

Data Mining Assignment 1

Anjishnu Mukherjee

January 2020

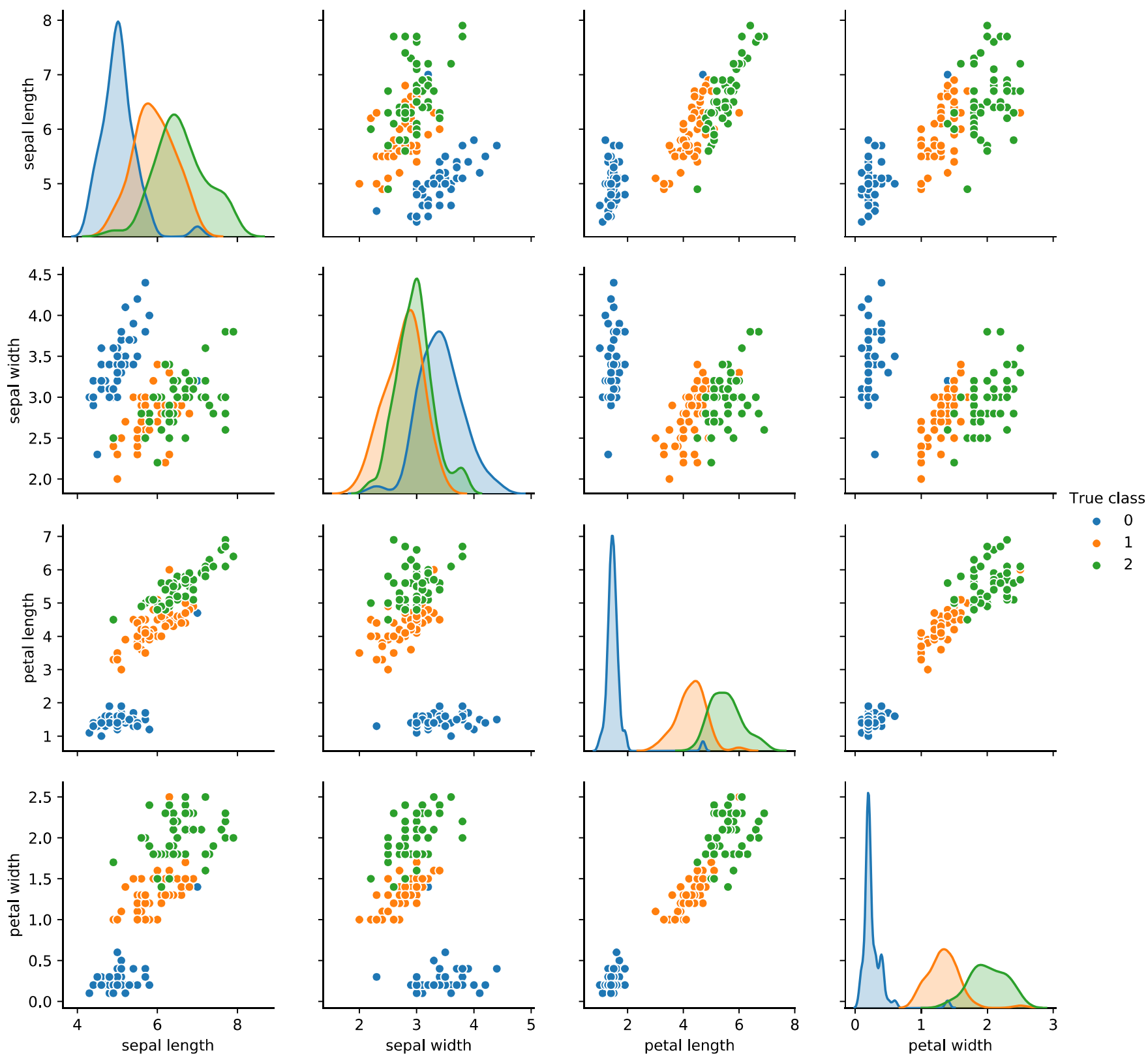
The solution does the analysis in two parts. First it considers $k=3$, then it considers $k=2$. Both of these when analyzed using pair-plots, clearly reveal that Iris-setosa is linearly separable from the other 2 classes. Iris-versicolour and Iris-virginica display a strong degree of correlation to the extent that some of the members of Iris-versicolor are misclassified by the algorithm developed as belonging to the Iris-virginica class.

