ELEC70073 Pattern Recognition Coursework Report

Peng Yunan, (kp620, 01868105), Xue Zihao, (xz1420, 01874225)

1. Section A: Data Preparation

1.1. A.1

The quality of pattern recognition highly depends on the raw data. If the data is non-distinguishable across objects, the effectiveness of our algorithms may not be visible; so it is crucial to select a time instance that we can differentiate between the data for different objects. By visualizing the data from all 10 trails across 6 objects, we choose **Time = 50** to be the time instance. The reasons are followed:

1. We can spot from the PVT and Electrode measurement(figure7 to 14) that at t=50 the data shows visible separability.

2. Even though a smaller time instance such as t=10 shows a bigger separation between different objects, choosing them is unrealistic because the data are not stabilized or converged. This means the robot may just start touching the object and will bring additional variation to our measurement as we cannot tell whether the separation in measurement is due to physical properties or random fluctuations. Moreover, pattern recognition algorithms such as PCA or LDA contain statistical measurements like standard deviation and mean; these statistical properties are more consistent under steady data.

1.2. A.2

We choose measurements from finger F0 and the PVT and Electrode data are stored in F0_PVT.mat and F0_Electrodes.mat files.

1.3. A.3

The 3D scatter plot of the complete contents of the PVT mat file is shown in figure1. In the following sections, acrylic will be in dark blue, black foam will be in red, car sponge will be in yellow, flour sack will be in purple, kitchen sponge will be in green and steel vase will be in light blue.

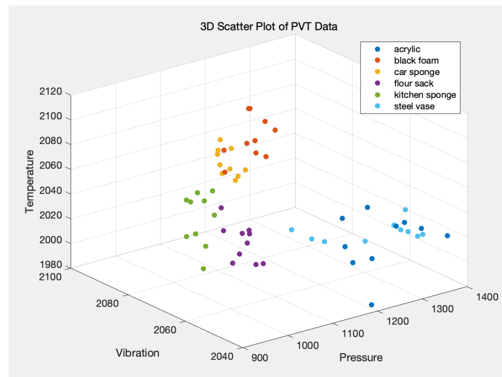

Figure 1. 3D scatter plot of complete PVT measurement

2. Section B: Principal Component Analysis

2.1. B.1

a. The **covariance matrix**, **eigenvalues** and **eigenvectors** for the data are:

$$CovarianceMatrix = \begin{bmatrix} 1.0000 & -0.3568 & -0.3185 \\ -0.3568 & 1.0000 & 0.2693 \\ -0.3185 & 0.2693 & 1.0000 \end{bmatrix}$$

$$Eigenvalues = \begin{bmatrix} 0.6337 & 0 & 0 \\ 0 & 0.7352 & 0 \\ 0 & 0 & 1.6311 \end{bmatrix}$$

$$Eigenvectors = \begin{bmatrix} 0.7803 & -0.1630 & -0.6038 \\ 0.5685 & 0.5870 & 0.5764 \\ 0.2605 & -0.7930 & 0.5507 \end{bmatrix}$$

b. The Standardised data with three PCs displayed are shown in figure2. The standardized data shown in figure2 has similar distribution to that in figure1, meaning the standardization process does not change data properties.

c. The reduced standardised data in 2D is shown in figure15. We can see clusters in different objects, which supports the fact that PCA preserves objects' informative properties.

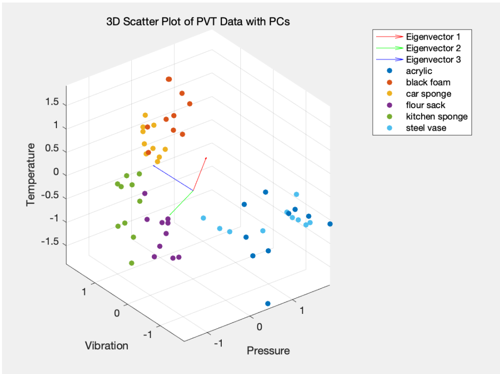

Figure 2. Standardised PVT data with the Principal components displayed

d. Figure3 shows how the PVT data is distributed across all principal components.

