Activity 15 – Expectation Maximization

Kenneth V. Domingo
2015–03116
App Physics 186, 1st Semester, A.Y. 2019–20
Corresponding author: kvdomingo@up.edu.ph

Results and Discussion

For this activity [1], I used the separated banana, apple, and orange feature data from a previous activity. The fruits form clear clusters in $a^* - b^*$ feature space and is suitable for this activity. Figure 1 shows the clustering of the fruit data.

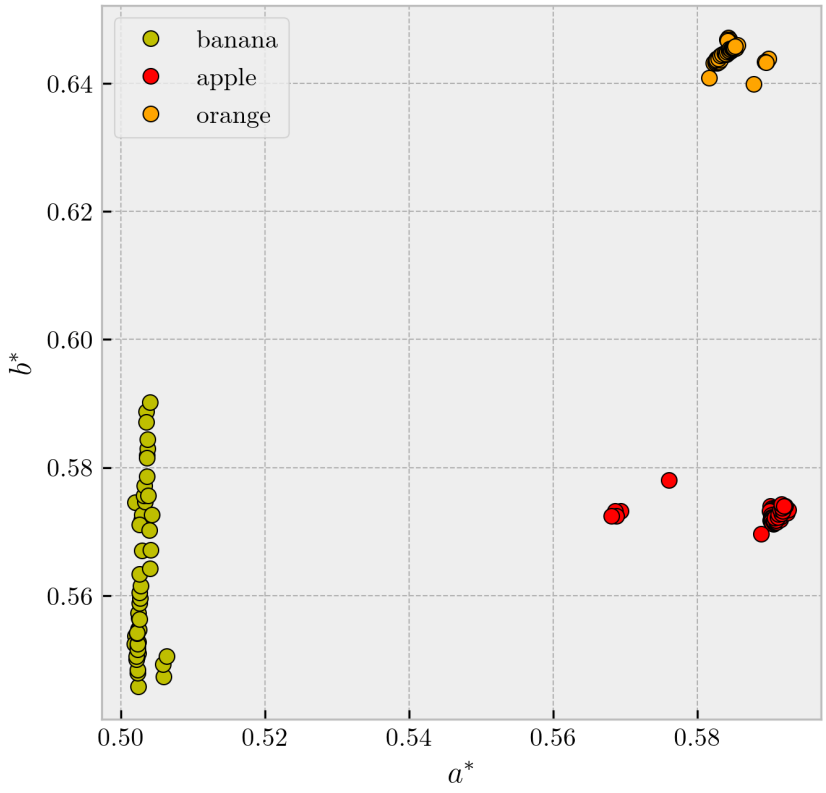

Figure 1: $a^* - b^*$ feature space of fruit data.

Since we are working only with two dimensions (two features), we assume a 2D Gaussian distribution, given by

$$p_l(\mathbf{x}|\boldsymbol{\mu}_l, \boldsymbol{\Sigma}_l) = \frac{1}{2\pi|\boldsymbol{\Sigma}_l|^{1/2}} \exp\left[-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu}_l)^\top \boldsymbol{\Sigma}_l^{-1}(\mathbf{x}-\boldsymbol{\mu}_l)\right] \tag{1}$$

In the interest of computational efficiency, we define an intermediate matrix $\boldsymbol{\omega}$ whose elements are given by

$$\omega_{li} = p_l(\mathbf{x}_i|\boldsymbol{\mu}_l, \boldsymbol{\Sigma}_l) \tag{2}$$

which are used throughout one entire iteration, in order to avoid redundant calculation of exponentials and matrix inversions. We then iterate with the update rules

$$P_l' = \frac{1}{N}\sum_{i=1}^{N} P(l|\mathbf{x}_i, \boldsymbol{\Theta}^g) \tag{3}$$

$$\boldsymbol{\mu}_l' = \frac{\sum_{i=1}^{N} \mathbf{x}_i P(l|\mathbf{x}_i, \boldsymbol{\Theta}^g)}{\sum_{i=1}^{N} P(l|\mathbf{x}_i, \boldsymbol{\Theta}^g)} \tag{4}$$

$$\boldsymbol{\Sigma}_l' = \frac{\sum_{i=1}^{N} P(l|\mathbf{x}_i, \boldsymbol{\Theta}^g)(\mathbf{x}_i - \boldsymbol{\mu}_l')(\mathbf{x}_i - \boldsymbol{\mu}_l')^\top}{\sum_{i=1}^{N} P(l|\mathbf{x}_i \boldsymbol{\Theta}^g)} \tag{5}$$

until the log-likelihood goes above some pre-set value. The log-likelihood is given by

$$L = \ln\left[\sum_i \sum_l P_l' p_l(\mathbf{x}_i|\boldsymbol{\mu}_l, \boldsymbol{\Sigma}_l)\right] \tag{6}$$

The PDF shown in Fig. 2 is obtained at an average of 31 epochs.

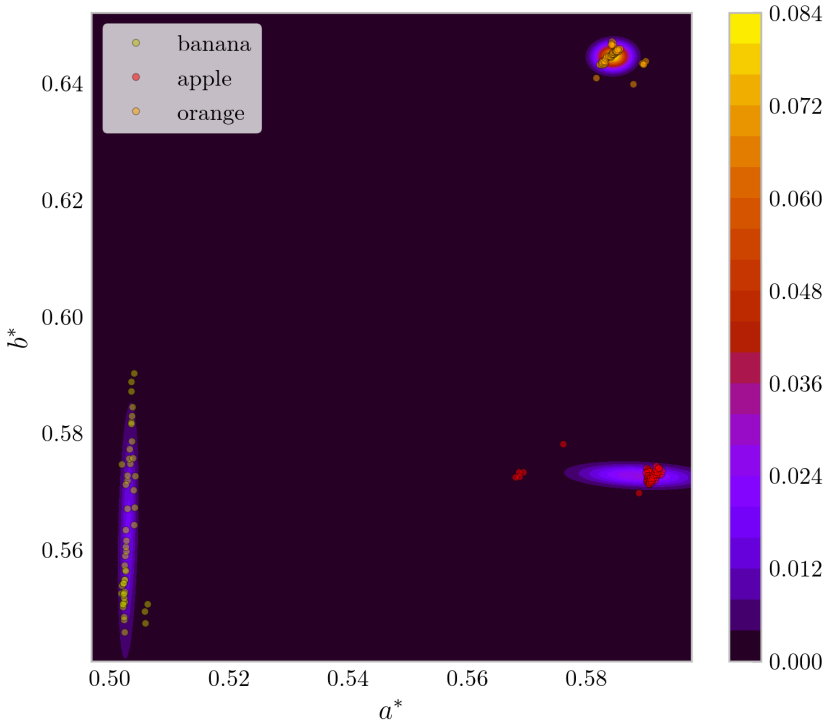

Figure 2: Estimated PDF in the a^*b^* feature space of bananas, apples, and oranges.

Table 1: Self-evaluation.

Technical correctness	5
Quality of presentation	5
Initiative	0
TOTAL	**10**

References

[1] M. N. Soriano, *A15 – Expectation maximization* (2019).