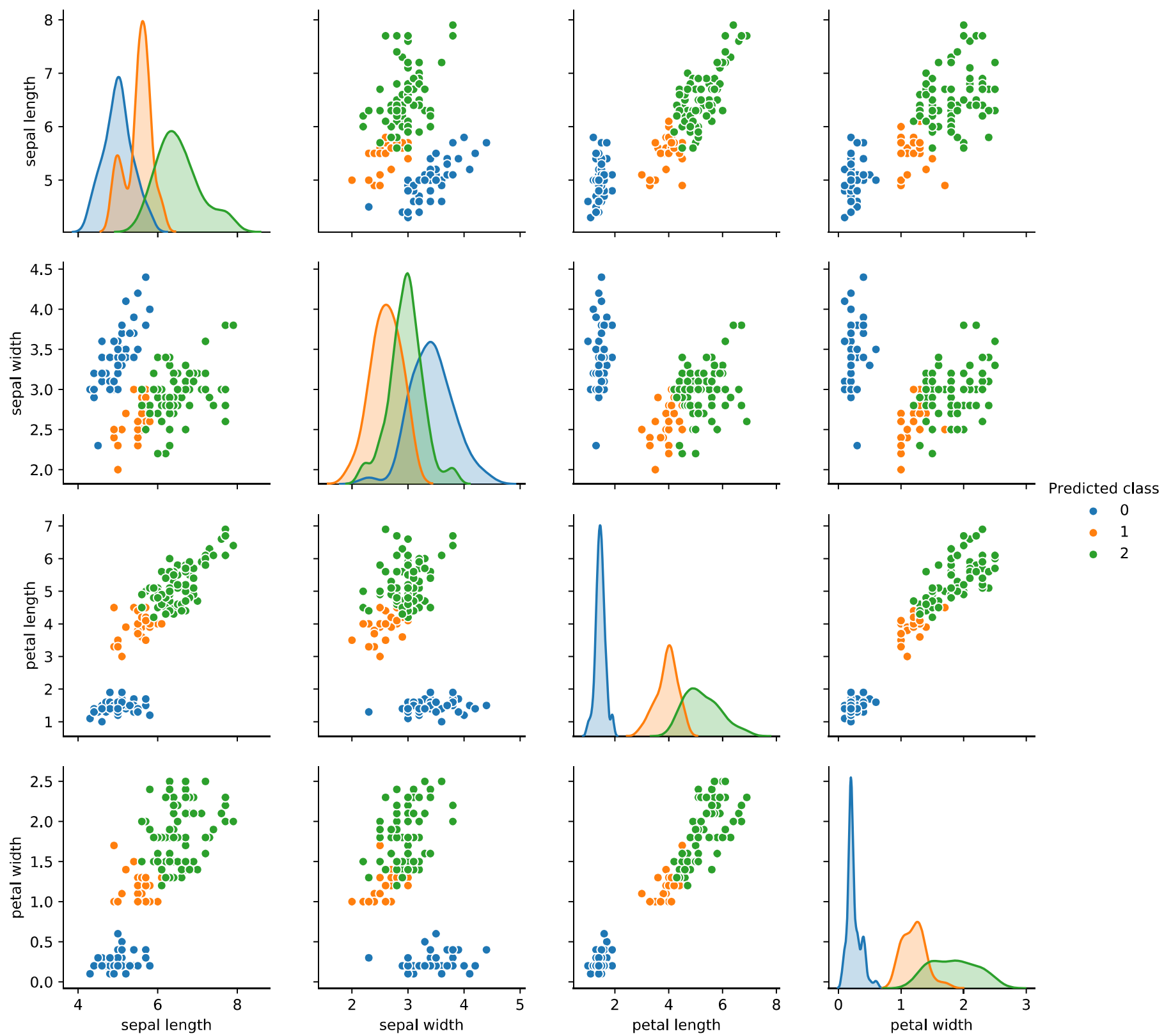
6 graphs are included next.

