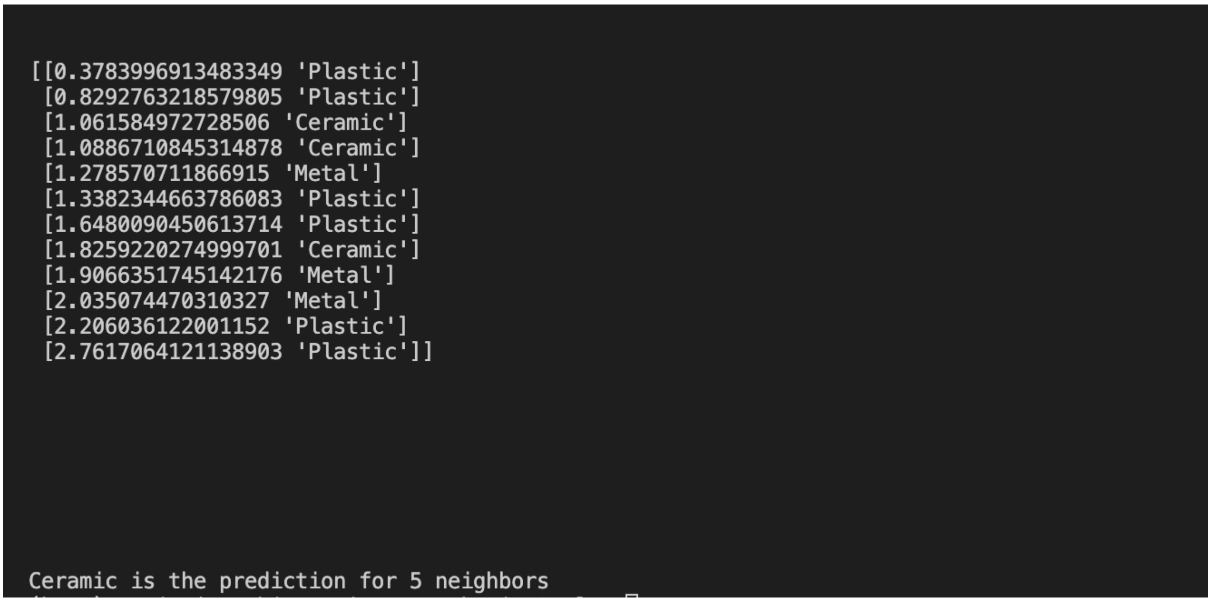
Machine Learning Homework 1

The k-nearest neighbors algorithm is a supervised learning classifier that uses proximity to predict the grouping of a single data point. Distance calculation is used to determine the nearest neighbors, which can be calculated using algorithms such as Euclidean, Manhattan, and Haversine distances.

Essentially, the algorithm calculates the distance between the data point and the provided dataset, identifies the nearest neighbors based on that distance, and makes a prediction accordingly. An example of finding the k-nearest neighbors using Cartesian distance is shown using a 2-D array that displays the distance between the given sample data point and its corresponding label. The algorithm selects the top k elements based on a specified value of k and returns the class with the highest number of occurrences.

