ME278 - A Practical Introduction to Data Analysis
Assignment 5

N Surya (22938)

November 28, 2024

Chapter 1

Dynamic Mode Decomposition

1.1 Linear dynamics

How does matrix A compare with original linear dynamics matrix?

The true matrix is:

$$\dot{\mathbf{x}}(t) = \mathbf{A}_{true}\mathbf{x}(t)$$

$$\mathbf{A}_{true} = \begin{bmatrix} 0 & 1 \\ \dfrac{-g}{l} & 0 \end{bmatrix}$$

Here, the DMD matrix is expected to approximate:

$$\mathbf{A}_{DMD} \approx e^{\mathbf{A}_{true}\Delta t}$$

So, in order to compare $\mathbf{A}_{DMD}$ and $\mathbf{A}_{true}$, we verify the 2-norm of $\mathbf{A}_{DMD} - e^{\mathbf{A}_{true}\Delta t}$, which was found to be **2.0137**.

Comparison of DMD and original solution:

From the plot, it can be seen that the DMD approximation serves very well for the small angle approximated linear solution of the pendulum problem.

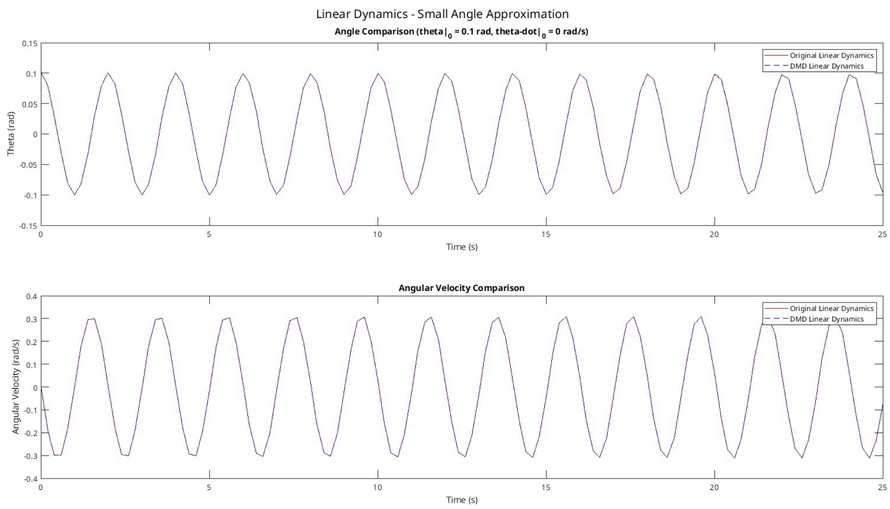

Figure 1.1: DMD (linear) vs linear solution for $t \in [0, 25]$

1.2 Nonlinear dynamics

Comparison of DMD for nonlinear dynamics and original solution:

It can also be seen that the DMD also captures nonlinear dynamics to a decent extend as can be seen from the low RMSE values.

