Creating Classification Rules to Distinguish Between Cherry Tree and Pear Tree Leaves

Presented to Dr. Steven Vamosi

Professor of Ecology and Evolutionary Biology

University of Calgary

Prepared by Jana Osea

Undergraduate Student in Statistics

University of Calgary

March 14, 2021

Contents

Summary

As requested by Dr. Steven Vamosi, I have prepared a report on developing a classification system to distinguish between cherry and pear tree leaves. Properly classifying items as a part of a group is becoming increasingly important.

In order to generate the data for the classification systems, a supplied pdf that contains 16 leaves from both cherry and pear trees were printed out on A4 paper. Each leaf was measured according to its width and length using a straight ruler to the closest millimeter.

Once the data was recorded, 2 classification systems were implemented and analyzed. The first classification system assumed that the width and length of each species of leaves followed a bivariate normal distribution in which both species had a pooled covariance matrix. The second classification system assumed that the width and length of each species of leaves followed a bivariate normal distribution with their own respective covariance matrix.

The findings from the results show that the decision boundary of the first classification system followed a linear equation while the second classification system followed a quadratice linear equation. It was also shown that the number of parameters that were estimated increased from the first to the second classifcation rule. Both rules appeared to have the same misclassifications. However, the second classification rule had on average lower lambda values compared to the first.

Overall, I suggest using the second classification rule as it better fits the data.

Introduction

In many fields of study, classifying items or individuals as belonging to one of two or more population or groups is an integral part of analysis. In many cases, it is often the goal of a research or study to classify each sample item correctly. In fields like finance and medical research, correctly classifying items can imply stopping transaction fraud or discovering deadly tumors. Hence, it is important to understand how classification procedures work.

Background

As the importance of classification is becoming more evident, Professor Steven Vamosi, a Doctor in Ecology and Evolutionary Biology has entrusted an undergraduate student in Statistics at the University of Calgary, Jana Osea, with the task to develop a classification rule to distinguish between cherry and pear tree leaves. This allows Dr. Vamosi a simple method to classify between the two species and for us demonstrate our knowledge of classification.

Goal

Using width and length measurements taken from cherry and pear tree leaves, our *goal* is to create a classification method to distinguish between the two species.

Data Generation Process

Data Source

- Cherry leaves: Figure 7.2 of the pdf file printed on a standard A4 paper
- Pear leaves: Figure 7.3 of the pdf file printed on a standard A4 paper

Data Input

In an Microsoft Excel Sheet (2020), I prepared 3 empty columns with the following headers: species, width, and length.

For each leaf a new row with 3 columns is recorded in the excel sheet that contains the species, width, and length measured according to the procedure outlined below. After recording each value, I saved the data as a csv file named "data.csv". In addition, the full raw data can be found in A1 in the appendix.

- species: If the leaf is part of figure 7.2, then species contains string input "cherry." If the leaf is part of figure 7.3, then species contains string input "pear."
- width: Measured the widest part of each leaf using a straight ruler to the closest millimeter
- length: Measured from the bottom tip to the top tip using a straight ruler of each leaf to the closest millimeter

Methods

Overview of Methods

After inputting the entire data set, I imported the csv file into my program. I made 2 classifications: (1) with equal variance assumption and (2) with no equal variance assumption. Densities and lambda values were calculated for each leaf and visualizations of classifcations were made. 3 new leaf measurements were provided and classified according to the first classification. In addition, misclassifications of each method were recorded.

Software and Packages

I used R version 4.0.3 (2020-10-10) (R Core Team (2020)) to perform all my classification programming. I also used the following R package to help me visualize and aid my density calculations

- ggplot2 (H. Wickham (2016))
- gridarrange (Baptiste Auguie (2017))
- mtvnorm (Alan Genz, et. al (2020))

First Classification

Assumptions

The first classification rule assumes the following:

1. For each species k = cherry or pear, the distribution of the width and length measurements follow a bivariate normal distribution as follows

$$\begin{pmatrix} X_k \\ Y_k \end{pmatrix} \sim N_2\left(\mu_k, \Sigma\right)$$

   where

$$\begin{aligned} X_k &= \text{ width (mm) of the } k \text{ species} \\ Y_k &= \text{ length (mm) of the } k \text{ species} \\ \mu_k &= \begin{pmatrix} \mu_{kx} \\ \mu_{ky} \end{pmatrix} \text{ of the } k \text{ species} \\ \Sigma &= \begin{pmatrix} \sigma_x^2 & \sigma_{xy} \\ \sigma_{xy} & \sigma_y^2 \end{pmatrix} \end{aligned}$$

   Hence, the density of a leaf given the x width and y length according to the k = cherry or pear species is given by

$$f_k(x, y \mid \mu_k, \Sigma) = \frac{1}{2\pi\sqrt{|\Sigma|}} \exp\left[\frac{-1}{2} \begin{pmatrix} x - \mu_{kx} \\ y - \mu_{ky} \end{pmatrix}^T \Sigma^{-1} \begin{pmatrix} x - \mu_{kx} \\ y - \mu_{ky} \end{pmatrix}\right]$$

   where

$$|\Sigma| = \text{ determinant of the covariance matrix.}$$

2. The covariance matrix Σ of the k = cherry or pear species is the same with possible differences in the mean vectors μ_k.

Parameter Estimation

Given the data collected, for $k =$ cherry or pear, we estimate the unknown parameters μ_k and Σ as $\hat{\mu}_k$ and $\hat{\Sigma}$ by using the method of moments as follows.

$$\hat{\mu}_k = \begin{pmatrix} \frac{1}{n_k}\sum_{i=1}^{n_k} x_{ki} \\ \frac{1}{n_k}\sum_{i=1}^{n_k} y_{ki} \end{pmatrix} \quad \text{and} \quad \hat{\Sigma} = \frac{1}{2}\left(\hat{\Sigma}_{\text{cherry}} + \hat{\Sigma}_{\text{pear}}\right)$$

where

$$\begin{aligned} x_{ki} &= \text{ width (mm) of the } i\text{-th leaf for the } k\text{-th species} \\ y_{ki} &= \text{ length (mm) of the } i\text{-th leaf for the } k\text{-th species} \\ n_k &= \text{ number of total leaves gathered for the } k\text{-th species} \\ \hat{\Sigma}_k &= \begin{pmatrix} \frac{1}{n_k}\sum_{i=1}^{n_k} x_{ki}^2 - \left(\frac{1}{n_k}\sum_{i=1}^{n_k} x_{ki}\right)^2 & \frac{1}{n_k}\sum_{i=1}^{n_k} x_{ki}y_{ki} - \frac{1}{n_k}\sum_{i=1}^{n_k} x_{ki}\frac{1}{n_k}\sum_{i=1}^{n_k} y_{ki} \\ \frac{1}{n_k}\sum_{i=1}^{n_k} x_{ki}y_{ki} - \frac{1}{n_k}\sum_{i=1}^{n_k} x_{ki}\frac{1}{n_k}\sum_{i=1}^{n_k} y_{ki} & \frac{1}{n_k}\sum_{i=1}^{n_k} y_{ki}^2 - \left(\frac{1}{n_k}\sum_{i=1}^{n_k} y_{ki}\right)^2 \end{pmatrix} \end{aligned}$$

