MACHINE LEARNING:

Expectation Maximization

Professor: Dr. Maricor Soriano
Instructor: Jayson Cubero

Rene L. Principe Jr.
2015-04622

dataset

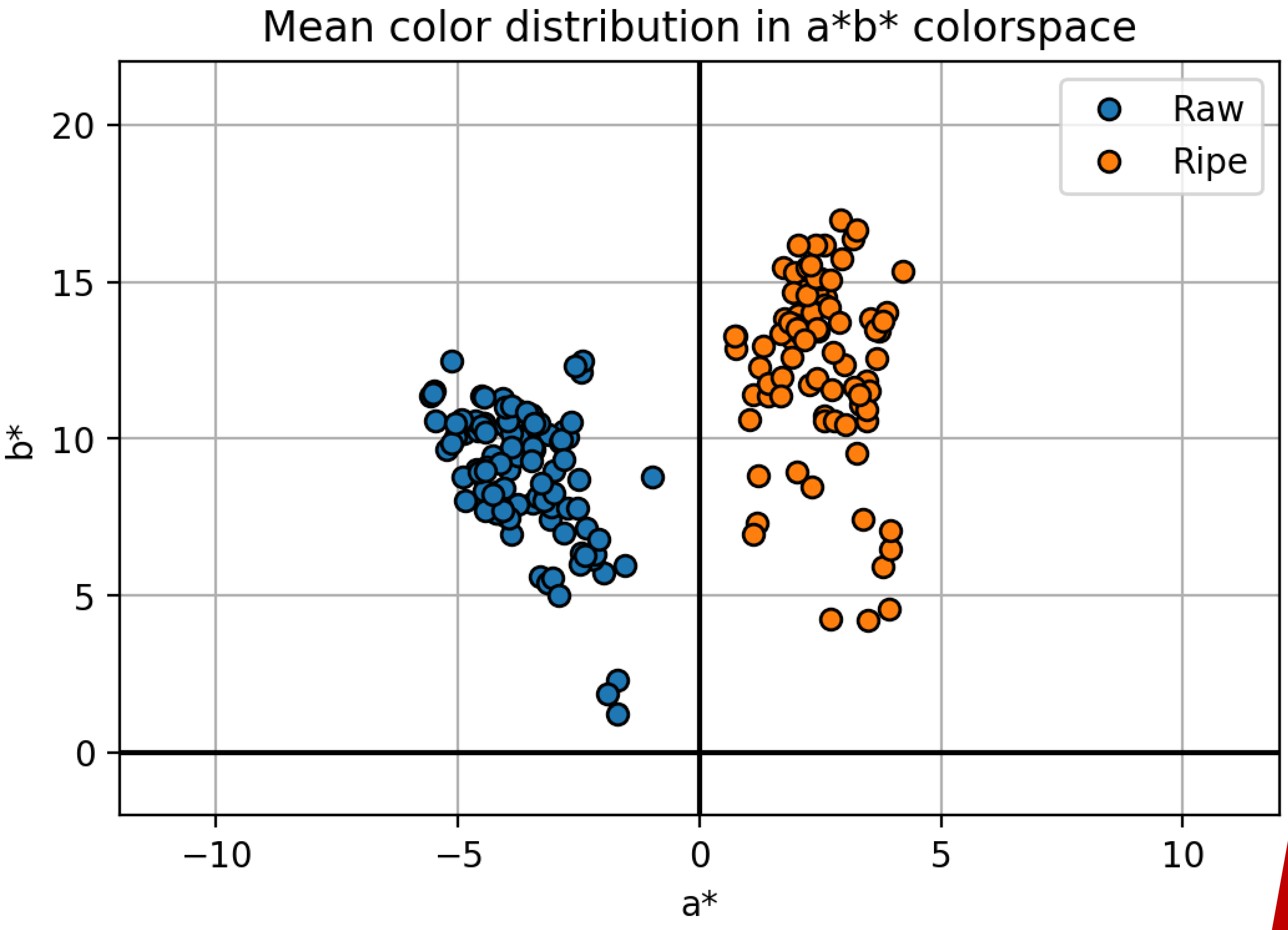
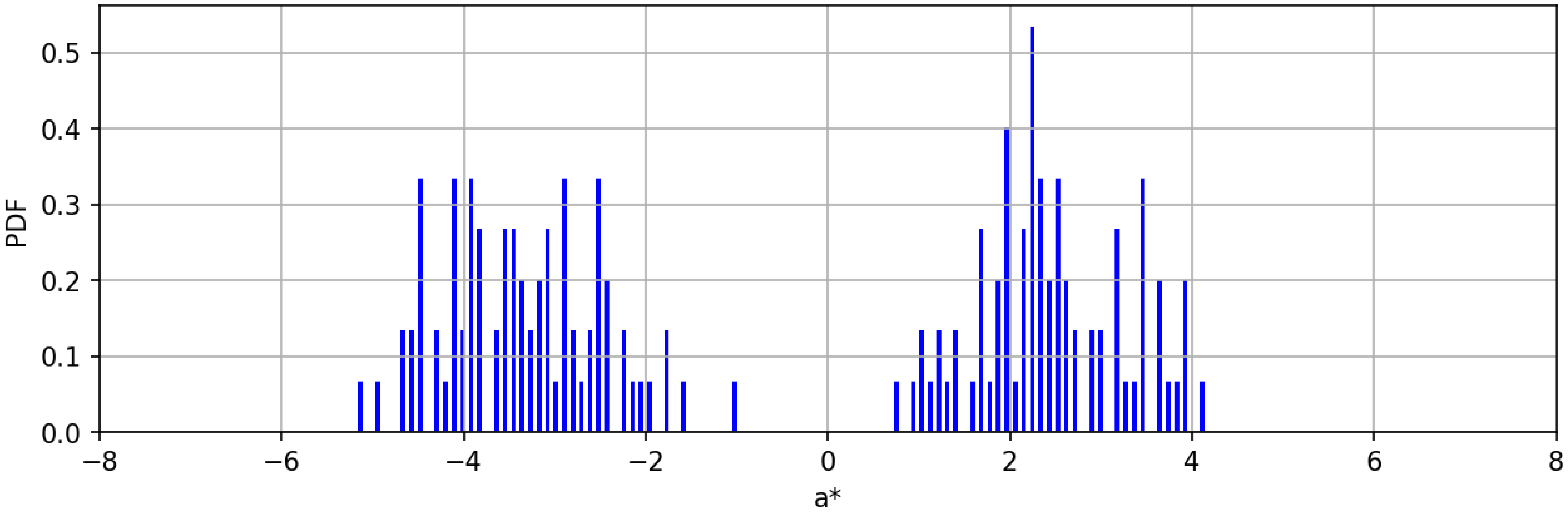


