Machine Learning Exercise 3

Laszlo Korte, MtrNr. 6329857

Universität Hamburg — November 19, 2019

Resulting model

$$\phi = 0.5 \qquad \mu_0 = \begin{pmatrix} 0.4107189542 \\ 0.6483660131 \\ 0.5604604666 \\ 0.6908728214 \\ 17.5510170262 \end{pmatrix} \qquad \mu_1 = \begin{pmatrix} 0.4054901961 \\ 0.5881045752 \\ 0.4993784041 \\ 0.6420118464 \\ 1.3978883483 \end{pmatrix}$$

$$\sigma = \begin{pmatrix} 0.0028755778 & 0.0015319920 & 0.0019780707 & 0.0019845095 & -0.0784795009 \\ 0.0015319920 & 0.0019355718 & 0.0017154186 & 0.0017486707 & 0.0007077625 \\ 0.0019780707 & 0.0017154186 & 0.0020140054 & 0.0021424077 & -0.0202835657 \\ 0.0019845095 & 0.0017486707 & 0.0021424077 & 0.0026534871 & 0.0018864772 \\ -0.0784795009 & 0.0007077625 & -0.0202835657 & 0.0018864772 & 22.9337304741 \end{pmatrix}$$

Features

As features I choose $\text{Red}_{\min}$, $\text{Red}_{\max}$, Red_{avg}, Blue_{avg}, and Edge score.

The edge score is calculated by applying a convolution matrix C to the image and summing the the square of the resulting values.

$$C = \begin{pmatrix} -1 & -1 & -1 \\ -1 & +8 & -1 \\ -1 & -1 & -1 \end{pmatrix}$$

The edge score is high for images containing high contrast edges. Applied to the 60 samples the edge score is already be enough to detect parasites.

But in a real world example there would probably be images containing edges but no parasites. So I choose 4 additional features based on the the image color. I used the min, max and average of the red channel because the parasites are marked by red color. In RGB coding white pixels have high red values as well. That's why we need a second color channel for comparison. I choose the average value of the blue channel. Assuming all images have about the same total lumiance the average green channel can be ommitted because it's inversely corelated to the sum of average red and average blue.

Result

All samples are classified correctly even if only half of the samples are used for training.