Using the formulas above and the sample data, the estimated mean and covariance matrix are shown below.

$$\hat{\mu}_{\text{cherry}} = \begin{pmatrix} 36.44 \\ 79.56 \end{pmatrix}$$

$$\hat{\mu}_{\text{pear}} = \begin{pmatrix} 41.19 \\ 70 \end{pmatrix}$$

$$\hat{\Sigma} = \begin{pmatrix} 22.51 & 31.28 \\ 31.28 & 120.5 \end{pmatrix}$$

Hence using the estimated parameters, the density formula for a given x width and y length for the $k =$ cherry or pear species is shown below.

$$f_{\text{cherry}}(x, y \mid \hat{\mu}_{\text{cherry}}, \hat{\Sigma}) = \frac{1}{2\pi\sqrt{\left|\begin{pmatrix} 22.51 & 31.28 \\ 31.28 & 120.5 \end{pmatrix}\right|}} \exp\left[\frac{-1}{2}\begin{pmatrix} x - 36.44 \\ y - 79.56 \end{pmatrix}^T \begin{pmatrix} 22.51 & 31.28 \\ 31.28 & 120.5 \end{pmatrix}^{-1} \begin{pmatrix} x - 36.44 \\ y - 79.56 \end{pmatrix}\right]$$

$$f_{\text{pear}}(x, y \mid \hat{\mu}_{\text{pear}}, \hat{\Sigma}) = \frac{1}{2\pi\sqrt{\left|\begin{pmatrix} 22.51 & 31.28 \\ 31.28 & 120.5 \end{pmatrix}\right|}} \exp\left[\frac{-1}{2}\begin{pmatrix} x - 41.19 \\ y - 70 \end{pmatrix}^T \begin{pmatrix} 22.51 & 31.28 \\ 31.28 & 120.5 \end{pmatrix}^{-1} \begin{pmatrix} x - 41.19 \\ y - 70 \end{pmatrix}\right]$$

Classification Rule

Let n be the total number of observations in the training data, i be any integer from 0 to n, and (x_i, y_i) be the i-th observation where x_i is the width measurement and y_i is the length measurement. The respective λ values for each observation (x_i, y_i) is as follows.

$$\lambda_i = \frac{f_{\text{cherry}}(x_i, y_i|\hat{\mu}_{\text{cherry}}, \hat{\Sigma})}{f_{\text{pear}}(x_i, y_i|\hat{\mu}_{\text{pear}}, \hat{\Sigma})}$$

$$= \frac{\frac{1}{2\pi\sqrt{\left|\begin{pmatrix} 22.51 & 31.28 \\ 31.28 & 120.5 \end{pmatrix}\right|}} \exp\left[\frac{-1}{2}\begin{pmatrix} x - 36.44 \\ y - 79.56 \end{pmatrix}^T \begin{pmatrix} 22.51 & 31.28 \\ 31.28 & 120.5 \end{pmatrix}^{-1} \begin{pmatrix} x - 36.44 \\ y - 79.56 \end{pmatrix}\right]}{\frac{1}{2\pi\sqrt{\left|\begin{pmatrix} 22.51 & 31.28 \\ 31.28 & 120.5 \end{pmatrix}\right|}} \exp\left[\frac{-1}{2}\begin{pmatrix} x - 41.19 \\ y - 70 \end{pmatrix}^T \begin{pmatrix} 22.51 & 31.28 \\ 31.28 & 120.5 \end{pmatrix}^{-1} \begin{pmatrix} x - 41.19 \\ y - 70 \end{pmatrix}\right]}$$

The following classification rule described below is used to determine whether observation (x_i, y_i) belongs to a certain species. The lambda values and classifications of all the observations is recorded in A2 in the appendix.

- if $\lambda_i > 1$, then observation (x_i, y_i) is a cherry leaf
- if $\lambda_i < 1$, then observation (x_i, y_i) is a pear leaf
- if $\lambda_i = 1$, then observation (x_i, y_i) is undetermined

Classification Errors

There is a total of 4 misclassifications. This is evident in table 1 as well as the rey circles and orange triangles in figure 2.

Table 1. Observation Points Where Misclassification Occurs

	species	width (mm)	length (mm)	classification
5	cherry	37	67	pear
7	cherry	47	88	pear
25	pear	40	90	cherry
27	pear	32	63	cherry