Figure 1. Color features a* and b* of raw and ripe bananas shall be used to determine a continuous ripeness variation.

The goal of this activity is to provide a comprehensive probability map by deriving the optimized gaussian fits from the calculated mean and covariance representation. In this activity, we'll be using the color features a* and b* as shown in Figure 1 and employ the expectation maximization algorithm to derive the optimum mean and covariance. From this, respective probability distribution functions (assuming it is Gaussian) shall be computed. In short, from a discrete number of sample points, we can predict the maximum-likelihood of all possible points.

I plotted the feature histogram in Figure 2 to show how a* feature is separable into two gaussians but for b*, classes are indistinguishable. Hence, we use this data to constantly update our parameters.



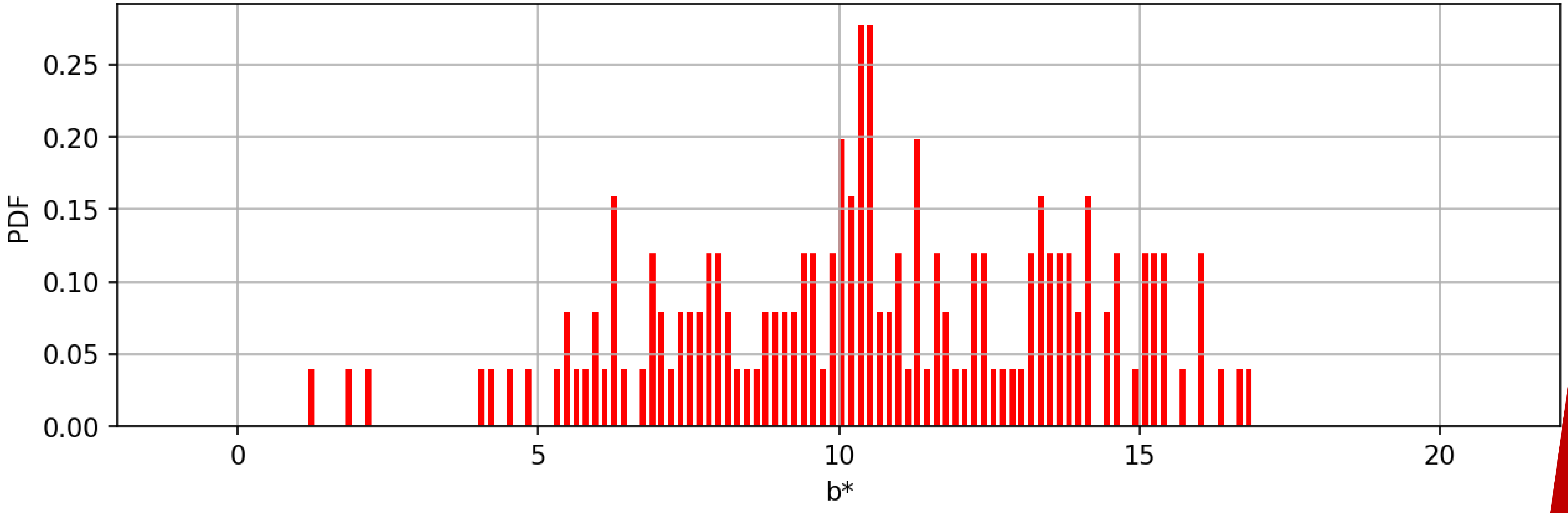
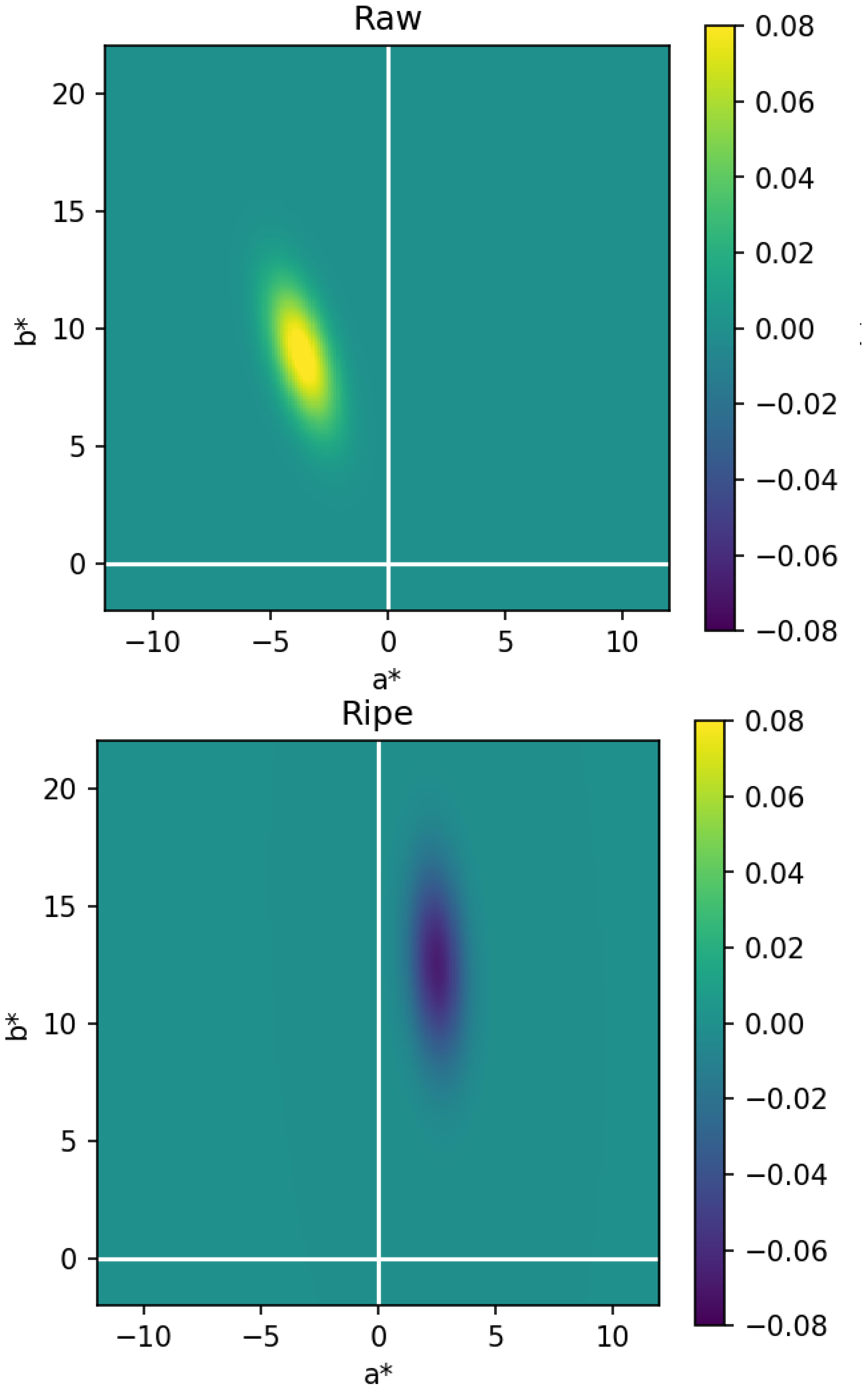


