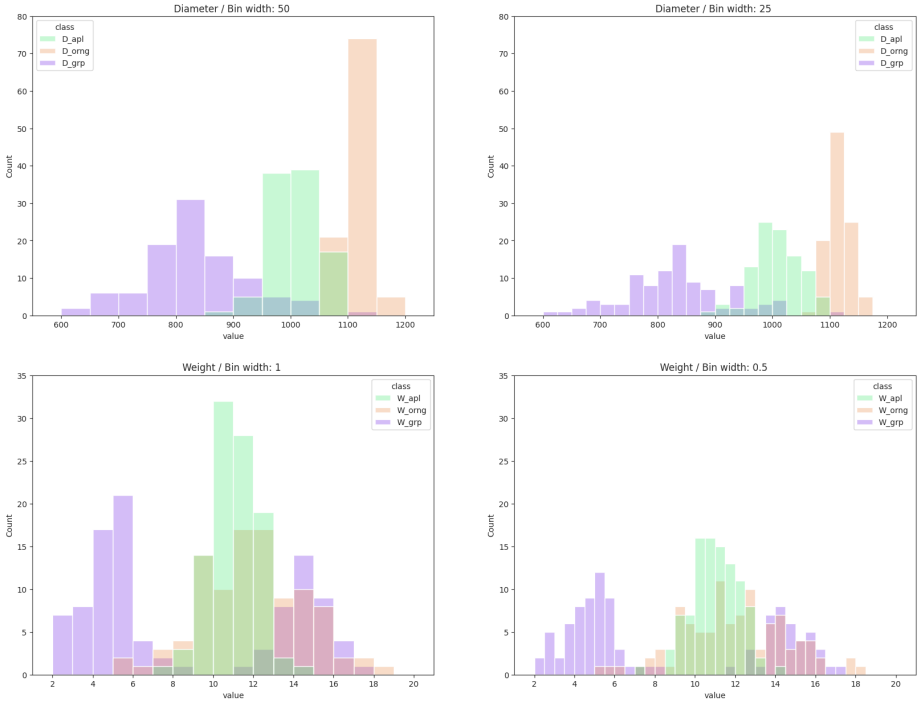
BIOM/SYSC5405 – Pattern Classification and Experiment Design

Assignment 1— Peer Evaluation Guide

Question A

For Question A, please grade Q1b, where students were asked to provide histograms for each of the two features. There should be a total of four plots a pair of plots for **weight** (with different bin widths used in each plot) and a pair of plots for **diameter** (again, with different bin widths). Each histogram should clearly show all three fruit classes (e.g., using transparency to permit viewing of overlapping classes). Example histograms follow. Finally, the student correctly states that **diameter is the preferred feature**, since the classes have the least overlap.

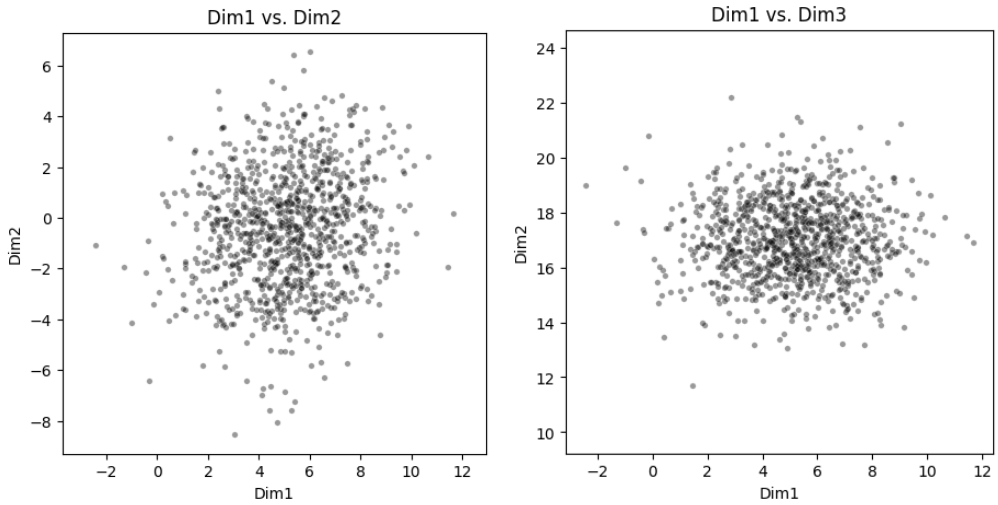


Common errors:

- Overlapping classes cannot be viewed (transparency was not used)
- Only a single set of histograms were provided (2 sets are required with different bin widths)
- The wrong type of plot is used (e.g., line graph, scatter plot)
- Separate plots are used for each fruit type (should be a single plot for each feature showing the distribution of all three fruit types)
- The student does not guess which feature is preferred
- The student guess incorrectly about which feature is preferred.
- The student does not state WHY diameter is preferred (something along the lines of "less overlap between the classes")

Question B

For Question A, please grade Q2b, where students were asked to create two scatter plots (dim1 vs. dim2 and dim1 vs. dim3).