Figure 1. Scatter Plot of Observation Data

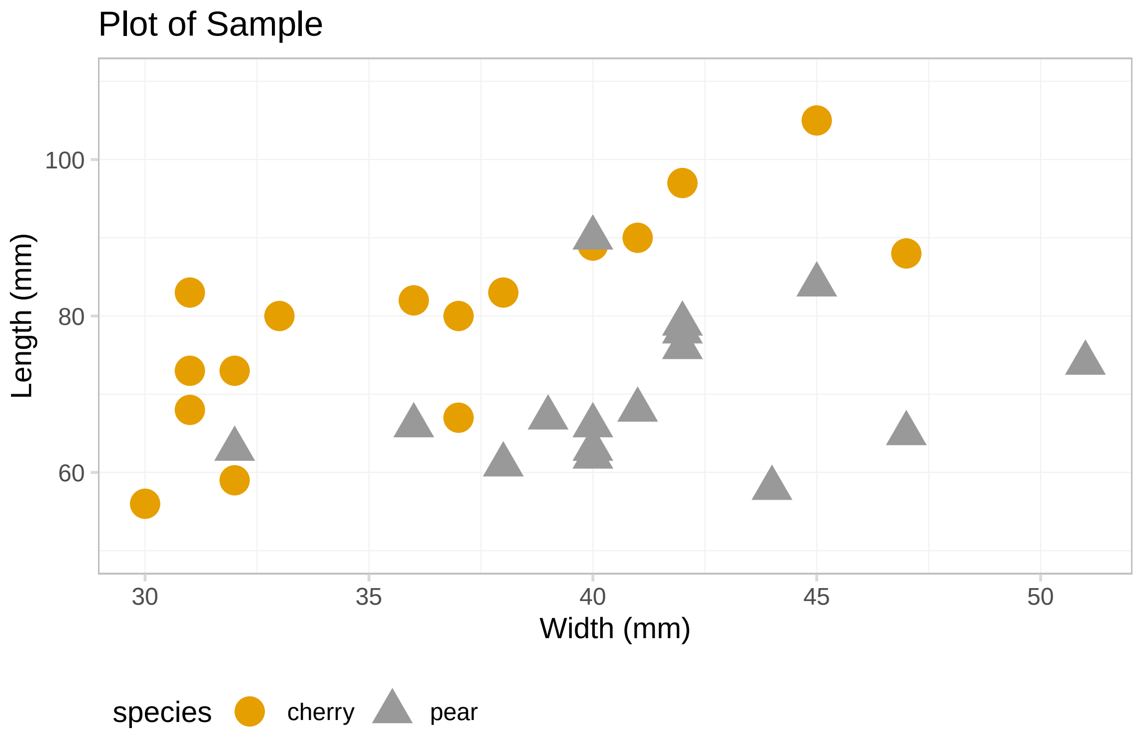
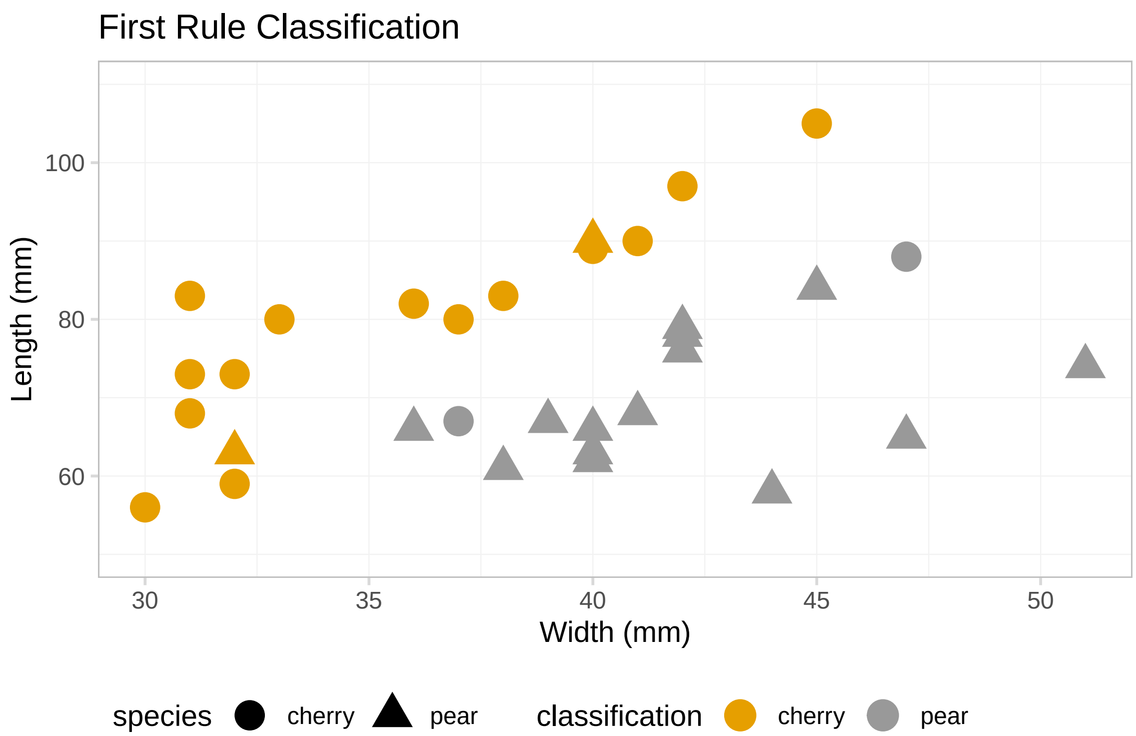


Figure 2. Scatter Plot of First Classification Rule



New Classification

Using the classification rules outlined above, we are tasked to classify new leaves with the following meassurements

id	width (mm)	length (mm)
u	32	82
v	38	52
w	52	76

We calculate λ using the formula stated above and get the following resuts

id	lamda	classification
u	139.61	cherry
v	0.01	pear
w	0.71	pear

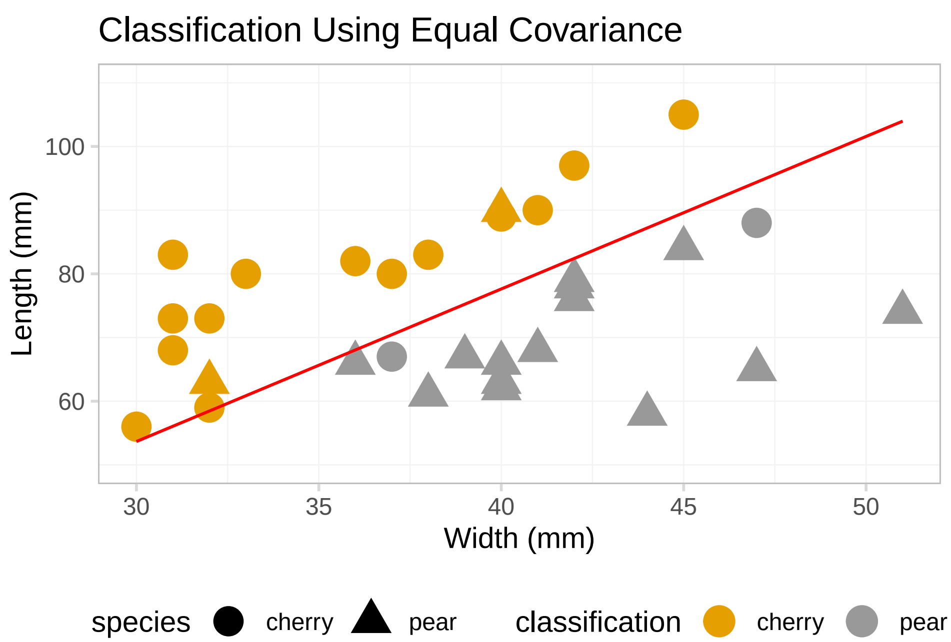
Decision Boundary

Using the derivation of the decision boundary found in A4 in the appendix, the decision boundary for the given in the formula below.

$$y = 2.4x - 18.18$$

The equation above is clearly a linear polynomial equation with degree 1. The decision boundary is plotted on the observation space as shown below. All the points above the decision boundary are classified as the cherry species and below are the pear species.