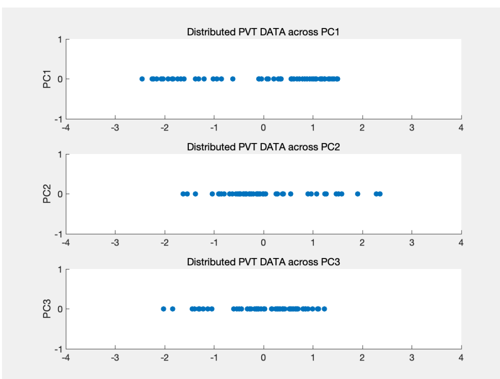

Figure 3. PVT data distribution across all principal components

e. PCA seeks to reduce the dimensionality of a dataset while preserving as much of the data's variation as possible. It does this by identifying the directions along which the variance of the data is maximized. These directions are the principal components, and they define a new coordinate system for the data. Theoretically, the PVT data will have 3 principal components, which we have confirmed from the calculations. The three principle components are ranked by the corresponding eigenvalues, which quantifies the variance. So the first principle component(PC1) will be the one that captures greatest variance, or most informative feature of PVT data. We can see from the figure3 that the projected data from PC1 indeed shows greatest variance, with projected data from PC2 and PC3 show gradually decreased variance. Furthermore, we observe from figure2 that the three principle components are orthogonal from each other. The orthogonality of principal components ensures that the variance captured by each component is unique and does not overlap with the variance captured by other components. This is important because it means that each principal component contributes a distinct aspect of the data's overall variance, allowing for a more interpretable reduction in dimensionality.

2.2. B.2

a. The Scree plot showing the variances of each principal components is shown in figure4.

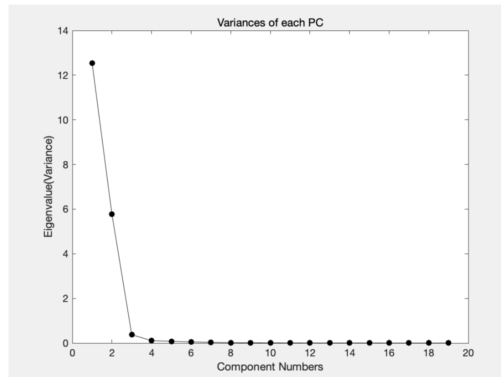

Figure 4. Scree plot showing the variances of each principal components of the electrode data

b. The visualization of the electrode data using the three principal components with largest variance is shown in figure 5.

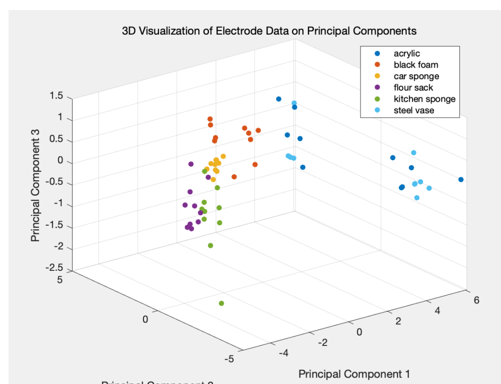

Figure 5. Electrode data visualization using the three PCs with largest variance

c. Scree plot is a line segment plot that shows how variance is distributed across the principal components. We can infer that most variance concentrate on the first four principal components, meaning they are most significant in capturing information of Electrodes data and additional components do not explain much variance. Therefore, choosing the first three PCs as instructed is an effective way to achieve dimensionality reduction. Indeed, from the visualization of the Electrode data projected onto the three PCs, we can see data are generally well clustered and the clusters are well separated. Besides the above observation, the clusters of acrylic and steel vase are very close, meaning they may share similar properties. Furthermore, clusters like car sponge are more compact, meaning they have similar electrode measurement, while kitchen sponge has more variability because its cluster is more spread out. Overall, we can conclude that by only using three PCs, the lower-dimensional representation of data will retain key information needed to differentiate between different objects, and the well separability supports the fact from the Scree plot that the variance captured is meaningful and related to the inherent differences between the objects.