Figure 2. Histogram of the a* and b* color features of my raw and ripe banana fruit dataset

Initializing the algorithm to random parameters, the log likelihood on the first iteration is -1130.74 and it took 23 iterations for it to reach a maximum likelihood of -811.91. Here I have two classes and two features which means a 2x2 Mean matrix and a 2x2x2 Standard Deviation matrix shall be returned on the final iteration. Using these parameters, the gaussian probability was plotted as shown in the plots in Figure 3. I set one class to have a negative probability to accentuate the disparity of two classes. To check, I simply over-layed these probability plots in the scatter plot of my dataset as shown in Figure 4.



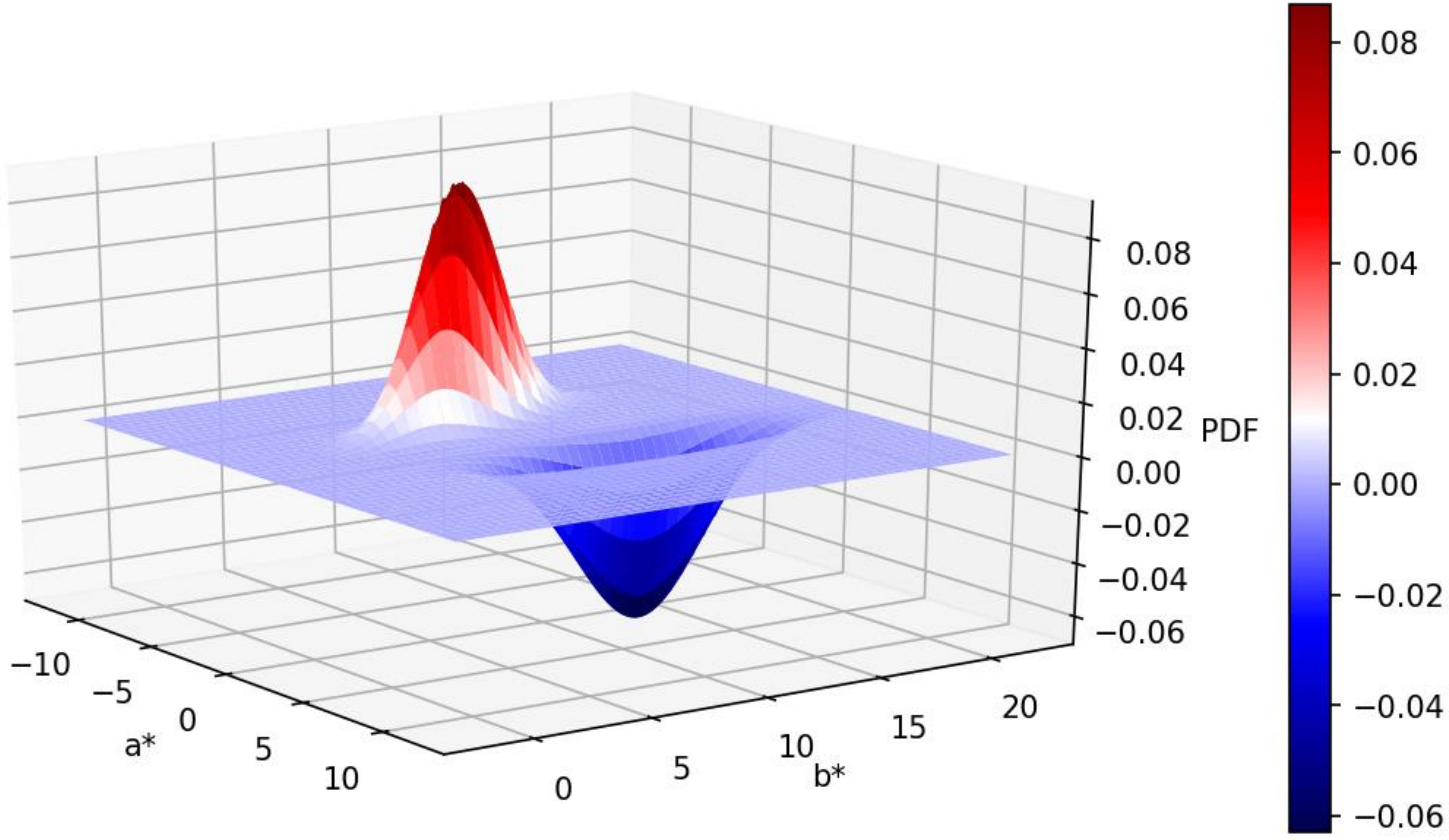


Figure 3. Probability maps of the raw (+) and ripe (-) banana class, and the 3D representation of the convolved probability distribution functions. Points situated at the zero is indiscernible.

In this activity, I'd give myself a 10

For this activity, I utilized the codes shown in the references to facilitate the implementation of the EM Algorithm.

References:

[1] M. Soriano, "Expectation Maximization", 2019.