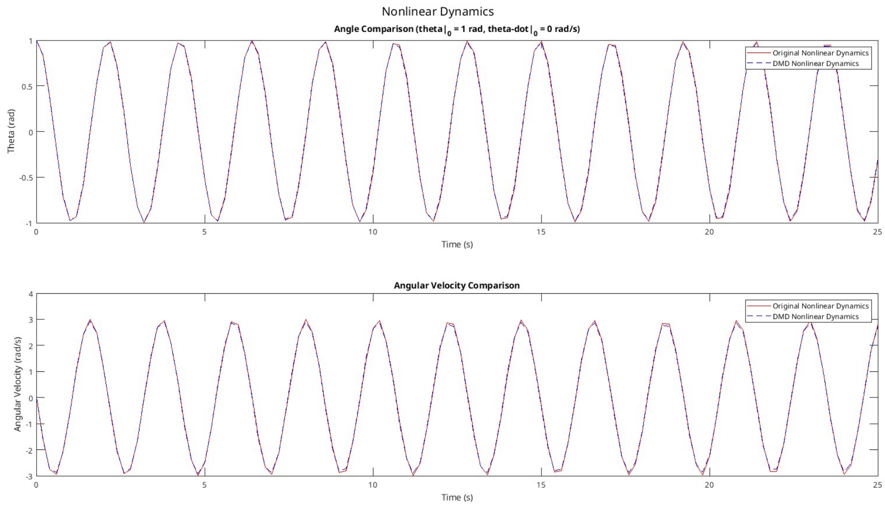

Figure 1.2: DMD (Nonlinear) vs nonlinear solution for $t \in [0, 25]$

Parameter	**RMSE**
Angular displacement	0.014892
Angular Velocity	0.05459

Table 1.1: Root Mean Squared Errors for θ and $\dot{\theta}$

Comparison of DMD for nonlinear dynamics and linear solution:

It is immediately evident that the linear solution diverges from the DMD solution (which is correct!) with time.

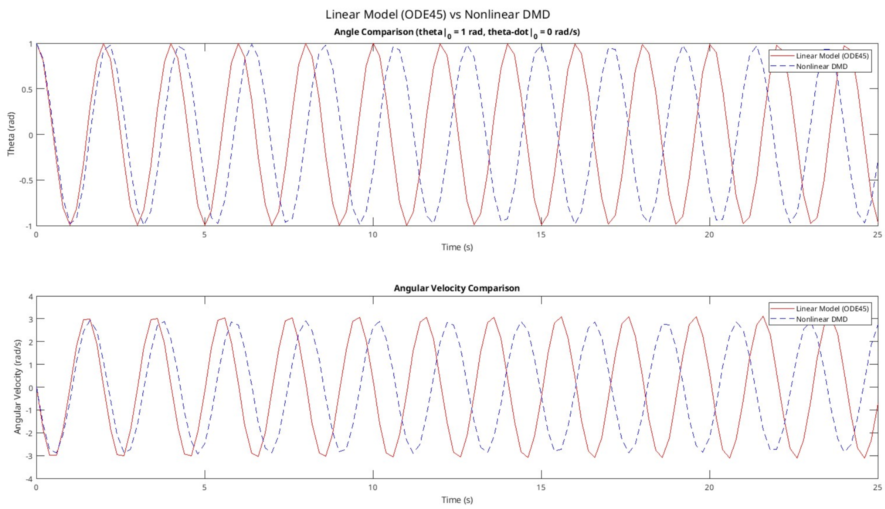

Figure 1.3: DMD (Nonlinear) vs linear solution for $t \in [0, 25]$

Chapter 2

Regression

2.1 $2nd$ degree polynomial regression

Polynomial coefficients:

$$y = 0.0607 + 1.9707 \cdot x_1 - 0.3066 \cdot x_2 + 0.0704 \cdot x_1^2 + 900.3247 \cdot x_2^2 - 4.1131 \cdot x_1 \cdot x_2$$

RMSE values:

Error on	Mean Squared Error
Test Data	0.9907
Training Data	1.0036

Since the MSE is higher on the training data, it indicates that the model might be underfitting.