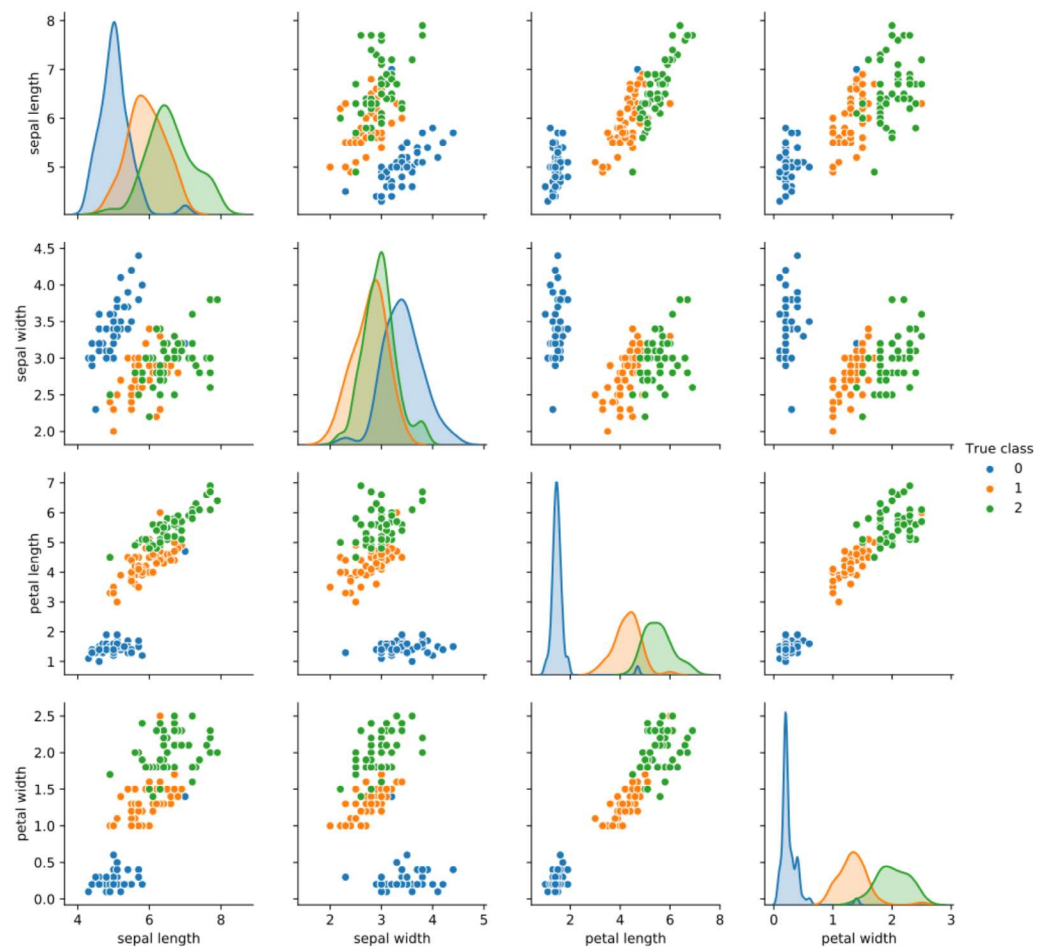
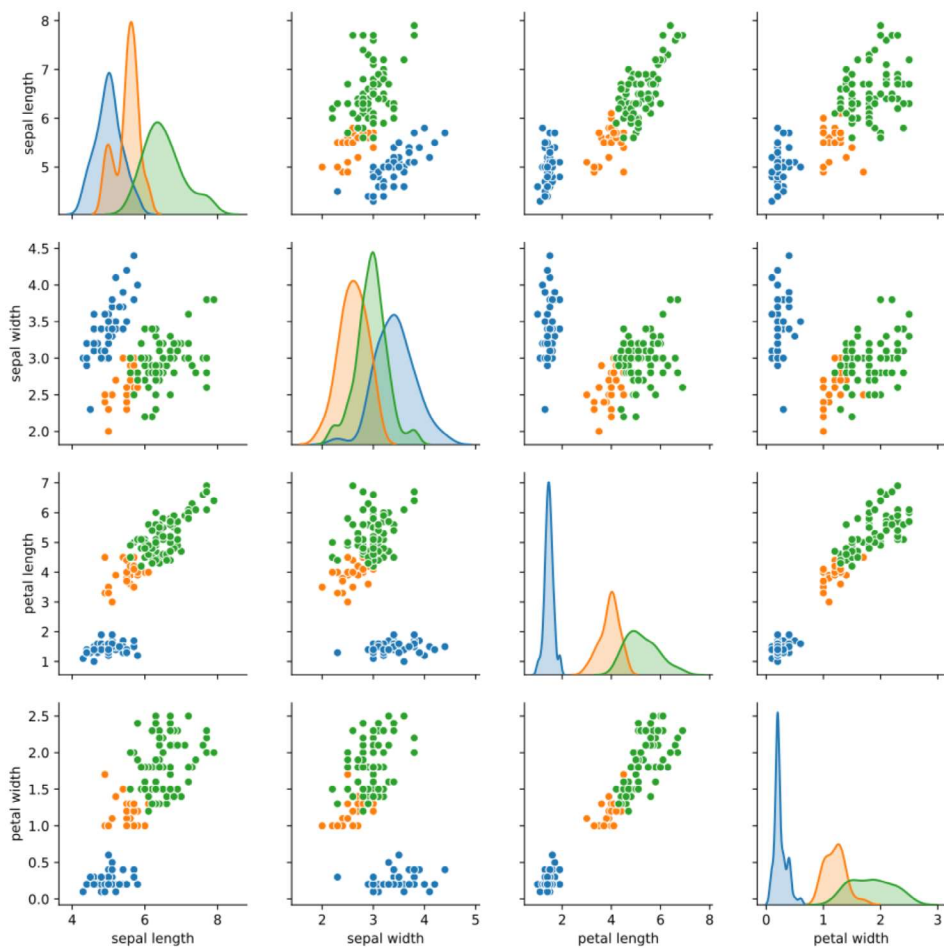
1. For $k = 3$, the pair-plots using the original dataset