Lastly, students are asked to comment on why the scatter plots look different. They should mention that dim1 and dim3 are uncorrelated (independent), whereas dim1 and dim2 are correlated (non-zero off-diagonal values in the covariance matrix at positions (1,2) and (2,1). This causes the scatter plot to 'rotate' such that the principal axes are no longer aligned with the x and y axis. The variances along the principal axes of each scatter plot are also different, leading to different sized clusters. **Student must mention at least one of these two things** (correlated vs. uncorrelated → rotated vs. circular) or cluster/distribution variance is greater in Dim2 direction in first plot vs. Dim3 direction in second plot.

Common Errors

- Plots are missing x- or y-axis labels
- Plots are not titled. If there is a descriptive caption, no title is required.
- Scale of x- and y-axis are not equal (i.e., more stretched in one direction). The scale between the left and right scatter plots can be different.
- Student did not comment on difference between plots mentioning one of the two points above
- Wrong plot type was used (e.g., line graph, bar graph, etc)

Question C

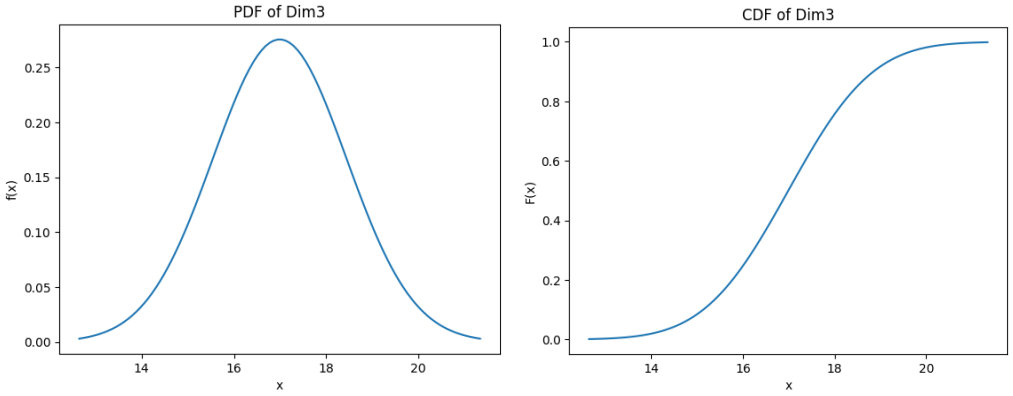
Please grade questions 2c) and 2e).

Q2c) The trace is the sum of the diagonal values and is = **11**. The student should state that the covariance matrix **IS positive definite** SINCE it is square, symmetric, and **one of the following reasons**:

- All its eigenvalues are positive
- All its 'pivots' are positive
- All its 'upper left determinants' are positive: $|4|$, $\left|\begin{bmatrix} 4 & 0.5 \\ 0.5 & 5 \end{bmatrix}\right|$, $\left|\begin{bmatrix} 4 & 0.5 & 0 \\ 0.5 & 5 & -0.2 \\ 0 & -0.2 & 2 \end{bmatrix}\right|$.

Reasons above come from
https://www.math.utah.edu/~zwick/Classes/Fall2012_2270/Lectures/Lecture33_with_Examples.pdf

Q2e) PDF and CDF are both plotted and look like the following. Note that some students will estimate the mean and variance from their sample of 1000 points, leading to slightly different estimates and therefore slightly different heights/positions of the PDF and CDF curves. Please accept some slight differences in plot height and x-shift/position.



Common Errors:

- 2c) Trace is incorrect
- 2c) Wrong conclusion about positive definite test
- 2c) No explanation given for conclusion about positive definite test
- 2e) The wrong type of plot is used for PDF/CDF (should be line graphs)
- 2e) PDF or CDF missing
- 2e) PDF or CDF don't match sample plots above (range of x should be similar to plots above; mean of PDF should match; height of curves should match above).
- 2e) Mean or variance is incorrect in PDF
- 2e) CDF should rise smoothly and monotonically from 0 to 1