Fabian Meyer, 6816480
Theresa Naß, 7122946

Gaussian Discriminant Analysis

In this exercise, we solved a Binary Classification problem using Gaussian Discriminant Analysis (GDA). We aim to classify images wether they contain Chagas parisite or not as shown in Image 1. To build a classifier, there are 30 examples available each. From these example we create six features for each image, which represent the image and are utilized to draw classification rules from. As the examples are relatively easy to distinguish, we decide to use simple features, even though there ae very likely features to create which construct a more robust model.

We decide to test the features minimal, maximal, and average color value for each

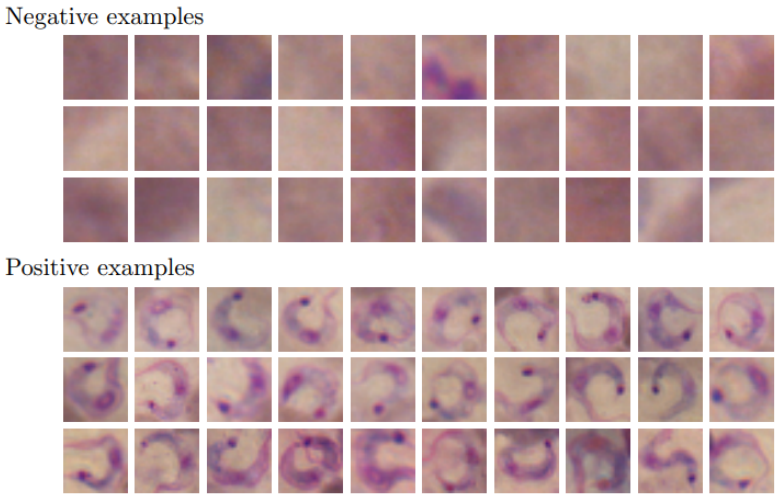

Figure 1: Chagas parasites

channel (RGB), mainly because the the positive examples of Chagas parasites have a very color intensive spot, which distinguishes them from the non-chagas examples.

Final Model

We decided to test combination of features, but each time we use six. The following table illustrates four examples. We evaluated our model with the classification accuracy.

Features	**Accuracy**
R_max, G_max, B_max, R_avg, G_avg, B_avg	0.93
R_max, G_max, B_max, R_min, G_min, B_min	0.98
R_avg, G_avg, B_avg, R_min, G_min, B_min	0.98
R_max, R_avg, G_avg, B_avg, G_min, B_min	0.98

https://github.com/gitfabianmeyer/ml_ss20

Each set of these features were able to reach good classification results, e.g. Accuracy ≥ 0.9 and many were able to reach the maximum of .98. In the following, we report the parameter of the final model with R_max, R_avg, G_avg, B_avg, G_min, B_min:

$$\Sigma = \begin{bmatrix} 221.08277 & 133.87888 & 118.72046 & 176.06197 & 113.03316 & 48.43388 \\ 133.87888 & 186.98444 & 129.04273 & 170.19188 & 128.62404 & 67.59777 \\ 118.72046 & 129.04273 & 172.54299 & 192.66638 & 139.31006 & 108.64623 \\ 176.06197 & 170.19188 & 192.66638 & 252.26656 & 170.48462 & 109.60448 \\ 113.03316 & 128.62404 & 139.31006 & 170.48462 & 130.96070 & 91.46402 \\ 48.43388 & 67.597777 & 108.64623 & 109.60448 & 91.46402 & 121.16555 \end{bmatrix}$$

$$\mu_1 = \begin{bmatrix} 47.1667 & 104.73333 & 176.17256 & 139.52424 & 142.91741 & 209.03333 \end{bmatrix}$$

$$\mu_0 = \begin{bmatrix} 102.8 & 103.4 & 163.7130 & 126.8280 & 127.3424 \end{bmatrix}$$

$$\Phi = 0.5$$

Conclusion

The implementation of the GDA is straightforward and the implementation runs extremely fast, for six features and 60 examples $t < 1s$. Even though we did not use sophisticated features, the results are very good, which leads to the conclusion, that the classifier is very suited for this classification problem.