Abhijit Kulkarni

CS 190I Homework 3

L2 Regularization Report

I did not work with any collaborators on this homework assignment.

In creating our logistic regression model, there were two steps. First was following the pseudocode in order to implement the model, and the other step was changing the values of our 3 hyper-parameters in order to increase the accuracy of the model. I utilized 10 cross validation over the training set in order to figure out the best values, and tweaked them slightly from there to obtain my final /f-measure of around 73%.

For this report, I kept my learning rate at 0.855, my learning rate decay at 0.95, and varied the L2 decay value as stated in the homework description. I did this over 10 s of the training set, which means that I segmented the training set into 10 different sets, and then proceeded to train over each one individually while testing over all the rest. This allows me to see how changing the values affect over varied training documents (reviews). Here are the statistics over all the folds, where the best L2 value is shown in bold for that fold.

**Fold 1**

|  |  |
| --- | --- |
| L2\_Decay Value | Accuracy |
| 0 | 0.5989 |
| 1e-07 | 0.6011 |
| 1e-06 | 0.6059 |
| 1e-05 | 0.5981 |
| 0.0001 | 0.6111 |
| 0.001 | **0.6178** |
| 0.01 | 0.6056 |
| 0.1 | 0.5782 |
| 1 | 0.5374 |

**Fold 2**

|  |  |
| --- | --- |
| L2\_Decay Value | Accuracy |
| 0 | 0.5983 |
| 1e-07 | 0.5967 |
| 1e-06 | 0.5774 |
| 1e-05 | 0.5926 |
| 0.0001 | 0.5989 |
| 0.001 | **0.6015** |
| 0.01 | 0.5859 |
| 0.1 | 0.5848 |
| 1 | 0.5740 |

**Fold 3**

|  |  |
| --- | --- |
| L2\_Decay Value | Accuracy |
| 0 | 0.5611 |
| 1e-07 | 0.5678 |
| 1e-06 | 0.5552 |
| 1e-05 | 0.5726 |
| 0.0001 | 0.5659 |
| 0.001 | **0.5744** |
| 0.01 | 0.5730 |
| 0.1 | 0.5585 |
| 1 | 0.5374 |

**Fold 4**

|  |  |
| --- | --- |
| L2\_Decay Value | Accuracy |
| 0 | 0.5781 |
| 1e-07 | **0.5791** |
| 1e-06 | 0.5748 |
| 1e-05 | 0.5756 |
| 0.0001 | 0.5770 |
| 0.001 | 0.5730 |
| 0.01 | 0.5700 |
| 0.1 | 0.5496 |
| 1 | 0.5319 |

**Fold 5**

|  |  |
| --- | --- |
| L2\_Decay Value | Accuracy |
| 0 | 0.5356 |
| 1e-07 | 0.5452 |
| 1e-06 | 0.5448 |
| 1e-05 | 0.5389 |
| 0.0001 | 0.5467 |
| 0.001 | **0.5537** |
| 0.01 | 0.5511 |
| 0.1 | 0.5456 |
| 1 | 0.5363 |

**Fold 6**

|  |  |
| --- | --- |
| L2\_Decay Value | Accuracy |
| 0 | 0.5470 |
| 1e-07 | 0.5441 |
| 1e-06 | 0.5404 |
| 1e-05 | **0.5519** |
| 0.0001 | 0.5463 |
| 0.001 | 0.5459 |
| 0.01 | 0.5478 |
| 0.1 | 0.5296 |
| 1 | 0.5163 |

**Fold 7**

|  |  |
| --- | --- |
| L2\_Decay Value | Accuracy |
| 0 | 0.5215 |
| 1e-07 | **0.5289** |
| 1e-06 | 0.5252 |
| 1e-05 | 0.5278 |
| 0.0001 | 0.5263 |
| 0.001 | 0.5281 |
| 0.01 | 0.5270 |
| 0.1 | 0.5181 |
| 1 | 0.4937 |

**Fold 8**

|  |  |
| --- | --- |
| L2\_Decay Value | Accuracy |
| 0 | 0.5219 |
| 1e-07 | 0.5267 |
| 1e-06 | 0.5215 |
| 1e-05 | 0.5270 |
| 0.0001 | 0.5274 |
| 0.001 | **0.5307** |
| 0.01 | 0.5219 |
| 0.1 | 0.5111 |
| 1 | 0.4993 |

**Fold 9**

|  |  |
| --- | --- |
| L2\_Decay Value | Accuracy |
| 0 | 0.5356 |
| 1e-07 | 0.5356 |
| 1e-06 | **0.5437** |
| 1e-05 | 0.5322 |
| 0.0001 | 0.5344 |
| 0.001 | 0.5289 |
| 0.01 | 0.5285 |
| 0.1 | 0.5119 |
| 1 | 0.4981 |

**Fold 10**

|  |  |
| --- | --- |
| L2\_Decay Value | Accuracy |
| 0 | 0.5034 |
| 1e-07 | **0.5043** |
| 1e-06 | 0.4956 |
| 1e-05 | 0.4967 |
| 0.0001 | 0.4970 |
| 0.001 | 0.5007 |
| 0.01 | 0.4956 |
| 0.1 | 0.4974 |
| 1 | 0.4956 |

**Average of all Folds**

|  |  |
| --- | --- |
| L2\_Decay Value | Accuracy |
| 0 | 0.5501 |
| 1e-07 | 0.5529 |
| 1e-06 | 0.5485 |
| 1e-05 | 0.5513 |
| 0.0001 | 0.5531 |
| 0.001 | **0.5555** |
| 0.01 | 0.5506 |
| 0.1 | 0.5384 |
| 1 | 0.5220 |

Now, here are graphs of everything. In order to graph this properly, I took the log of the l2\_decay’s and added 8 to them. Since log(0) is undefined, I kept it as 0. This means 0 corresponds to 0, 1:1e-07, 2:1e-06, 3:1e-05, 4:0.0001, 5:0.001, 6:0.01, 7:0.1, and 8:1, as can be seen in table below.

|  |  |
| --- | --- |
| L2\_Decay Value | Value on X-Axis |
| 0 | 0 |
| 1e-07 | 1 |
| 1e-06 | 2 |
| 1e-05 | 3 |
| 0.0001 | 4 |
| 0.001 | 5 |
| 0.01 | 6 |
| 0.1 | 7 |
| 1 | 8 |

The x-axis plots lambda/L2 decay values, while y-axis plots accuracy for each fold. Here all the graphs, starting off with the averages across all folds. Something around .001 has the best overall accuracy, as can be seen in majority of graphs and also the average of all folds. 1e-7 is also good.