[Codes]

https://stackoverflow.com/questions/28342968/how-to-plot-a-2d-gaussian-with-different-sigma/28343236

https://towardsdatascience.com/simple-example-of-2d-density-plots-in-python-83b83b934f67

https://zhiyzuo.github.io/EM/

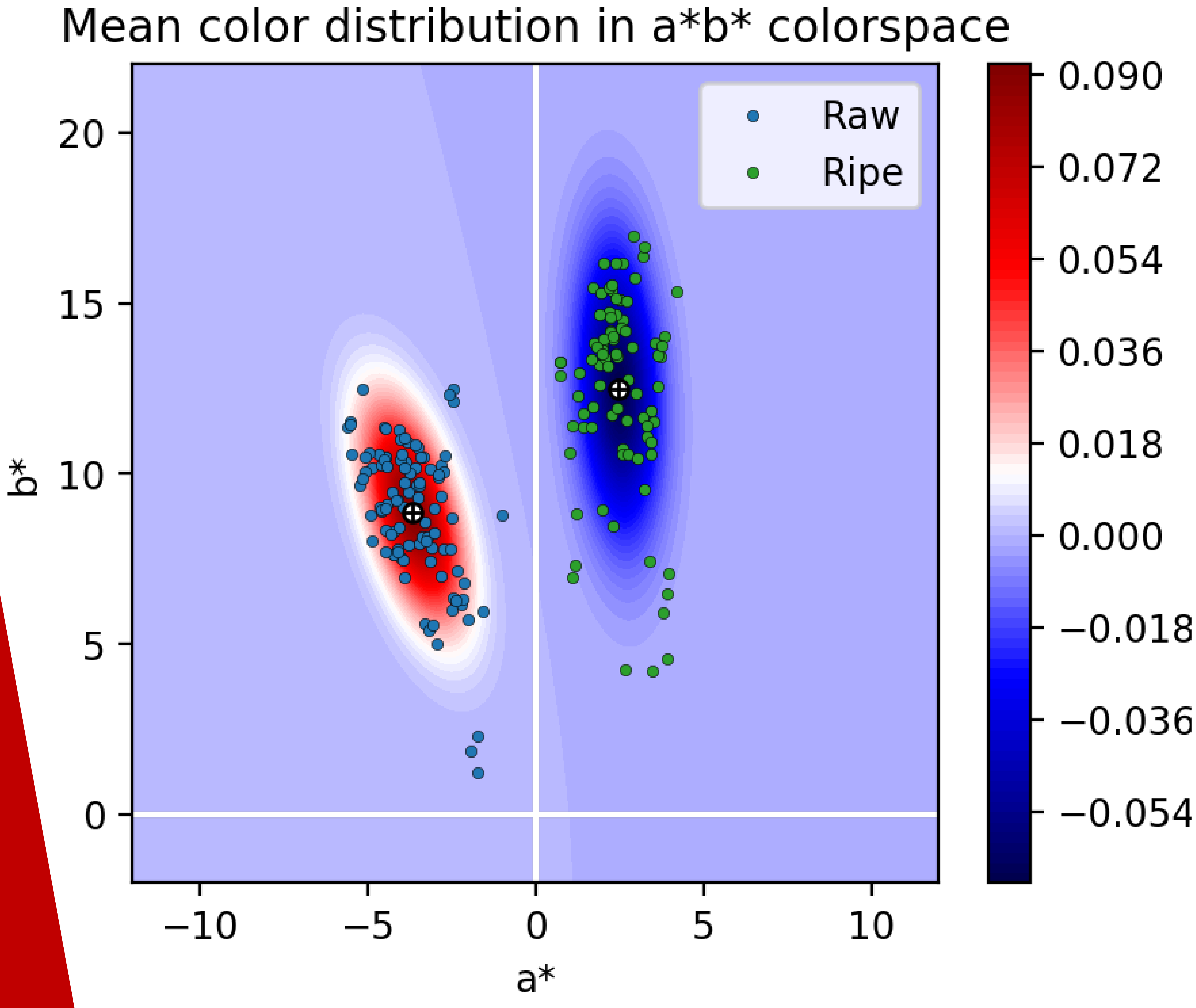


Figure 4. Over-layed plot of the gaussian PDFs with maximum likelihood and the scatter plot of the datapoints.