2. For $k = 3$, the pair-plots using the clustering algorithm developed
3. Side-by-side comparison of the two images above
4. For $k = 2$, the pair-plots using the original dataset
5. For $k = 2$, the pair-plots using the clustering algorithm developed
6. Side-by-side comparison of the two images above

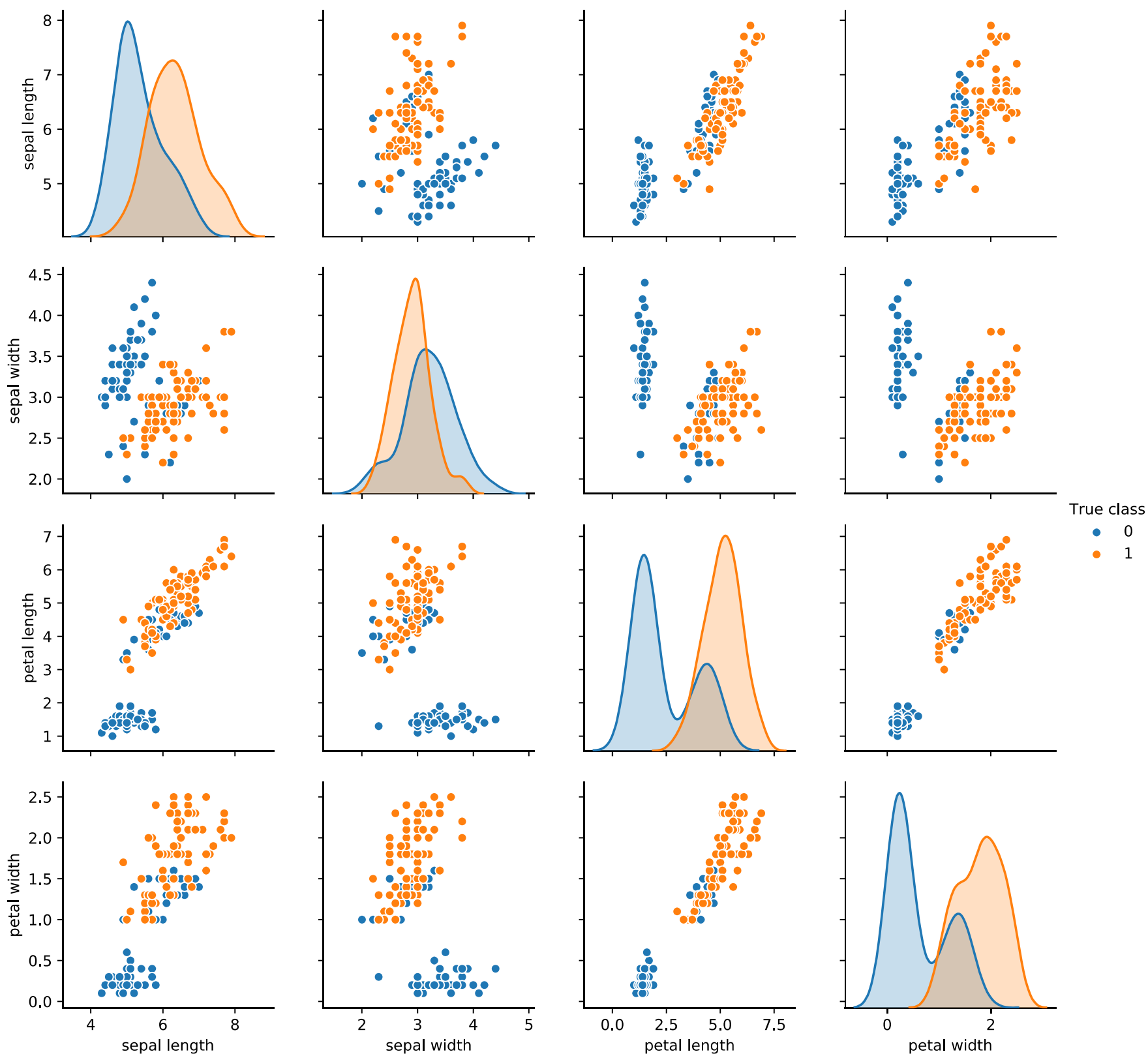