Figure 3. Scatter Plot with Decision Boundary



Furthermore, I predicted all the possible points in the observation space to the closest 0.1 mm using the first classification rules and got the following results in the plot below. From figure 4, it is clear that the decision boundary is a linear equation that divides the observation space into 2 distinct regions.

Figure 4. Grid Plot of a Predicted Points in the Observation Space

Second Classification

Assumptions

The assumptions are similar to the first classification. However, the only difference is that the covariance matrices are no longer equal. Instead, for each species $k =$ cherry or pear, the distribution of the width and length measurements follow a bivariate normal as follows

$$\begin{pmatrix} X_k \\ Y_k \end{pmatrix} \sim N_2\left(\mu_k, \Sigma_k\right)$$

where

$$\Sigma_k = \begin{pmatrix} \sigma_{kx}^2 & \sigma_{kxy} \\ \sigma_{kxy} & \sigma_{ky}^2 \end{pmatrix} \text{ covariance matrix of the } k \text{ species}$$

Hence, the density of a leaf given the x width and y length according to the $k =$ cherry or pear species is given by

$$f_k(x, y \mid \mu_k, \Sigma) = \frac{1}{2\pi\sqrt{|\Sigma_k|}} \exp\left[\frac{-1}{2}\begin{pmatrix} x-\mu_{kx} \\ y-\mu_{ky} \end{pmatrix}^T \Sigma_k^{-1} \begin{pmatrix} x-\mu_{kx} \\ y-\mu_{ky} \end{pmatrix}\right]$$

where

$$|\Sigma_k| = \text{ determinant of the } k \text{ species covariance matrix.}$$

Parameter Estimation

For $k =$ cherry or pear, $\hat{\mu}_k$ is the same as derived in the first classification. $\hat{\Sigma}_k$ is no longer the pooled covariance matrix. For each $k =$ cherry or pear the estimation of $\hat{\Sigma}_k$ is the method of moments covariance matrix. The derivation of this can be found in the first classification parameter estimation.

$$\hat{\Sigma}_{\text{cherry}} = \begin{pmatrix} 27.12 & 52.07 \\ 52.07 & 162.87 \end{pmatrix}$$

$$\hat{\Sigma}_{\text{pear}} = \begin{pmatrix} 17.9 & 10.5 \\ 10.5 & 78.12 \end{pmatrix}$$

Hence using the estimated parameters, the density formula for a given x width and y length for the $k =$ cherry or pear species is shown below.

$$f_{\text{cherry}}(x, y \mid \hat{\mu}_{\text{cherry}}, \hat{\Sigma}_{\text{cherry}}) = \frac{1}{2\pi\sqrt{\left|\begin{pmatrix} 27.12 & 52.07 \\ 52.07 & 162.87 \end{pmatrix}\right|}} \exp\left[\frac{-1}{2}\begin{pmatrix} x - 36.44 \\ y - 79.56 \end{pmatrix}^T \begin{pmatrix} 27.12 & 52.07 \\ 52.07 & 162.87 \end{pmatrix}^{-1} \begin{pmatrix} x - 36.44 \\ y - 79.56 \end{pmatrix}\right]$$

$$f_{\text{pear}}(x, y \mid \hat{\mu}_{\text{pear}}, \hat{\Sigma}_{\text{pear}}) = \frac{1}{2\pi\sqrt{\left|\begin{pmatrix} 17.9 & 10.5 \\ 10.5 & 78.12 \end{pmatrix}\right|}} \exp\left[\frac{-1}{2}\begin{pmatrix} x - 41.19 \\ y - 70 \end{pmatrix}^T \begin{pmatrix} 17.9 & 10.5 \\ 10.5 & 78.12 \end{pmatrix}^{-1} \begin{pmatrix} x - 41.19 \\ y - 70 \end{pmatrix}\right]$$

Classification Rule

Similar to the first classification rule, we use the same cutoff λ value and test cases. However, each corresponding λ_i is now calculated as follows.

$$\lambda_i = \frac{f_{\text{cherry}}(x_i, y_i | \hat{\mu}_{\text{cherry}}, \hat{\Sigma}_{\text{cherry}})}{f_{\text{pear}}(x_i, y_i | \hat{\mu}_{\text{pear}}, \hat{\Sigma}_{\text{pear}})}$$

$$= \frac{\frac{1}{2\pi\sqrt{\left|\begin{pmatrix} 27.12 & 52.07 \\ 52.07 & 162.87 \end{pmatrix}\right|}} \exp\left[\frac{-1}{2}\begin{pmatrix} x - 36.44 \\ y - 79.56 \end{pmatrix}^T \begin{pmatrix} 27.12 & 52.07 \\ 52.07 & 162.87 \end{pmatrix}^{-1} \begin{pmatrix} x - 36.44 \\ y - 79.56 \end{pmatrix}\right]}{\frac{1}{2\pi\sqrt{\left|\begin{pmatrix} 17.9 & 10.5 \\ 10.5 & 78.12 \end{pmatrix}\right|}} \exp\left[\frac{-1}{2}\begin{pmatrix} x - 41.19 \\ y - 70 \end{pmatrix}^T \begin{pmatrix} 17.9 & 10.5 \\ 10.5 & 78.12 \end{pmatrix}^{-1} \begin{pmatrix} x - 41.19 \\ y - 70 \end{pmatrix}\right]}$$

Classification Errors

Similar to the previous classification rule, there are also 4 missclassifications in the data. These are the same exact misclassifications of the previous rule. See figure 1 and 2.

New Classification

Using the classification rules outlined above, we classify the same new values as found in the first classification rule. The results of the new classification compared to the first classification are shown below.

			First Rule		Second Rule	
id	width (mm)	length (mm)	lamda	classification	lambda	classification
u	32	82	139.61	cherry	20.10	cherry
v	38	52	0.01	pear	0.00	pear
w	52	76	0.71	pear	0.41	pear

The results show that the classifications of the same u, v, w points are the same. However, notice that the lambda values for the second classification are not as high expecially for u.