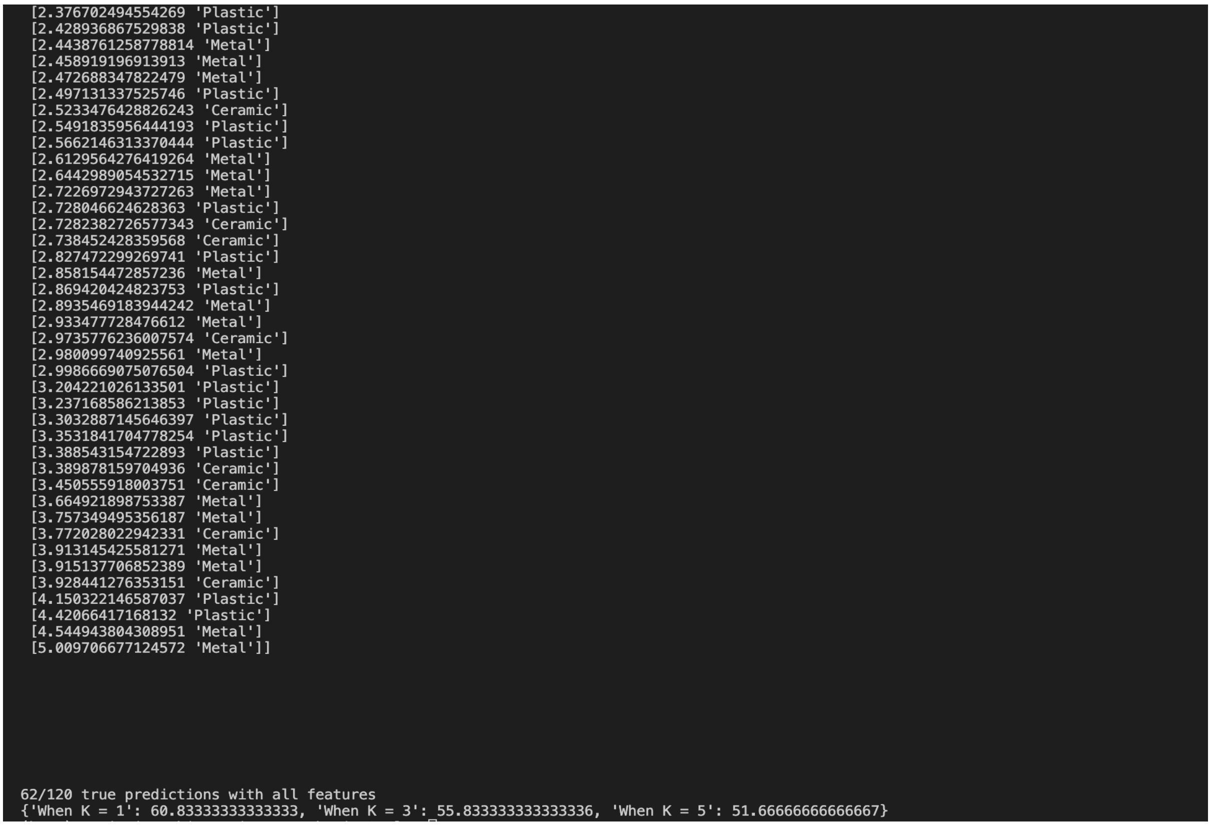
Details

Question 2a produces the following output



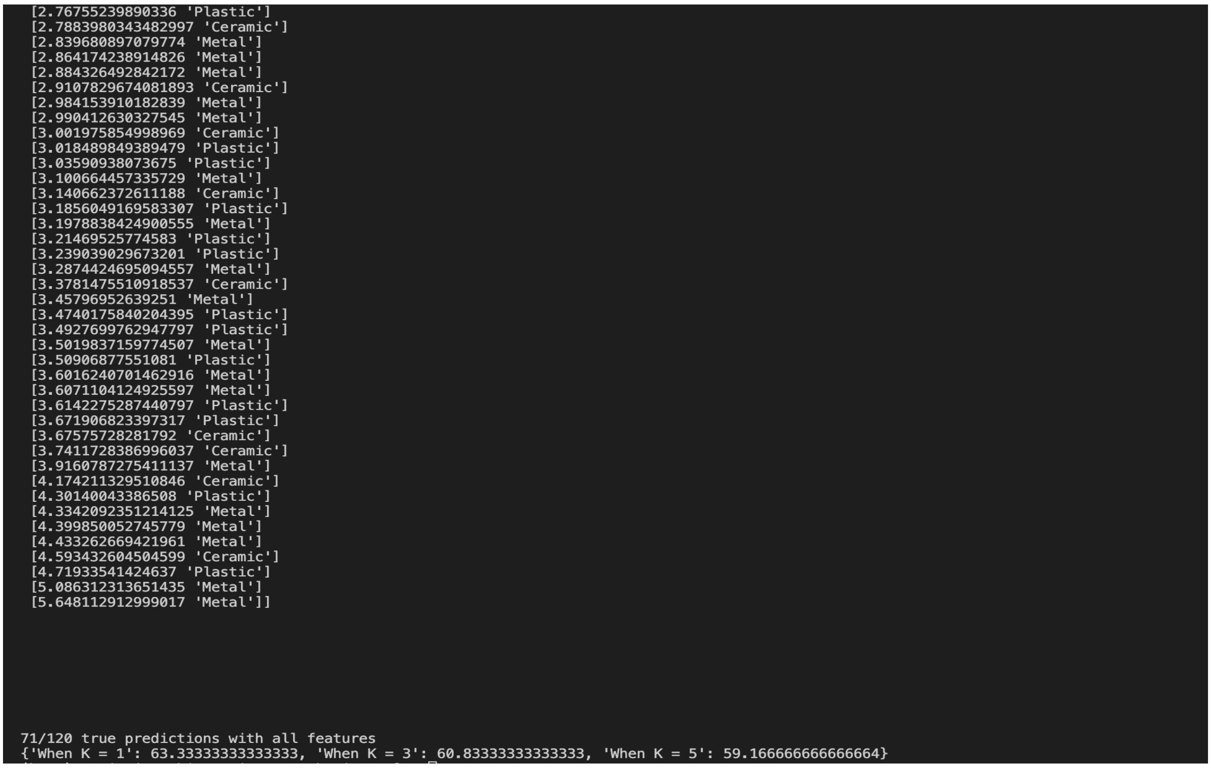
```
[[0.3783996913483349 'Plastic']
 [0.8292763218579805 'Plastic']
 [1.061584972728506 'Ceramic']
 [1.0886710845314878 'Ceramic']
 [1.278570711866915 'Metal']
 [1.3382344663786083 'Plastic']
 [1.6480090450613714 'Plastic']
 [1.8259220274999701 'Ceramic']
 [1.9066351745142176 'Metal']
 [2.035074470310327 'Metal']
 [2.206036122001152 'Plastic']
 [2.7617064121138903 'Plastic']]







Ceramic is the prediction for 5 neighbors
```

