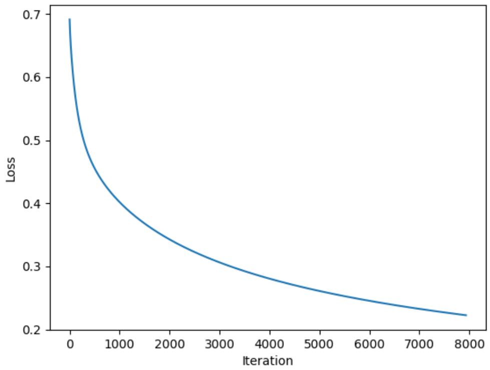

Figure 4: Train loss for step size 0.01 and mini batch gradient descent

2.2.2 Medium Step Size

0.1 is chosen as the medium step size. Followings are the number of iterations and total time needed to converge.

- Cross Validation 1: 4419 iterations and 14.471 seconds.
- Cross Validation 2: 4615 iterations and 15.311 seconds.
- Cross Validation 3: 4304 iterations and 13.965 seconds.
- Cross Validation 4: 4154 iterations and 13.723 seconds.
- Cross Validation 5: 4265 iterations and 13.940 seconds.

On Average, 4351 iterations and 14.282 seconds are needed to converge. Additionally average accuracy on cross validation sets is 0.966.

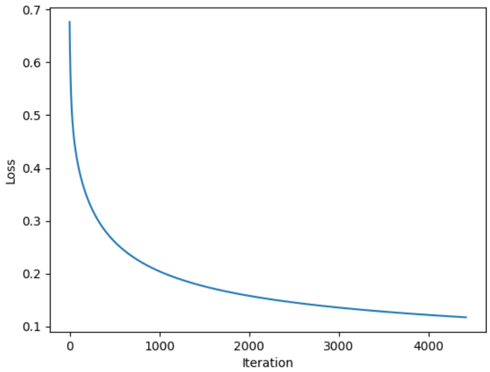

Figure 5: Train loss for step size 0.1 and mini batch gradient descent

2.2.3 Large Step Size

1 is chosen as the large step size. Followings are the number of iterations and total time needed to converge.

- Cross Validation 1: 2297 iterations and 8.152 seconds.
- Cross Validation 2: 2516 iterations and 8.956 seconds.
- Cross Validation 3: 2420 iterations and 8.159 seconds.
- Cross Validation 4: 2293 iterations and 7.919 seconds.
- Cross Validation 5: 2246 iterations and 7.373 seconds.

On Average, 2354 iterations and 8.112 seconds are needed to converge. Additionally average accuracy on cross validation sets is 0.973.

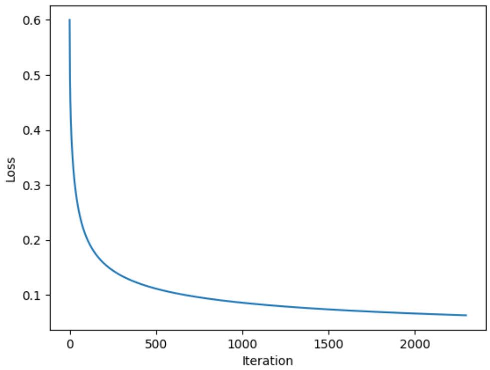

Figure 6: Train loss for step size 1 and mini batch gradient descent

2.3 Analysis of the Results

As we can see from the results, higher learning rate works better for our data. When the step size is higher, the accuracy is higher, we need less number of iterations and less time. This is valid for both full batch and mini-batch gradient descent. Hence, we can infer that 1 is the best step size among 0.01,0.1 and 1 for our data. Secondly, it seems that mini-bath gradient descent approach is better for this data. This is because, again for each step size we need less iterations and time and we get better accuracy score. To sum up, we can conclude the followings:

- Higher step size is better for this data.
- Mini-batch gradient descent with batch size 66 is better than full batch gradient descent for this data.
- In general, we have a stable training because our loss always decreases for each configuration.

3 Part 2

It is asked to find
$P(class = mammals|GiveBirth = yes, canFly = no, LiveInWater = yes, HaveLegs = no)$ and
$P(class = non-mammals|GiveBirth = yes, canFly = no, LiveInWater = yes, HaveLegs = no)$, let's call them X and Y respectively. We assume Naive Bayes that states features are conditionally independent given the class. Followings are needed to calculate X and Y.

- $P(class = mammals) = 7/20$
- $P(class = non - mammals) = 13/20$
- $P(GiveBirth = yes|class = mammals) = 6/7$
- $P(GiveBirth = yes|class = non - mammals) = 1/13$
- $P(canFly = no|class = mammals) = 6/7$
- $P(canFly = no|class = non - mammals) = 10/13$
- $P(LiveInWater = yes|class = mammals) = 2/7$
- $P(LiveInWater = yes|class = non - mammals) = 3/13$

- $P(HaveLegs = no|class = mammals) = 2/7$
- $P(HaveLegs = no|class = non-mammals) = 4/13$

Hence, $X = P(GiveBirth = yes|class = non-mammals) * P(canFly = no|class = non-mammals) * P(LiveInWater = yes|class = non-mammals) * P(HaveLegs = no|class = non-mammals * P(class = non-mammals) = \frac{6}{7} * \frac{6}{7} * \frac{2}{7} * \frac{2}{7} * \frac{7}{20} = 0.02099$

$Y = P(GiveBirth = yes|class = mammals) * P(canFly = no|class = mammals) * P(LiveInWater = yes|class = mammals) * P(HaveLegs = no|class = mammals * P(class = mammals) = \frac{1}{13} * \frac{10}{13} * \frac{3}{13} * \frac{4}{13} * \frac{13}{20} = 0.00273$

X>Y so it is classified as <u>mammals</u>.