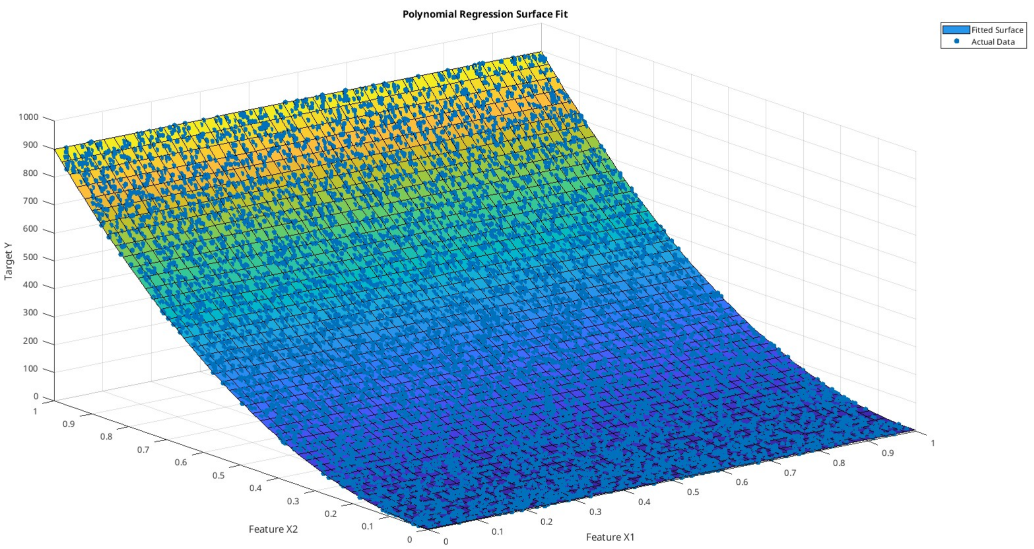

Figure 2.1: 2nd degree polynomial regression - surface

2.2 Regularization with 10^{th} degree polynomial regression

Mean Squared Error	Value
No regularization (Train)	63.9284
No regularization (Test)	116.9684
Ridge (Train)	63.9285
Ridge (Test)	116.9991
Lasso (Train)	88.7242
Lasso (Test)	128.0981

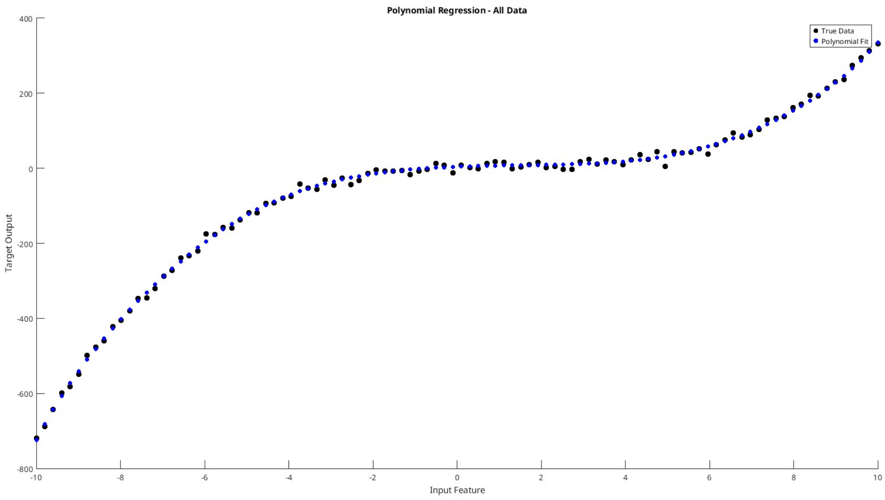

Figure 2.2: Unregularized polynomial regression

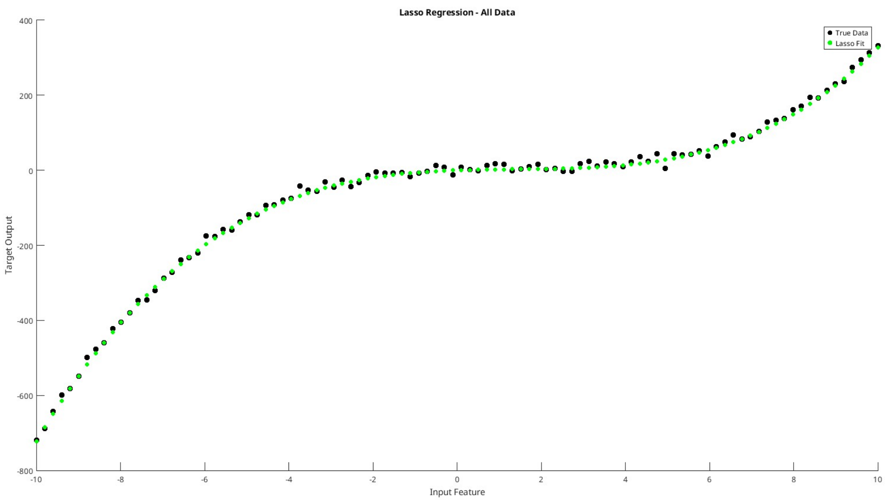

Figure 2.3: Lasso regression

In all three cases, it can be seen that the training MSE is less than the test MSE. This indicates that all the three models may be overfitting. Based on the discrepancy in MSE, we can claim that the Lasso regularization is the closes to aptly fitting the data. Additionally, Ridge regularization seems to have no beneficial effect over the plain polynomial fit.