3. Section C: Linear Discriminant Analysis

3.1. C.1

a. LDA is designed to create a model that maximizes the separation between classes. It does this by finding a linear combination of features. The results of using LDA to split the training data in terms of pairs of features are shown in figure16 to 18. To observe the effective of LDA, we also show the projected data in figure 19 to 21.

b. The result of applying LDA to three-dimensional PVT data is shown in figure22, the green hyperplane represents the LDA decision boundary. To interpret the effectiveness, the 2D and 1D projections are shown in figure23 and 24.

c. From the projection figures, we can see there are lots of overlapping. This means the selected features are not perfectly separable on the LDA component. Black foam and car sponge shares many similarities on physical properties. For example, they are all likely to be highly deformable. This means that when pressure is applied, they compress, potentially in a non-linear and non-uniform way. The similarity of physical properties would be a significant reason of the relatively poor result of LDA here.

d. To better observe the effectiveness of LDA on object separation, we need to find objects that are distinct in some features. For example, the materials made up of steel vase and flour sack are very different because the former is hard and the latter is relatively soft. So we will anticipate the result of separating the data based on pressure is satisfactory. Indeed, from figure25 and 26, we see a clear boundary between two classes when taking pressure into account. Figure27 is the splitting result in terms of vibration and temperature, which shows overlap, meaning the two classes share more similarities in vibration and temperature than pressure.

4. Section D: Clustering & Classification

4.1. D.1

a. Figure 6 shows our choice of clustering algorithm, which takes Euclidean distance as the distance metric.

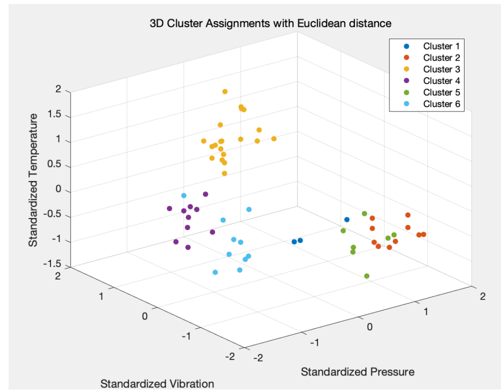

Figure 6. K-means clustering result using Euclidean distance

b. The k-means algorithm using Euclidean distance as distance metric seeks to minimize the within-cluster variance, which is the sum of the squared distances between each point and the centroid of its assigned cluster. From the result plot(figure6), we can say that the clusters correspond well to the actual clusters(figure28) despite some wrong clustering in acrylic, kitchen sponge and steel vase, this may be due to the limitation of Euclidean distance as the algorithm assumes the clusters are spherical and that the features contribute equally to the cluster formation.

c. By changing using Euclidean distance to using Manhattan distance, we observe a different clustering result, as shown in figure29. The Manhattan distance measures the absolute differences along each dimension. From the result, we observe that Manhattan distance tends to produce a better clustering result than Euclidean, especially in clustering kitchen sponge. This may be due to the property of Manhattan distance that all dimensions now contribute linearly to the total distance, meaning large differences in one dimension won't dominate the distance measure as much as they would with Euclidean distance.

4.2. D.2

a. The number of trees is chosen to be 800. The reason is that we need to balance between model performance and computational efficiency. We will observe the accuracy of 800 trees are very good, due to the fact that increasing the number of trees helps stabilizing the model's predictions by reducing variance. On the other hand, we do not want the training to consumes too much resource, so we have empirically chosen 800 trees.

b. The two generated decision trees are shown in figure30 and figure31.

c. A confusion matrix is a tool often used in classification to visualize the performance of an algorithm. Each row of the matrix represents the instances in an actual class, while each column represents the instances in a predicted class. From the confusion matrix(32), we can see most identification lies in the diagonal of the matrix, meaning the accuracy is very satisfactory despite some misclassification. The model performs poorest on classifying label 5, as it only achieves a 50% accuracy(2 correct 2 incorrect.) Overall, there are 20 correct predictions out of 24 total predictions, so the accuracy is 83.3%.

d. If the objects have inherently similar properties, the features extracted may not be sufficiently distinctive, leading to misclassifications. PCA can reveal if there is overlap in the clusters of the data, which may explain why misclassification happen as classifier struggles to find a clear boundary in overlapping clusters. Even though PCA reduces dimensionality by focusing features that capture the most variance, it may mislead the classification. The reasons are followed:

1. If the principal components that capture the most variance in the data do not correspond to the features that best discriminate between certain classes, then PCA may not be helpful.

2. PCA is a linear technique and may not capture nonlinear relationships between features. If the distinction between object properties is nonlinear, PCA might not be effective.

Overall, PCA helps separate classes by capturing the most relevant variance for classification. However, it may oversimplify the data by projecting it onto a space that doesn't capture the nuances necessary for classification.