Decision Boundary

Using the derivation in A5 in the appendix, the decision boundary for the given classification is the function below

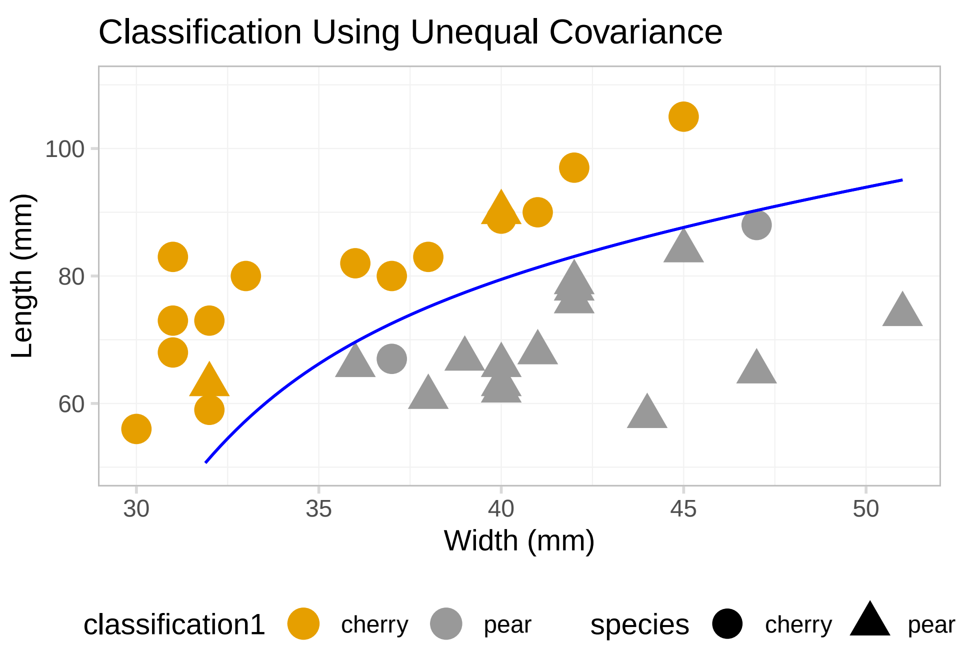
$$y = \frac{1}{2A}\left(-B - \sqrt{B^2 - 4AC}\right)$$

where

$$\begin{aligned}
A &= 0.002 \\
B &= 0.96849 \\
C &= 0.03481x^2 + 1.75362x - 73.25789
\end{aligned}$$

The equation above is clearly a quadratic polynomial equation with degree 2. The decision boundary is plotted on the observation space as shown below. All the points above the decision boundary are classified as the cherry species and below are the pear species.

Figure 5. Scatter Plot with Decision Boundary

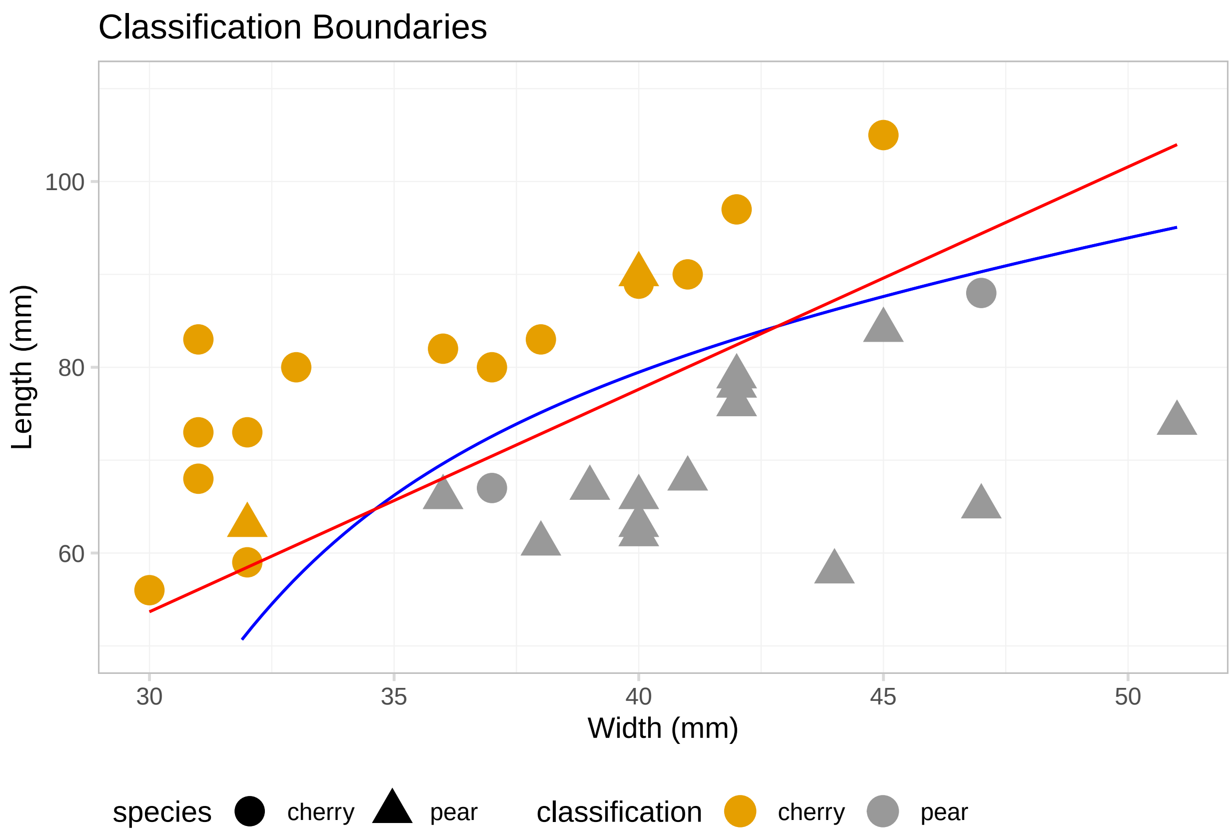


Furthermore, similar to the decision boundary in the first classification, I predicted every possible sample in the observation space to the closest 0.1 mm and plotted the predictions as shown in figure 6.

Figure 6. Grid Plot of Predicted Points in the Observation Space

In addition, when we compare the two different classifications, we observe that the decision boundary between the two are different. The first classification has a decision boundary that is linear and the second classification has a decision boundary that is quadratic. We can observe the difference in the figure 7 below.

