

Maria Babaeva

Assignment 3 Machine Learning

Practice: Implementation Logistic Regression

Write a Code that takes as input a weight vector w and a dataset and outputs the following:

a) Logistic Regression Cross-Entropy (negative log-likelihood) for different learning rates (etas). Here is part of the full result:

```
256
  0.1050252  0.08936254 -0.0388346 -0.02176368  0.03206196  0.01024886
  0.06557524  0.04502599  0.03986678  0.03837857 -0.08157235  0.04587493
  0.00065423 -0.0154914  0.01397267  0.00403812 -0.12022072]]
eta=0.01, epoch 297, negative log-likelihood 0.6405, w=[[ 0.11287143 -0.05874976  0.10670063 -0.22117744  0.04102681  0.05748
064
  0.10515255  0.08943741 -0.03898574 -0.02188838  0.03206221  0.01015348
  0.06564662  0.04502579  0.03992402  0.03843455 -0.08173084  0.04598562
  0.00061901 -0.01547834  0.01400288  0.00402979 -0.12049822]]
eta=0.01, epoch 298, negative log-likelihood 0.6404, w=[[ 0.11296453 -0.05891004  0.10702087 -0.22168694  0.04103937  0.05747
836
  0.10527943  0.08951183 -0.03913679 -0.022013  0.03206241  0.01005802
  0.06571781  0.04502544  0.03998125  0.03849043 -0.08188895  0.04609601
  0.00058383 -0.01546528  0.01403309  0.00402144 -0.12077509]]
eta=0.01, epoch 299, negative log-likelihood 0.6404, w=[[ 0.11305784 -0.05907036  0.10734069 -0.22219539  0.04105173  0.05747
572
  0.10540584  0.08958581 -0.03928777 -0.02213754  0.03206256  0.00996248
  0.06578882  0.04502494  0.04003846  0.0385462  -0.08204668  0.04620609
  0.00054866 -0.01545221  0.01406328  0.00401308 -0.12105134]]
```

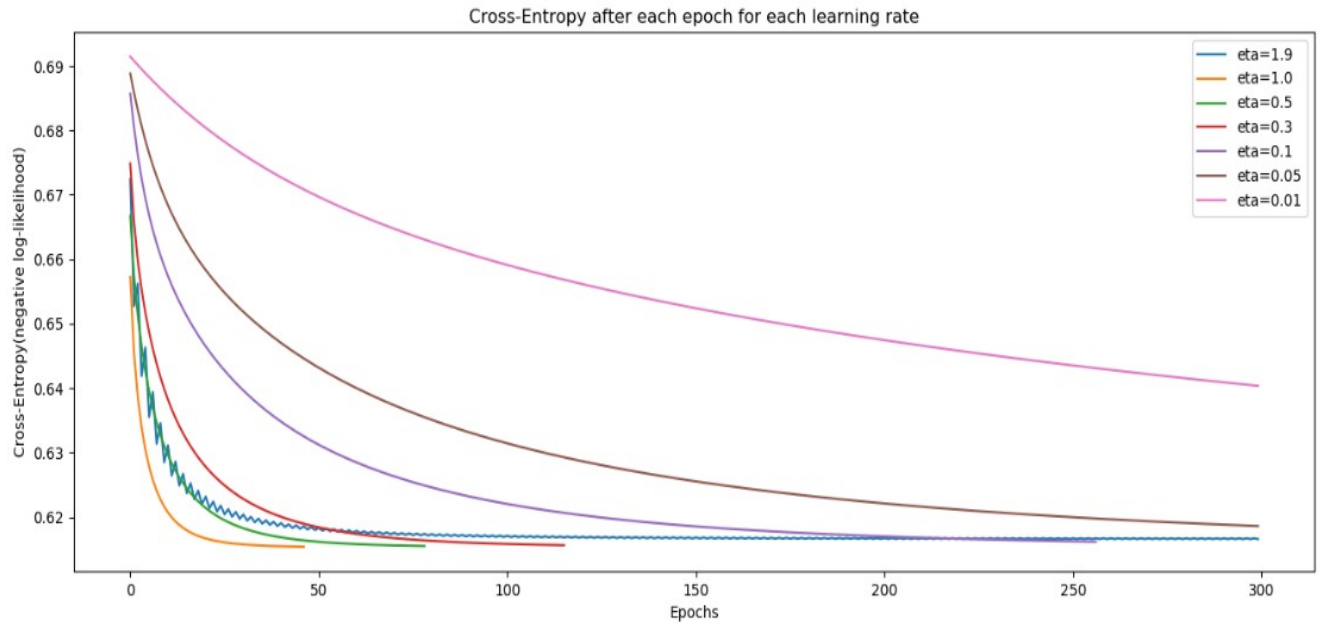
b) The classification accuracy of w . Output the percentage of correctly labelled examples with different learning rate. The best accuracy on the test data is 0.63 with learning rate = 0.05.

```
{0.01: 0.62303664921465973,
0.05: 0.63350785340314131,
0.1: 0.61780104712041883,
0.3: 0.62303664921465973,
0.5: 0.61780104712041883,
1.0: 0.61780104712041883,
1.9: 0.61780104712041883}
```

Experiments:

1) the Bigger a learning rate the faster our Gradient descent converge. However if a learning rate is more than 1.9 our gradient start having problems to converge because a step is too big and it starting having problems in finding the global minimum.

The minimum possible Cross - Entropy is about 0.62.



2) The best learning rate from the plot is 1. It converges to the weight vector with the best fit to the training data (lowest Cross-entropy). Fast convergence and we don't need to make too many iterations.

The best accuracy of w on the test data is 0.63 with learning rate $= 0.05$.

We can compare this accuracy with other methods:

Decision Tree ID3 showed 68.6% and Naive Bayes showed 56.5% accuracy .