5. Section E: Conclusion

a. In general, pattern recognition techniques including PCA, LDA, clustering and bagging have significantly advanced our ability to analyze and interpret complex datasets. PCA and LDA both reduce dimensions of the input features while PCA focuses on identifying the most relevant features that contribute to the variance in the dataset and LDA focuses on optimizing class separability. Clustering is an unsupervised algorithm aiming to find dimilarity between data points, and bagging helps classifying new unlabeled data samples. In conclusion, these techniques help us in simplifying complex data, identifying underlying patterns and improving the performance of predictive models.

b. It is possible to distinguish objects only using touch. As shown in previous experiments, we can leverage different properties in the sense of touch, such as temperature, vibration, electrode and pressure. These already built up complex feature space of different objects. However, the effectiveness may decrease for objects that share similar properties(e.g. black foam and kitchen sponge in previous experiment.) We have confirmed that by using techniques such as LDA and PCA and clustering, we can interpret samples with ease. If the data is abundant, we can further leverage complex machine learning models to help learning the data and classifying.

c. From previous observation, vibration and temperature are better properties for discrimination. Impedance measurement is helpful in section D, but it needs 19 nodes which are costly. We should prioritize these vibration and temperature when considering a cheaper tactile sensor with fewer sensing capabilities as this ensures that the most discriminative features can be captured. This is consistent with the principle of maximizing information yield from the most informative and relevant sensing modalities, thus ensuring the effectiveness of tactile sensing systems in our experiment.

d. We can try to do pattern recognition with dynamic time wrapping which mentioned in the lecture. It is used to measure similarity between two temporal sequences varies in speed. By warping two time series trajectories to try and align them, comparing a new sequence of sensor data against a library of known patterns to identify the closest match(with Euclidean distance for example), DTW is very useful before performing clustering, classification and good for sensor data analysis.

DTW can accurately match sequences with temporal variations and is robust to time shifts and distortions between sequences, ensuring reliable pattern recognition. But on the other hand, with its computationally intensive, there are limitations with its real time application or on large dataset.

6. Appendix

Our GitHub repository is available at `https://github.com/ZihaoX3/CVPR`.

The repository contains the following files organized by section:

Section A:

- `A_1.m` - Question A.1: View the time series sensor data for PVT and Electrodes.
- `A_2.m` - Question A.2: For finger F0, sample PVT and Electrode data for each object and trail.
- `A_3.m` - Question A.3: 3D scatter plot of PVT data.
- `F0_PVT.mat` - Collected PVT data at $t = 50$.
- `F0_Electrodes.mat` - Collected Electrode data at $t = 50$.

Section B:

- `B_1.m` - Question B.1
  - Report covariance matrix, eigenvalues, eigenvectors of PVT data.
  - Plot standardized data with PCs.
  - Plot 2D dimension data.
  - Show how PVT data is distributed across all PCs.
- `B_2.m` - Question B.2
  - Determine PCs of electrode data using PCA.
  - Report variance of each PC using Scree plot.
  - Plot electrode data using PCs.

Section C:

- `C_1.m` and `C_1projection.m` - Question C.1.a: Use LDA to split training data, visualize them in 2D and 1D.
- `C_2.m` - Question C.1.b: Use LDA to three dimensional PVT data.
- `C_4.m` - Question C.1.d: LDA analysis of steel vase and flour sack.

Section D:

- `D_1.m` and `D_1_2.m` - Question D.1: Clustering algorithm with Euclidean and Manhattan distance metrics.
- `standardlized3D.m` - Real life object similarities used for comparison.
- `D_2.m` - Question D.2: Bagging algorithm to electrode.

Additional plots:

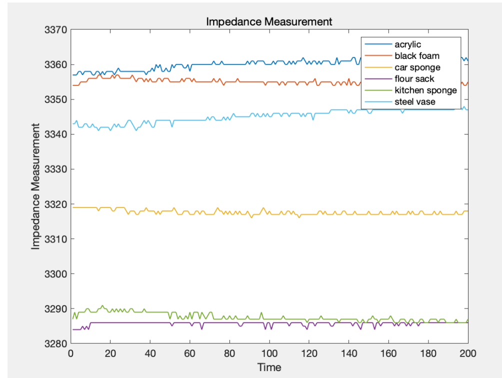

Figure 7. Measurement of one impedance of all objects from trial 3

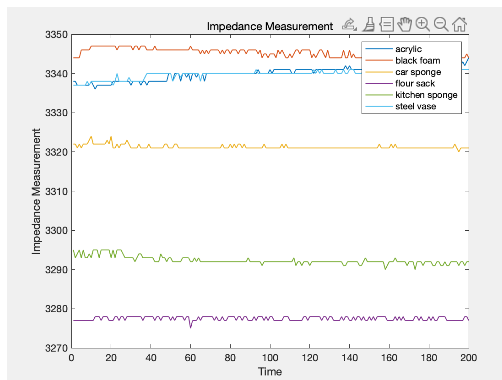

Figure 8. Measurement of one impedance of all objects from trial 8

Figure 9. Measurement of vibration of all objects from trial 3

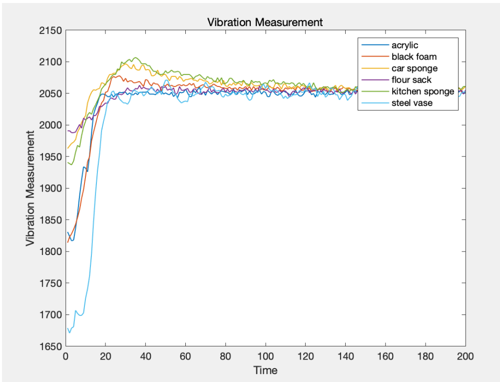

Figure 10. Measurement of vibration of all objects from trial 8

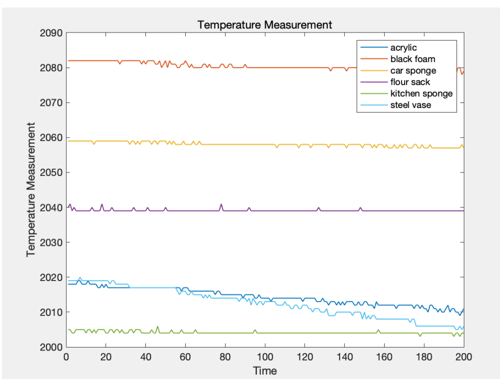

Figure 11. Measurement of temperature of all objects from trial 3

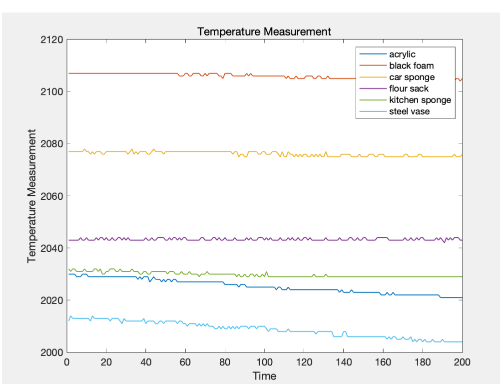

Figure 12. Measurement of temperature of all objects from trial 8

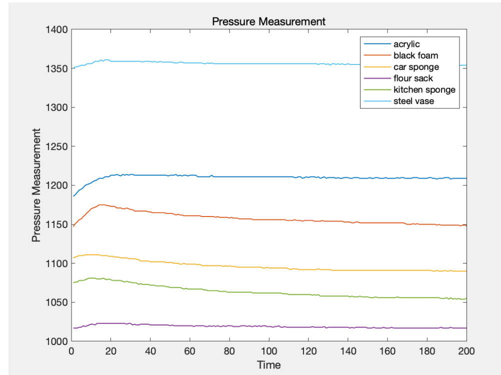