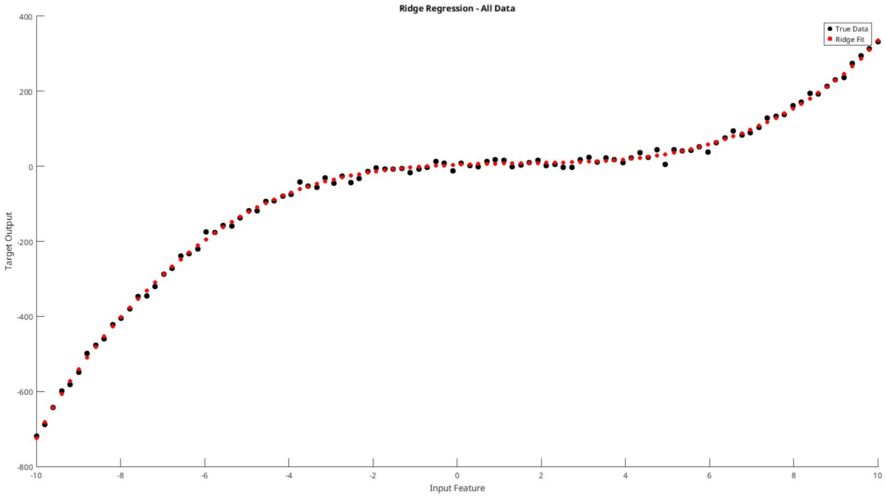

Figure 2.4: Ridge regression

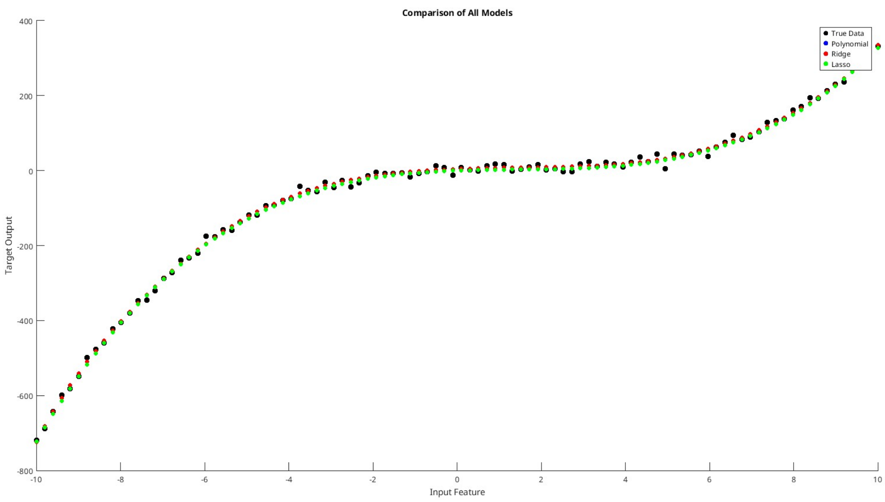

Figure 2.5: All models

Chapter 3

Clustering

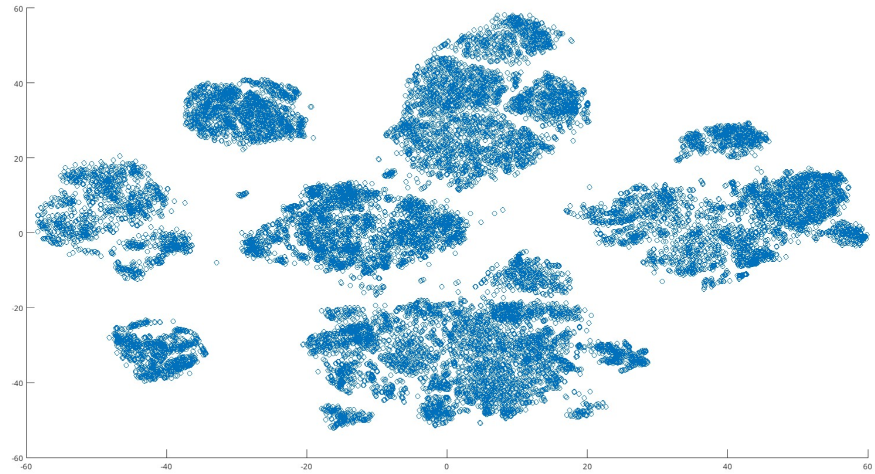

Figure 3.1: Data visualization

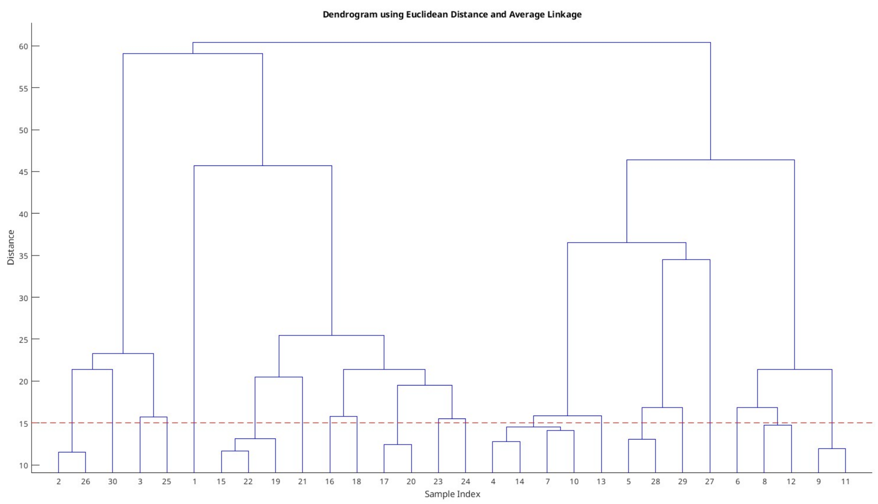

Figure 3.2: Dendrogram