Figure 7. Decision Boundary Comparison



Notice that getting rid of the assumption of equal covariance between the two species allows for the decision boundary to cater to the shape of the data. For example, if we look at the length of the cherry species in figure 7, we notice that the quadratic decision boundary is lowercompared to the linear quadratic boundary in order to have more allowance to correctly classify lower lengths as cherry leaves.

Conclusion and Recommendations

Conclusion

In this report, we created a data generation protocol of measuring the width and length of 2 different species of leaves (cherry and pear). Then we developed 2 classification rules and performed them on the data. The findings in the results show that the difference between the the two classifications is the decision boundary equation. This means that the shape of the decision boundary is determined by whether or not we assume equal covariance between the two species (ie. same covariance of leaf width and length between pear and cherry). Assuming equal covariance results in a linear decision boundary and assuming non-equal covariance results in a quadratic decision boundary. The number of parameters we estimate also changes depending on the equal covariance assumption. If we assume equal covaraince, the number of parameters we estimate is 8 (μ_k for k= cherry or pear and Σ) and if we do not assume equal covariance, the number of parameters we estimate is 12 (μ_k, and Σ_k for k= cherry or pear).

Recommendations

I recommend using the second classification rule when determining which species a leaf belongs to. Even though there are more parameters to estimate compared to the first classification rule, the decision boundary is more catered to the shape of the data and so will yield more accurate results.

References

- R Core Team (2020). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org./
- H. Wickham (2016). ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York, 2016.
- Baptiste Auguie (2017). gridExtra: Miscellaneous Functions for "Grid" Graphics. R package version 2.3. https://CRAN.R-project.org/package=gridExtra
- Alan Genz, Frank Bretz, Tetsuhisa Miwa, Xuefei Mi, Friedrich Leisch, Fabian Scheipl, Torsten Hothorn (2020). mvtnorm: Multivariate Normal and t Distributions. R package version 1.1-1. URL http://CRAN.R-project.org/package=mvtnorm
- Alan Genz, Frank Bretz (2009), Computation of Multivariate Normal and t Probabilities. Lecture Notes in Statistics, Vol. 195., Springer-Verlag, Heidelberg. ISBN 978-3-642-01688-2

Appendix

A1. Data Table with Density Calculations

```
##    species width length
## 1   cherry    30     56
## 2   cherry    32     59
## 3   cherry    31     68
## 4   cherry    41     90
## 5   cherry    37     67
## 6   cherry    45    105
## 7   cherry    47     88
## 8   cherry    31     73
## 9   cherry    32     73
## 10  cherry    33     80
## 11  cherry    36     82
## 12  cherry    40     89
## 13  cherry    42     97
## 14  cherry    37     80
## 15  cherry    31     83
## 16  cherry    38     83
## 17    pear    44     58
## 18    pear    41     68
## 19    pear    38     61
## 20    pear    40     62
## 21    pear    42     78
## 22    pear    40     66
## 23    pear    40     63
## 24    pear    47     65
## 25    pear    40     90
## 26    pear    42     76
## 27    pear    32     63
## 28    pear    45     84
## 29    pear    42     79
## 30    pear    36     66
## 31    pear    39     67
## 32    pear    51     74
```

A2. Densities, Lambda Calculations and First Rule Classification

```
##    species f.cherry.1 f.pear.1     l.1    c.1
## 1   cherry    3.8e-04  2.3e-04 1.6e+00 cherry
## 2   cherry    6.4e-04  5.7e-04 1.1e+00 cherry
## 3   cherry    1.8e-03  1.5e-04 1.2e+01 cherry
## 4   cherry    2.2e-03  2.7e-04 8.1e+00 cherry
## 5   cherry    1.2e-03  2.5e-03 4.9e-01   pear
## 6   cherry    2.3e-04  9.0e-06 2.5e+01 cherry
## 7   cherry    2.5e-04  9.5e-04 2.6e-01   pear
## 8   cherry    2.0e-03  5.6e-05 3.5e+01 cherry
## 9   cherry    2.5e-03  1.2e-04 2.1e+01 cherry
## 10  cherry    2.5e-03  4.4e-05 5.6e+01 cherry
## 11  cherry    3.6e-03  1.9e-04 1.9e+01 cherry
## 12  cherry    2.5e-03  2.3e-04 1.1e+01 cherry
## 13  cherry    1.0e-03  4.9e-05 2.1e+01 cherry
## 14  cherry    3.8e-03  5.1e-04 7.4e+00 cherry
## 15  cherry    9.0e-04  3.2e-06 2.8e+02 cherry
## 16  cherry    3.6e-03  4.2e-04 8.4e+00 cherry
## 17    pear    1.4e-06  6.2e-04 2.2e-03   pear
## 18    pear    3.0e-04  3.7e-03 8.0e-02   pear
## 19    pear    2.2e-04  2.7e-03 8.3e-02   pear
## 20    pear    1.1e-04  2.9e-03 3.8e-02   pear
## 21    pear    1.1e-03  2.8e-03 4.0e-01   pear
## 22    pear    3.1e-04  3.6e-03 8.7e-02   pear
## 23    pear    1.4e-04  3.1e-03 4.6e-02   pear
## 24    pear    1.2e-06  5.9e-04 2.1e-03   pear
## 25    pear    2.4e-03  1.8e-04 1.3e+01 cherry
## 26    pear    8.4e-04  3.2e-03 2.6e-01   pear
## 27    pear    1.2e-03  4.7e-04 2.6e+00 cherry
## 28    pear    5.2e-04  1.7e-03 3.1e-01   pear
## 29    pear    1.2e-03  2.5e-03 4.9e-01   pear
## 30    pear    1.3e-03  2.0e-03 6.5e-01   pear
## 31    pear    6.1e-04  3.4e-03 1.8e-01   pear
## 32    pear    4.6e-07  2.5e-04 1.9e-03   pear
```

where

`f.k.1` $= f_k(x, y|\hat{\mu}_k, \hat{\Sigma})$ of the k species

`l.1` $= \lambda = \dfrac{f_{\text{cherry}}(x, y|\hat{\mu}_{\text{cherry}}, \hat{\Sigma})}{f_{\text{pear}}(x, y|\hat{\mu}_{\text{pear}}, \hat{\Sigma})}$ calculation