Figure 13. Measurement of pressure of all objects from trial 3

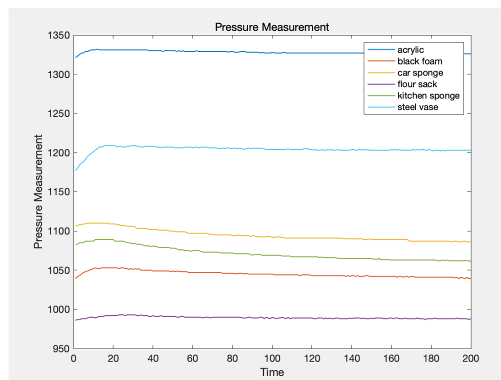

Figure 14. Measurement of pressure of all objects from trial 8

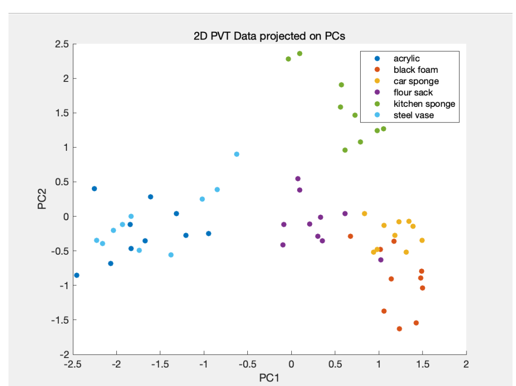

Figure 15. Reduced standardised PVT data in 2D

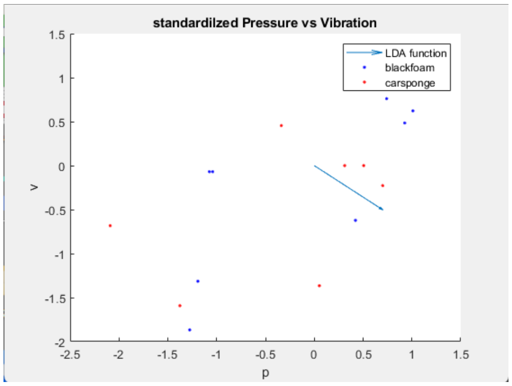

Figure 16. LDA separation Pressure vs. Vibration

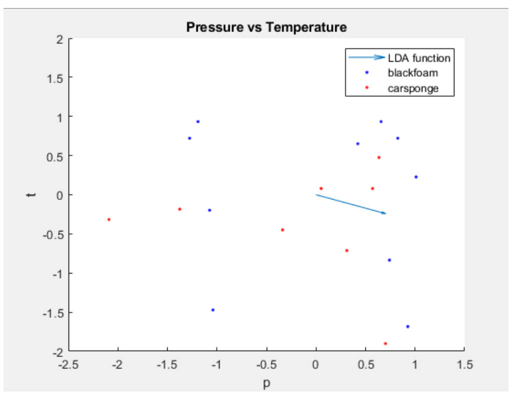

Figure 17. LDA separation Pressure vs. Temperature

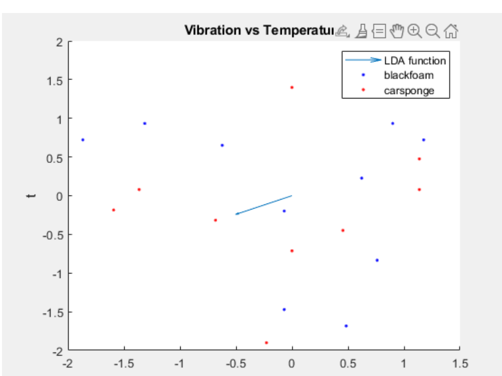

Figure 18. LDA separation Vibration vs. Temperature

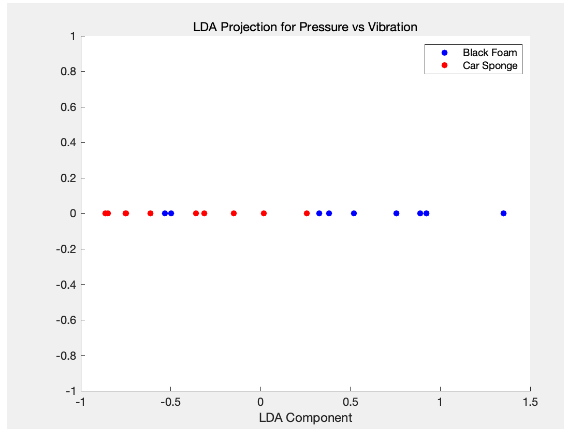

Figure 19. Projection LDA on 1D Pressure vs. Vibration

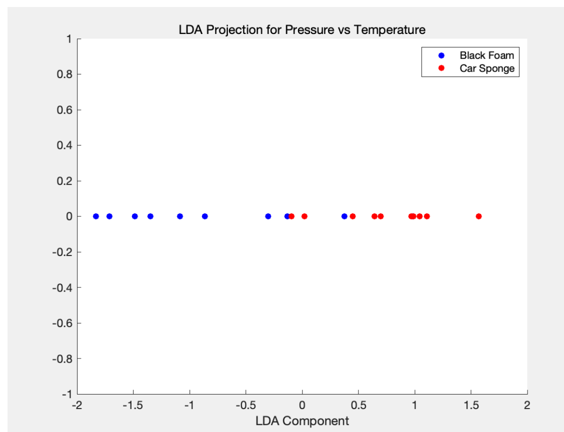

