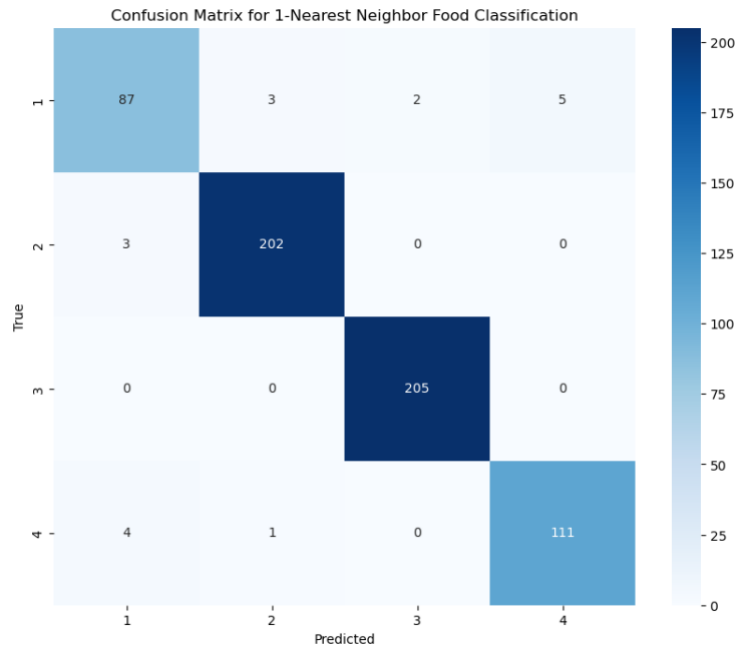


**Homework 2**  
**Problem 3**  
**Ayad Masud**  
**733009045**

*I certify that I have personally done the coding, generated the figures and written the report without aid from anybody else, and that I have not plagiarized, self-plagiarized, or used AI-generated text. I certify that I have acknowledged any sources I used to complete this assignment. ARM.*

## 1 Part 1: 1-KNN Classifier

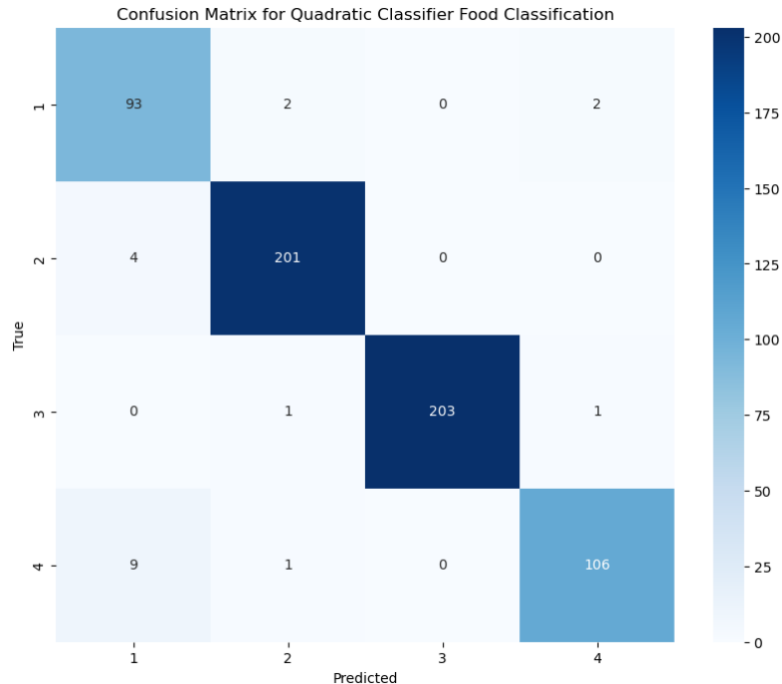
Figure 1 shows the confusion matrix for the 1-KNN classifier. It achieved a 97% accuracy. Along the diagonal we can see the correct predictions and off the diagonal are the incorrect predictions. If we analyze the matrix a little more, we can see that the most misclassified class is class 1. It is often misclassified with class 4. This may be because breakfast cereals and baked goods have some of the same ingredients and macros. Class 3 is misclassified the least amount of times which makes sense because class 3 is protein and is distinctly different from the other classes.



**Figure 1.** Confusion matrix for 1-KNN Classifier

## 2 Part 2: Quadratic Classifier

Figure 2 shows the confusion matrix for a quadratic classifier. The assumption that is made in the problem statement is not valid. This is because many features in food items are highly correlated. For example, a food with lots of fat will likely also be high in calories. So, it is unreasonable to assume a diagonal covariance matrix where the features are completely uncorrelated. The quadratic classifier achieved an accuracy of 96.7% which is slightly less than the 1-KNN classifier. However, the quadratic classifier is better at classifying class 1, worse at classifying class 4, and around the same performance for classes 2 and 3.



**Figure 2.** Confusion matrix for Quadratic Classifier

### 3 Part 3: Forward Feature Selection

The features selected after running forward feature selection on the dataset are as listed:

Selected features:

1. Feature\_3 (Protein)
2. Feature\_5 (Ash)
3. Feature\_9 (Calcium)
4. Feature\_18 (Selenium)
5. Feature\_25 (Total Folate)

The features picked make sense. Feature number 3 and that is protein which easily separates class 3 from the other classes, hence, why it is the first feature picked by the forward feature selection algorithm. Feature number 5 is also a good feature because the only class that would have ash is class 4, which makes it a good feature to separate class 4 from the other classes. Feature number 9 is calcium which can be used to separate class 2 because some of the leafy green vegetables have high calcium content compared to other classes. Feature number 25 is also used to separate class 2 from other classes because vegetables contain a high amount of folate.

### 4 Resources used to achieve this goal

**Canvas:** Homework template, lecture slides

**Python Libraries:** NumPy, pandas, seaborn, matplotlib, scikit learn

## 5 References

- “1.13. Feature Selection — Scikit-Learn 0.21.3 Documentation.” *Scikit-Learn.org*, 2019, [scikit-learn.org/stable/modules/feature\\_selection.html](https://scikit-learn.org/stable/modules/feature_selection.html).
- “Sklern.discriminant\_analysis.QuadraticDiscriminantAnalysis.” *Scikit-Learn*, [scikit-learn.org/stable/modules/generated/sklern.discriminant\\_analysis.QuadraticDiscriminantAnalysis.html#sklern.discriminant\\_analysis.QuadraticDiscriminantAnalysis](https://scikit-learn.org/stable/modules/generated/sklern.discriminant_analysis.QuadraticDiscriminantAnalysis.html#sklern.discriminant_analysis.QuadraticDiscriminantAnalysis).
- “Nearest Neighbors Classification.” *Scikit-Learn*, [scikit-learn.org/stable/auto\\_examples/neighbors/plot\\_classification.html#k-nearest-neighbors-classifier](https://scikit-learn.org/stable/auto_examples/neighbors/plot_classification.html#k-nearest-neighbors-classifier)