Question 2c tells us to use knn method by the leave-one-out method. The output using cartesian distance looks like -



```
 [2.376702494554269 'Plastic']
 [2.428936867529838 'Plastic']
 [2.4438761258778814 'Metal']
 [2.458919196913913 'Metal']
 [2.472688347822479 'Metal']
 [2.497131337525746 'Plastic']
 [2.5233476428826243 'Ceramic']
 [2.5491835956444193 'Plastic']
 [2.5662146313370444 'Plastic']
 [2.6129564276419264 'Metal']
 [2.6442989054532715 'Metal']
 [2.7226972943727263 'Metal']
 [2.728046624628363 'Plastic']
 [2.7282382726577343 'Ceramic']
 [2.738452428359568 'Ceramic']
 [2.827472299269741 'Plastic']
 [2.858154472857236 'Metal']
 [2.869420424823753 'Plastic']
 [2.8935469183944242 'Metal']
 [2.933477728476612 'Metal']
 [2.9735776236007574 'Ceramic']
 [2.980099740925561 'Metal']
 [2.9986669075076504 'Plastic']
 [3.204221026133501 'Plastic']
 [3.237168586213853 'Plastic']
 [3.3032887145646397 'Plastic']
 [3.3531841704778254 'Plastic']
 [3.388543154722893 'Plastic']
 [3.389878159704936 'Ceramic']
 [3.450555918003751 'Ceramic']
 [3.664921898753387 'Metal']
 [3.757349495356187 'Metal']
 [3.772028022942331 'Ceramic']
 [3.913145425581271 'Metal']
 [3.915137706852389 'Metal']
 [3.928441276353151 'Ceramic']
 [4.150322146587037 'Plastic']
 [4.42066417168132 'Plastic']
 [4.544943804308951 'Metal']
 [5.009706677124572 'Metal']]




62/120 true predictions with all features
{'When K = 1': 60.83333333333333, 'When K = 3': 55.833333333333336, 'When K = 5': 51.66666666666667}
```

Question 2d tells us to use knn method by the leave-one-out method. The output using manhattan distance looks like -



```
 [2.76755239890336 'Plastic']
 [2.7883980343482997 'Ceramic']
 [2.839680897079774 'Metal']
 [2.864174238914826 'Metal']
 [2.884326492842172 'Metal']
 [2.9107829674081893 'Ceramic']
 [2.984153910182839 'Metal']
 [2.990412630327545 'Metal']
 [3.001975854998969 'Ceramic']
 [3.018489849389479 'Plastic']
 [3.03590938073675 'Plastic']
 [3.100664457335729 'Metal']
 [3.140662372611188 'Ceramic']
 [3.1856049169583307 'Plastic']
 [3.1978838424900555 'Metal']
 [3.21469525774583 'Plastic']
 [3.239039029673201 'Plastic']
 [3.2874424695094557 'Metal']
 [3.3781475510918537 'Ceramic']
 [3.45796952639251 'Metal']
 [3.4740175840204395 'Plastic']
 [3.4927699762947797 'Plastic']
 [3.5019837159774507 'Metal']
 [3.50906877551081 'Plastic']
 [3.6016240701462916 'Metal']
 [3.6071104124925597 'Metal']
 [3.6142275287440797 'Plastic']
 [3.671906823397317 'Plastic']
 [3.67575728281792 'Ceramic']
 [3.7411728386996037 'Ceramic']
 [3.9160787275411137 'Metal']
 [4.174211329510846 'Ceramic']
 [4.30140043386508 'Plastic']
 [4.3342092351214125 'Metal']
 [4.399850052745779 'Metal']
 [4.433262669421961 'Metal']
 [4.593432604504599 'Ceramic']
 [4.71933541424637 'Plastic']
 [5.086312313651435 'Metal']
 [5.648112912999017 'Metal']]




71/120 true predictions with all features
{'When K = 1': 63.33333333333333, 'When K = 3': 60.83333333333333, 'When K = 5': 59.166666666666664}
```

Conclusion

It is observed by looking at the overall output that the Manhattan distance is more accurate for all the values of k, especially for k = 3 and k = 5