Figure 20. Projection LDA on 1D Pressure vs. Temperature

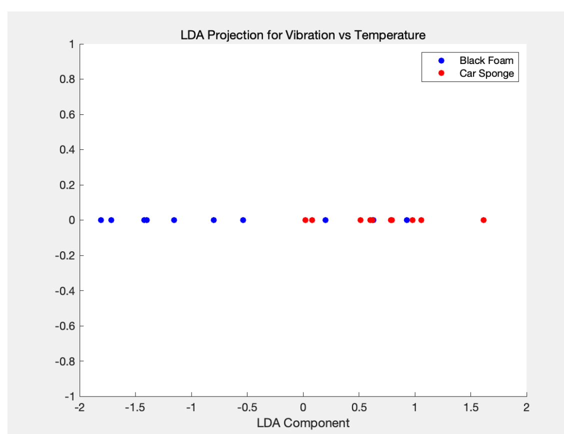

Figure 21. Projection LDA on 1D Vibration vs. Temperature

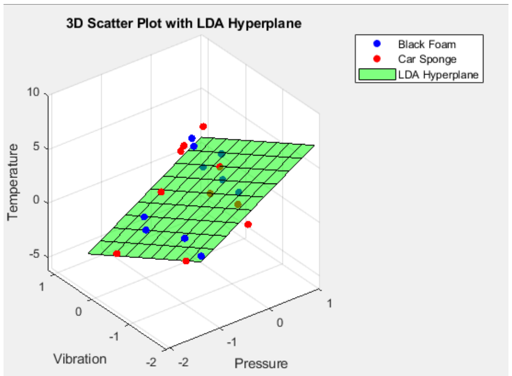

Figure 22. 3D Scatter Plot with LDA Hyperplane

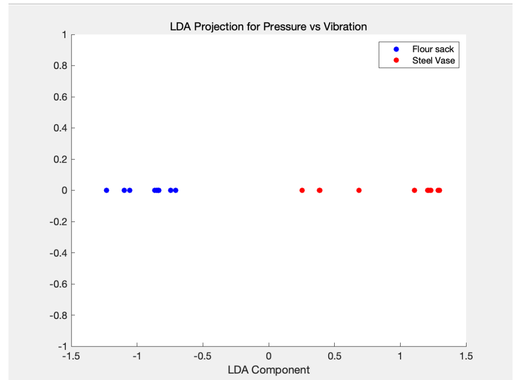

Figure 25. Flour sack and Steel vase LDA Projection for Pressure vs Vibration

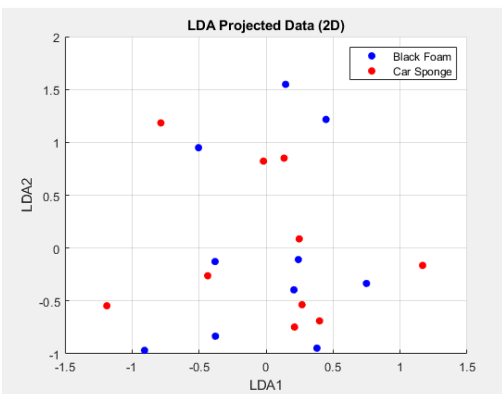

Figure 23. 2D Scatter Plot

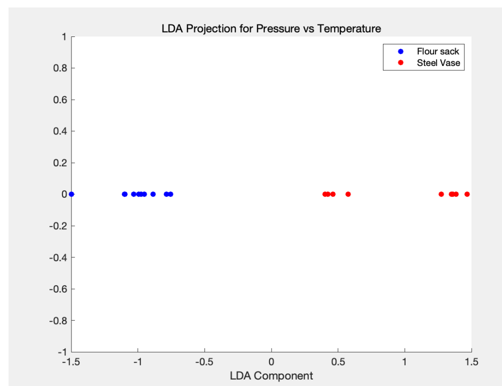

Figure 26. Flour sack and Steel vase LDA Projection for Pressure vs Temperature

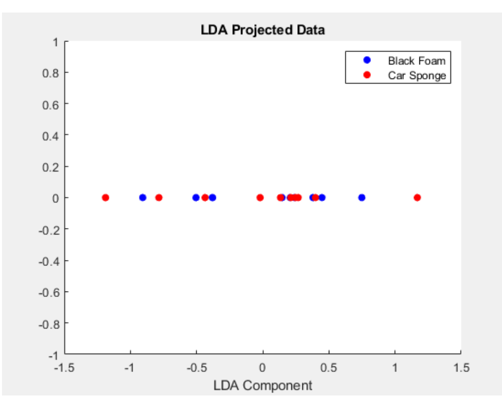

Figure 24. 1D Scatter Plot

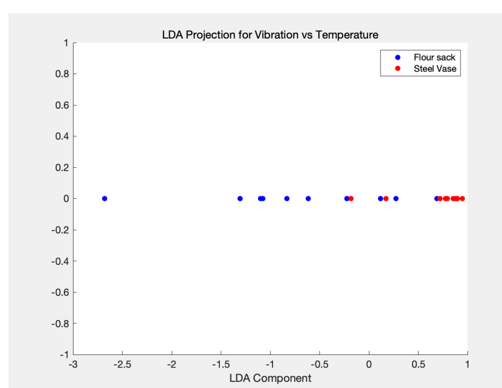