`c.1` = "cherry" or "pear" depending on the first classification rule

A3. Densities, Lambda Calculations and Second Rule Classification

```
##    species f.cherry.2 f.pear.2     l.2    c.2
## 1   cherry    6.6e-04  9.2e-05 7.2e+00 cherry
## 2   cherry    8.5e-04  3.4e-04 2.5e+00 cherry
## 3   cherry    2.2e-03  2.2e-04 1.0e+01 cherry
## 4   cherry    2.6e-03  2.7e-04 9.6e+00 cherry
## 5   cherry    8.7e-04  2.7e-03 3.2e-01   pear
## 6   cherry    5.2e-04  1.7e-06 3.1e+02 cherry
## 7   cherry    1.6e-04  3.9e-04 4.1e-01   pear
## 8   cherry    2.0e-03  1.4e-04 1.4e+01 cherry
## 9   cherry    2.6e-03  2.6e-04 1.0e+01 cherry
## 10  cherry    2.1e-03  1.5e-04 1.4e+01 cherry
## 11  cherry    3.5e-03  4.3e-04 8.1e+00 cherry
## 12  cherry    2.9e-03  2.9e-04 1.0e+01 cherry
## 13  cherry    1.5e-03  3.3e-05 4.6e+01 cherry
## 14  cherry    3.8e-03  9.2e-04 4.1e+00 cherry
## 15  cherry    4.8e-04  2.0e-05 2.4e+01 cherry
## 16  cherry    3.7e-03  7.2e-04 5.1e+00 cherry
## 17    pear    4.3e-08  9.7e-04 4.4e-05   pear
## 18    pear    9.9e-05  4.3e-03 2.3e-02   pear
## 19    pear    9.2e-05  2.3e-03 3.9e-02   pear
## 20    pear    2.7e-05  2.9e-03 9.1e-03   pear
## 21    pear    6.6e-04  2.9e-03 2.3e-01   pear
## 22    pear    1.1e-04  4.0e-03 2.8e-02   pear
## 23    pear    3.9e-05  3.2e-03 1.2e-02   pear
## 24    pear    3.2e-08  1.1e-03 3.0e-05   pear
## 25    pear    2.8e-03  2.2e-04 1.3e+01 cherry
## 26    pear    4.3e-04  3.5e-03 1.2e-01   pear
## 27    pear    1.6e-03  4.1e-04 3.9e+00 cherry
## 28    pear    3.2e-04  1.1e-03 2.8e-01   pear
## 29    pear    8.0e-04  2.6e-03 3.0e-01   pear
## 30    pear    1.1e-03  2.1e-03 5.1e-01   pear
## 31    pear    3.0e-04  3.8e-03 7.9e-02   pear
## 32    pear    1.0e-08  2.9e-04 3.5e-05   pear
```

where

`f.k.2` $= f_k(x, y|\hat{\mu}_k, \hat{\Sigma}_k)$ of the k species

`l.2` $= \lambda = \dfrac{f_{\text{cherry}}(x, y|\hat{\mu}_{\text{cherry}}, \hat{\Sigma}_{\text{cherry}})}{f_{\text{pear}}(x, y|\hat{\mu}_{\text{pear}}, \hat{\Sigma}_{\text{pear}})}$ calculation

`c.2` = "cherry" or "pear" depending on the second classification rule

A4. Calculating Decision Boundary for First Classification

The decision boundary is the equation $y = f(x)$ such that $\lambda = 1$. In order to have cleaner derivation, let $a = \begin{pmatrix} x \\ y \end{pmatrix}$. We calculate the decision boundary by the following

$$
\begin{aligned}
\lambda &= \frac{f_{\text{cherry}}(x_i, y_i | \hat{\mu}_{\text{cherry}}, \hat{\Sigma})}{f_{\text{pear}}(x_i, y_i | \hat{\mu}_{\text{pear}}, \hat{\Sigma})} \\
&= \frac{\frac{1}{2\pi\sqrt{|\hat{\Sigma}|}} \exp\left\{-\frac{1}{2}(a - \hat{\mu}_{cherry})^T \hat{\Sigma}^{-1}(a - \hat{\mu}_{cherry})\right\}}{\frac{1}{2\pi\sqrt{|\hat{\Sigma}|}} \exp\left\{-\frac{1}{2}(a - \hat{\mu}_{pear})^T \hat{\Sigma}^{-1}(a - \hat{\mu}_{pear})\right\}} \\
&= \exp\left\{-\frac{1}{2}\left((a - \hat{\mu}_{cherry})^T \hat{\Sigma}^{-1}(a - \hat{\mu}_{cherry}) - (a - \hat{\mu}_{pear})^T \hat{\Sigma}^{-1}(a - \hat{\mu}_{pear})\right)\right\}
\end{aligned}
$$

Since, the cutoff value is at $\lambda = 1$, we set the λ equal to 1 and simplify.

$$
\begin{aligned}
1 &= \exp\left\{-\frac{1}{2}\left((a - \hat{\mu}_{cherry})^T \hat{\Sigma}^{-1}(a - \hat{\mu}_{cherry}) - (a - \hat{\mu}_{pear})^T \hat{\Sigma}^{-1}(a - \hat{\mu}_{pear})\right)\right\} \\
\implies \log(1) &= log\left(\exp\left\{-\frac{1}{2}\left((a - \hat{\mu}_{cherry})^T \hat{\Sigma}^{-1}(a - \hat{\mu}_{cherry}) - (a - \hat{\mu}_{pear})^T \hat{\Sigma}^{-1}(a - \hat{\mu}_{pear})\right)\right\}\right) \\
\implies 0 &= -\frac{1}{2}\left((a - \hat{\mu}_{cherry})^T \hat{\Sigma}^{-1}(a - \hat{\mu}_{cherry}) - (a - \hat{\mu}_{pear})^T \hat{\Sigma}^{-1}(a - \hat{\mu}_{pear})\right)
\end{aligned}
$$

Since $\log(1) = 0$, we get the following equality

$$
\begin{aligned}
&(a - \hat{\mu}_{cherry})^T \hat{\Sigma}(a - \hat{\mu}_{cherry}) = (a - \hat{\mu}_{pear})^T \hat{\Sigma}^{-1}(a - \hat{\mu}_{pear}) \\
\implies &0 = (a - \hat{\mu}_{cherry})^T \hat{\Sigma}^{-1}(a - \hat{\mu}_{cherry}) - (a - \hat{\mu}_{pear})^T \hat{\Sigma}^{-1}(a - \hat{\mu}_{pear}) \\
&0 = a^T \hat{\Sigma}^{-1} a + 2a^T \hat{\Sigma}^{-1} \hat{\mu}_{cherry} - \hat{\mu}_{cherry}^T \hat{\Sigma}^{-1} \hat{\mu}_{cherry} - a^T \hat{\Sigma}^{-1} a - 2a^T \hat{\Sigma}^{-1} \hat{\mu}_{pear} + \hat{\mu}_{pear}^T \hat{\Sigma}^{-1} \hat{\mu}_{pear} \\
&0 = a^T \underbrace{2\hat{\Sigma}^{-1}}_{A} \underbrace{(\hat{\mu}_{cherry} - \hat{\mu}_{pear})}_{B} - \underbrace{(\hat{\mu}_{cherry}^T \hat{\Sigma}^{-1} \hat{\mu}_{cherry} - \hat{\mu}_{pear}^T \hat{\Sigma} \hat{\mu}_{pear})}_{C}
\end{aligned}
$$