The nomenclature for the legend used in the graphs is

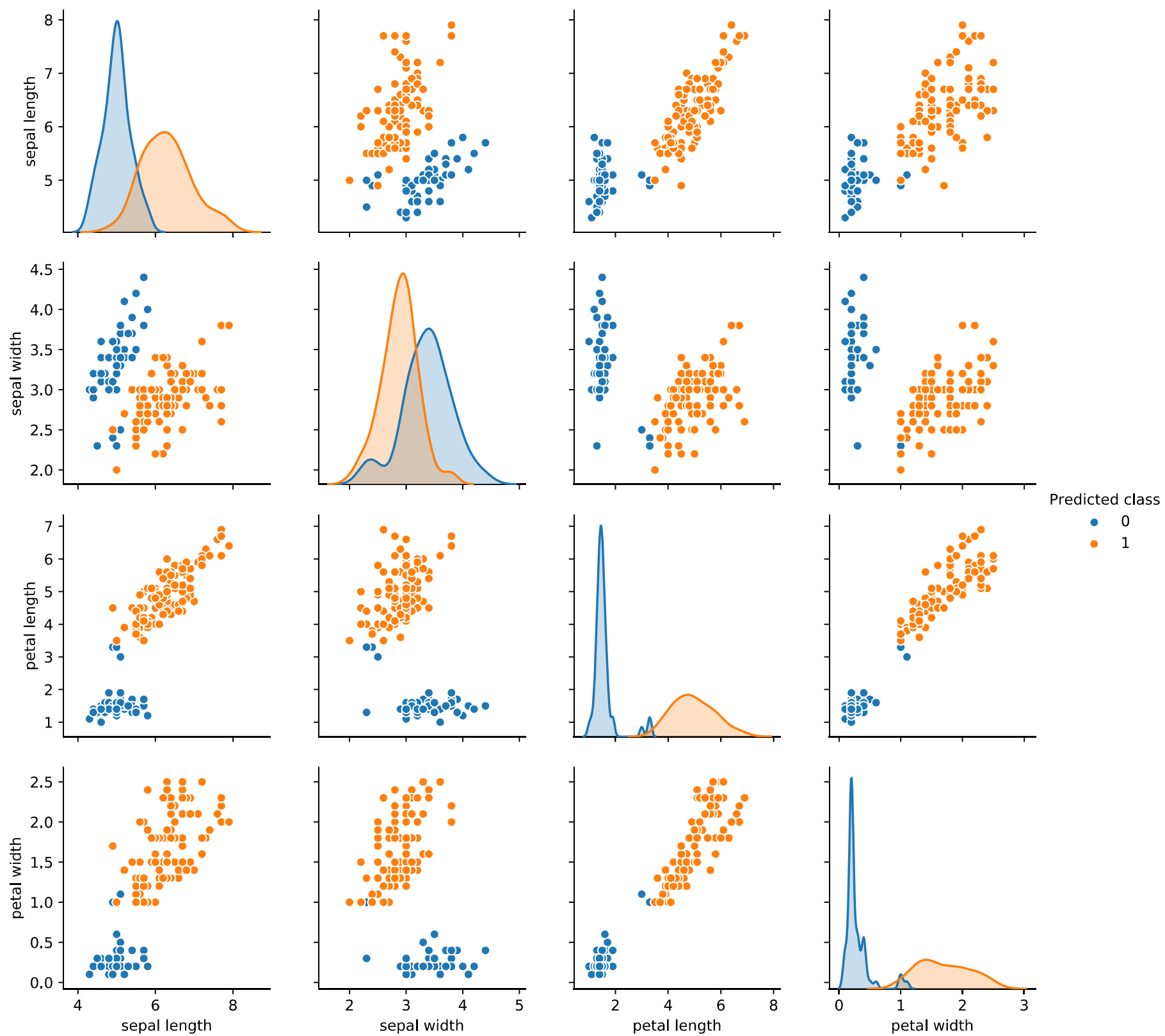
- 0 — Iris-setosa (blue)
- 1 — Iris-versicolour (yellow)