Figure 27. Flour sack and Steel vase LDA Projection for Vibration vs Temperature

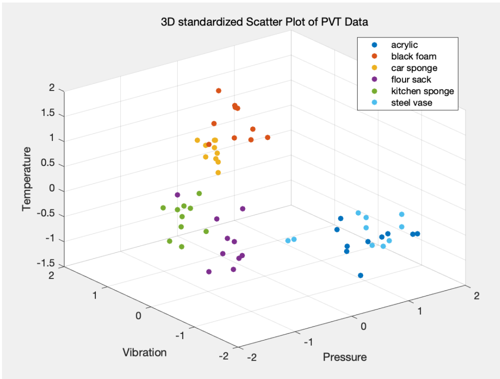

Figure 28. Real-life objects standardized scatter plot

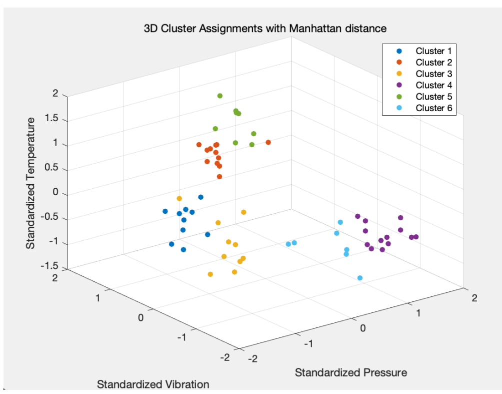

Figure 29. K-means clustering result using Mahattan distance

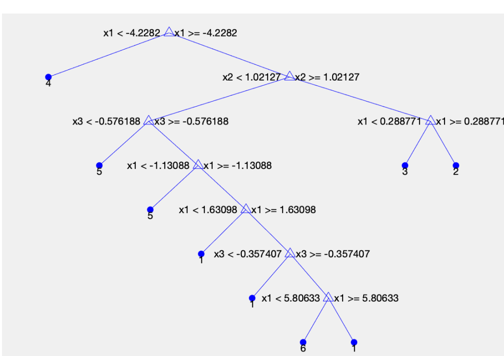

Figure 30. First decision tree by applying bagging to the electrode data that was processed with PCA

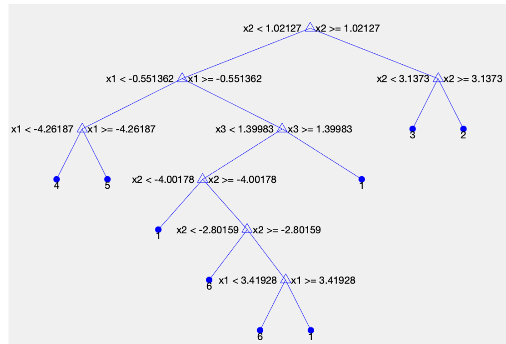

Figure 31. Second decision tree by applying bagging to the electrode data that was processed with PCA

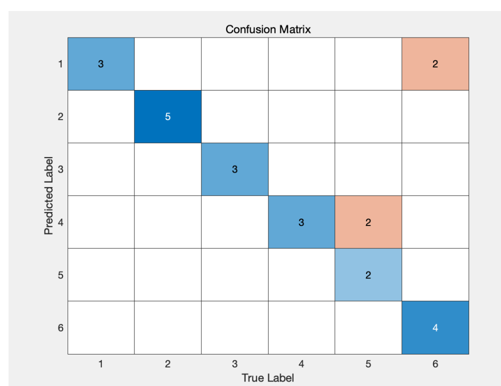

Figure 32. Confusion matrix of the trained model from bagging