**Logistic Regression on Cleveland Dataset**

**A close up of a person

Description automatically generated**

**A close up of a newspaper

Description automatically generated**

(a)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Maximum # of iterations | Ein (cross-entropy error) | Binary Classification Error on training data | Binary Classification Error on testing data | Time that training process took (sec) |
| 104 | 0.5847 | 0.3092 | 0.3172 | 0.129904 |
| 105 | 0.4937 | 0.2237 | 0.2069 | 1.011837 |
| 106 | 0.4354 | 0.1513 | 0.1310 | 9.306209 |

As we can see from the data above, the binary classification error on testing data is basically the same as the binary classification error on training data. Therefore, the generalization error is pretty small. That is because the model complexity is not very high, and we have enough number of data points.

Also, as the data above shown, when we increase the maximum number of iterations, although we spend more time in training, all of the cross-entropy error, the binary classification error on training data, and the binary classification error on testing data decrease. Namely, increasing the maximum number of iterations will increase the performance of the model.

(b)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Ein (cross-entropy error) | Binary Classification Error on training data | Binary Classification Error on testing data | Time that training process took (sec) |
| glmfit | 0.4074 | 0.1711 | 0.1103 | 0.066556 |

(c)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| eta | Ein (cross-entropy error) | Binary Classification Error on testing data | Number of iterations required to terminate | Time that training process took (sec) |
| 0.01 | 0.4074 | 0.1103 | 23368 | 0.255172 |
| 0.1 | 0.4074 | 0.1103 | 2333 | 0.032722 |
| 1 | 0.4074 | 0.1103 | 230 | 0.007798 |
| 4 | 0.4074 | 0.1103 | 54 | 0.005308 |
| 6 | 0.4074 | 0.1103 | 32 | 0.004477 |
| 7 | 0.4074 | 0.1103 | 43 | 0.004854 |

From the data above, we can see that normalizing the data will positively affect the performance of the model a little bit. Both the cross-entropy error and the binary classification error on testing data decrease a little bit.

When we increase the learning rate eta, the number of iterations decreases, and we spend less time in training. i.e. we find the optimal solution faster because each step we move becomes larger. However, we cannot increase eta too much. As illustrated in the last two rows, when we increase eta from 6 to 7, the number of iterations actually increases a little bit. Moreover, the cross-entropy error and the binary classification error on testing data remain the same as we increase eta.