**How kNN and decision trees work for classification and regression**

**For classification:**

In conclusion, DT is better than kNN for classification. The accurancy of DT is .80 and the accurancy of kNN is 0.785.

The KNN algorithm has a relatively high degree of adaptation to numerical data, and has less preprocessing. Generally, a single type of data can be normalized, and the formula for selecting distance can also be carried out based on the actual situation. One advantage of the KNN algorithm is that it is not sensitive to outliers, but during preprocessing, if the extreme data can be removed and then normalized, the classification effect will be better.

The decision tree algorithm has higher requirements for data preprocessing and requires pre-classification. The classification process will be better if it is oriented to the problem itself. If it is used to make the actual algorithm and put it into use, it will be better for a specific user to perform a scaling effect by the user. However, when there is too much data, more consideration should be given to the pruning of the decision tree, but it will inevitably reduce the accuracy.

**For regression:**

Decision trees are supervised and non-parametric, and have a structure similar to a flowchart. These trees are easily interpretable because you can see each "decision" the machine made and why it got what it got. kNN, or K-Nearest Neighbor, is also non-parametric and supervised. In kNN, the k is a constant defined by the user.

Both methods being non-parametric mean the distribution of data cannot be defined in a few parameters. (No assumptions) Decision trees are better are finding non-linear relationships, making them worse than say linear regression. In addition, bigger trees mean more overfitting, which means we must prune it. kNN on the other hand becomes slower with larger datasets, and becomes sensible to outliers.

kNN works with regression by approximating the relationship between independent variables and the continuous outcome via the average of the observations. Decision trees will regress the data by asserting true or false to specific questions, causing it to branch off – like a tree- into a true and false.

**How the 3 clustering methods of step 3 work?**

Given the ground truth in the data set, none of the algorithms seemed to work all that well in creating five distinct clusters. The intended learning process for this data is likely determining position from the relative distances and directions between each 3D coordinate point. Hierarchical clustering works best with small data sets, and even finding a subset that consisted of one user for maximum consistency resulted in poor results with no clear cut. K-Means clustering handled the large data set better, somewhat accurately finding 3 clusters and struggling with the remaining two. Using the same data for the model-based clustering, BIC recommended a VVV selection (varying shape, volume, and orientation) with nine different clusters, which is again totally different from the expected five clusters, as this function was not forced to adhere to the same parameters as K-Means clustering.

**How PCA and LDA work, and why they might be useful techiques for machine learning**

PCA (Principal Component Analysis) is a type of unsupervised learning, which removes correlation between independent variable and looks for hidden latent variable, and reduces noise.

LDA (Linear Discriminant Analysis) is a supervised algorithm that performs classification and dimensionality reduction at the same time. The key to LDA is that when classification is performed, the variance within the class is minimized, but the variance between classes is maximized.

The difference in these two comes from PCA finding a new axis that allows the data to be widely distributed, while the LDA finds a new axis that best classifies data.

PCA performs better when the difference in class comes from the difference in variance, and LDA performs better when the difference in class of data comes from the difference in mean. Also, When data is represented in a 3-dimensional plot, LDA shows a better performance than PCA as well in classification.