- 2 — Iris-virginica (green)

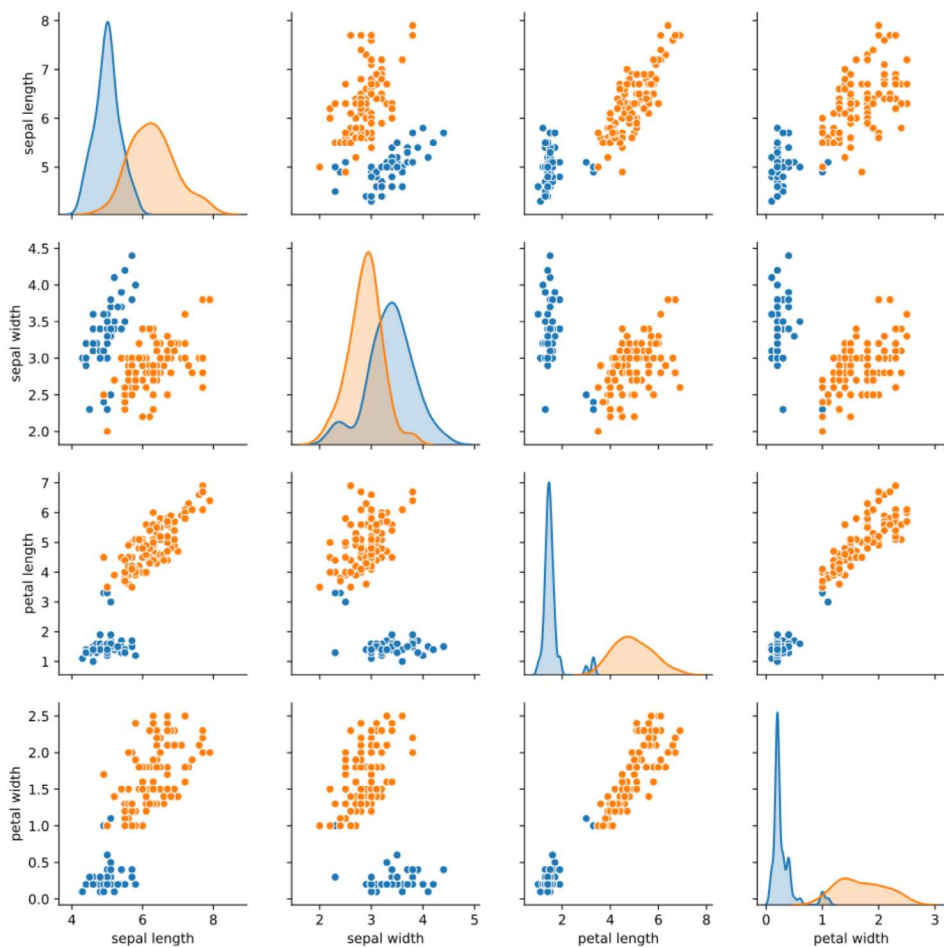




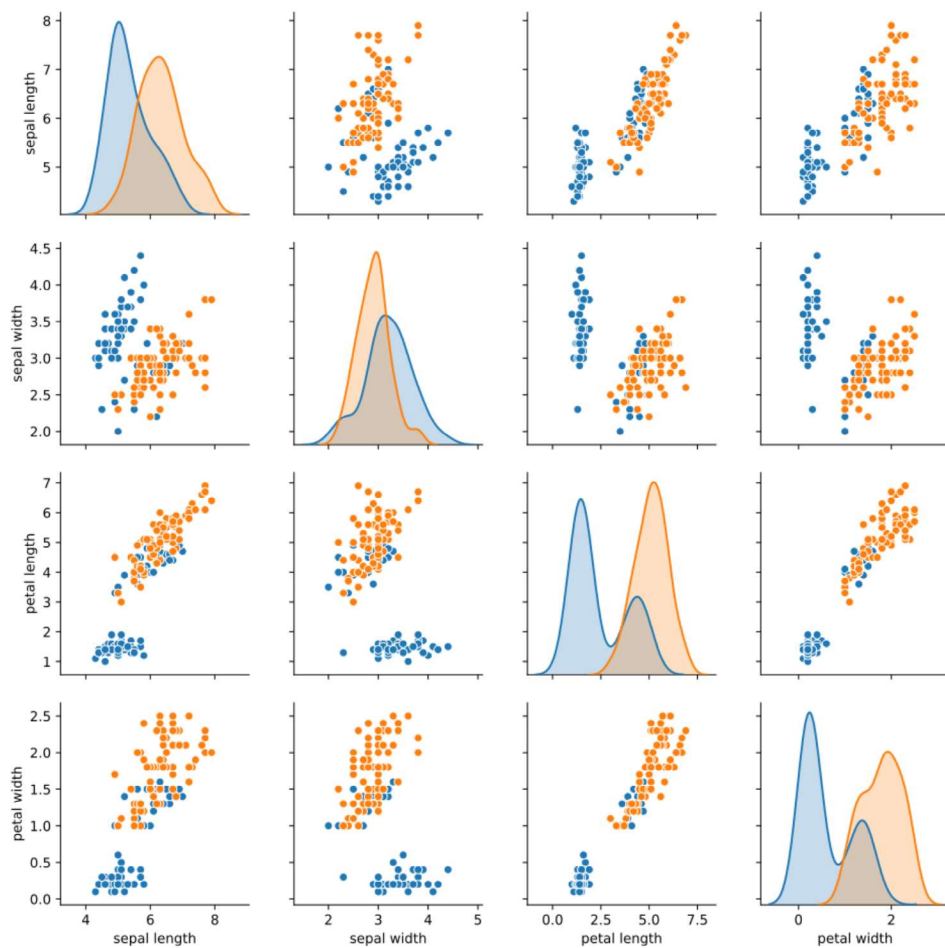








Predicted class
• 0
• 1



True class
• 0
• 1