Since, A is a 2×2 matrix, B is 2×1 and C is a constant then we can express the equality in terms of x and y then solve for y.

$$
\begin{aligned}
0 &= a^T AB - C \\
&= \begin{pmatrix} x & y \end{pmatrix} \begin{pmatrix} m & n \\ n & o \end{pmatrix} \begin{pmatrix} p \\ q \end{pmatrix} - c \\
&= q(mx + oy) + r(nx + py) - c \\
\implies y &= -\frac{qm + rn}{qo + rp} x + \frac{c}{qo + rp}
\end{aligned}
$$

$y = f(x)$ is the decision boundary where $\lambda = 1$ as desired.

A5. Calculating Decision Boundary for Second Classification

The decision boundary is the equation $y = f(x)$ such that $\lambda = 1$. In order to have cleaner derivation, let $a = \begin{pmatrix} x \\ y \end{pmatrix}$. We calculate the decision boundary by the following

$$\begin{aligned}
\lambda &= \frac{f_{\text{cherry}}(x_i, y_i|\hat{\mu}_{\text{cherry}}, \hat{\Sigma}_{\text{cherry}})}{f_{\text{pear}}(x_i, y_i|\hat{\mu}_{\text{pear}}, \hat{\Sigma}_{\text{pear}})} \\
&= \frac{\frac{1}{2\pi\sqrt{|\hat{\Sigma}_{\text{cherry}}|}} \exp\left\{-\frac{1}{2}(a - \hat{\mu}_{cherry})^T \hat{\Sigma}_{\text{cherry}}^{-1} (a - \hat{\mu}_{cherry})\right\}}{\frac{1}{2\pi\sqrt{|\hat{\Sigma}_{\text{pear}}|}} \exp\left\{-\frac{1}{2}(a - \hat{\mu}_{pear})^T \hat{\Sigma}_{\text{pear}}^{-1} (a - \hat{\mu}_{pear})\right\}} \\
&= \frac{\sqrt{|\hat{\Sigma}_{\text{pear}}|}}{\sqrt{|\hat{\Sigma}_{\text{cherry}}|}} \exp\left\{-\frac{1}{2}\left((a - \hat{\mu}_{cherry})^T \hat{\Sigma}_{\text{cherry}}^{-1} (a - \hat{\mu}_{cherry}) - (a - \hat{\mu}_{pear})^T \hat{\Sigma}_{\text{pear}}^{-1} (a - \hat{\mu}_{pear})\right)\right\}
\end{aligned}$$

Since, the cutoff value is at $\lambda = 1$, we set the λ equal to 1 and simplify.

$$\begin{aligned}
1 &= \frac{\sqrt{|\hat{\Sigma}_{\text{pear}}|}}{\sqrt{|\hat{\Sigma}_{\text{cherry}}|}} \exp\left\{-\frac{1}{2}\left((a - \hat{\mu}_{cherry})^T \hat{\Sigma}_{\text{cherry}}^{-1} (a - \hat{\mu}_{cherry}) - (a - \hat{\mu}_{pear})^T \hat{\Sigma}_{\text{pear}}^{-1} (a - \hat{\mu}_{pear})\right)\right\} \\
\implies \log(1) &= log\left(\frac{\sqrt{|\hat{\Sigma}_{\text{pear}}|}}{\sqrt{|\hat{\Sigma}_{\text{cherry}}|}} \exp\left\{-\frac{1}{2}\left((a - \hat{\mu}_{cherry})^T \hat{\Sigma}_{\text{cherry}}^{-1} (a - \hat{\mu}_{cherry}) - (a - \hat{\mu}_{pear})^T \hat{\Sigma}_{\text{pear}}^{-1} (a - \hat{\mu}_{pear})\right)\right\}\right) \\
\implies 0 &= log\left(\frac{\sqrt{|\hat{\Sigma}_{\text{pear}}|}}{\sqrt{|\hat{\Sigma}_{\text{cherry}}|}}\right) - \frac{1}{2}\left((a - \hat{\mu}_{cherry})^T \hat{\Sigma}_{\text{cherry}}^{-1} (a - \hat{\mu}_{cherry}) - (a - \hat{\mu}_{pear})^T \hat{\Sigma}_{\text{pear}}^{-1} (a - \hat{\mu}_{pear})\right) \\
&= a^T \underbrace{(\hat{\Sigma}_{\text{cherry}}^{-1} - \hat{\Sigma}_{\text{pear}}^{-1})}_{A} a - a^T \underbrace{2(\hat{\Sigma}_{\text{cherry}}^{-1} + \hat{\mu}_{\text{cherry}} - \hat{\Sigma}_{\text{pear}}^{-1}\hat{\mu}_{\text{pear}})}_{B} + \underbrace{log\left(\frac{\sqrt{|\hat{\Sigma}_{\text{cherry}}|}}{\sqrt{|\hat{\Sigma}_{\text{pear}}|}}\right)}_{C}
\end{aligned}$$

Since, A is a 2×2 matrix, B is 2×1 and C is a constant then we can express the equality in terms of x and y then solve for y.

$$\begin{aligned}
0 &= a^T Aa - aB + C \\
&= \begin{pmatrix} x & y \end{pmatrix} \begin{pmatrix} m & n \\ n & o \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} - \begin{pmatrix} x & y \end{pmatrix} \begin{pmatrix} p \\ q \end{pmatrix} - c \\
&= x(xm + yn) + y(xn + yo) + px + qy + c \\
\implies 0 &= (o)y^2 + (2nx + q)y + (mx^2 + px + c) \\
\implies y &= \frac{-(2nx + q)}{2o}\sqrt{(2nx + q)^2 - 4o(mx^2 + px + c)} \\
&y = f(x) \text{ is the decision boundary where } \lambda = 1 \text{ as desired.}
\end{aligned}$$