Based on the visualization of the data, it is clear that the clusters show less density and more variance inside, hence average linkage makes more sense. Euclidean distance is also appropriate for the same reason. The cut off is chosen at 15 based on the dendrogram and the data visualization.

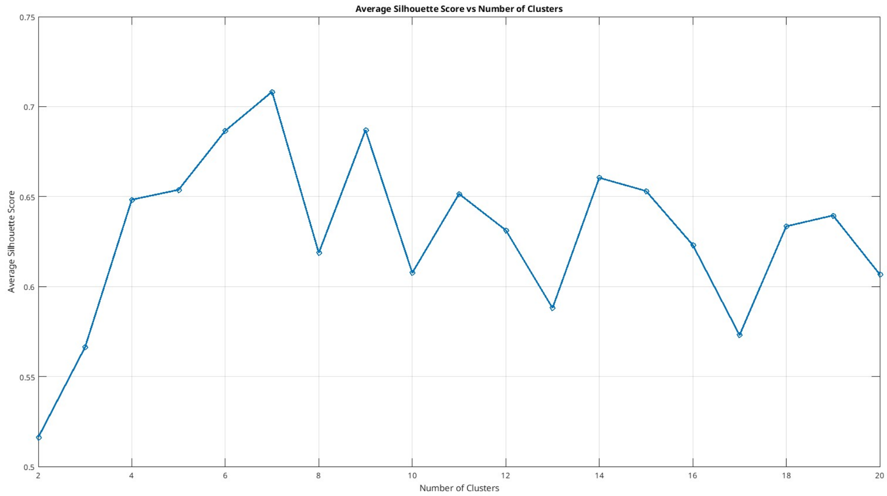

Figure 3.3: Silhoutte score vs number of clusters

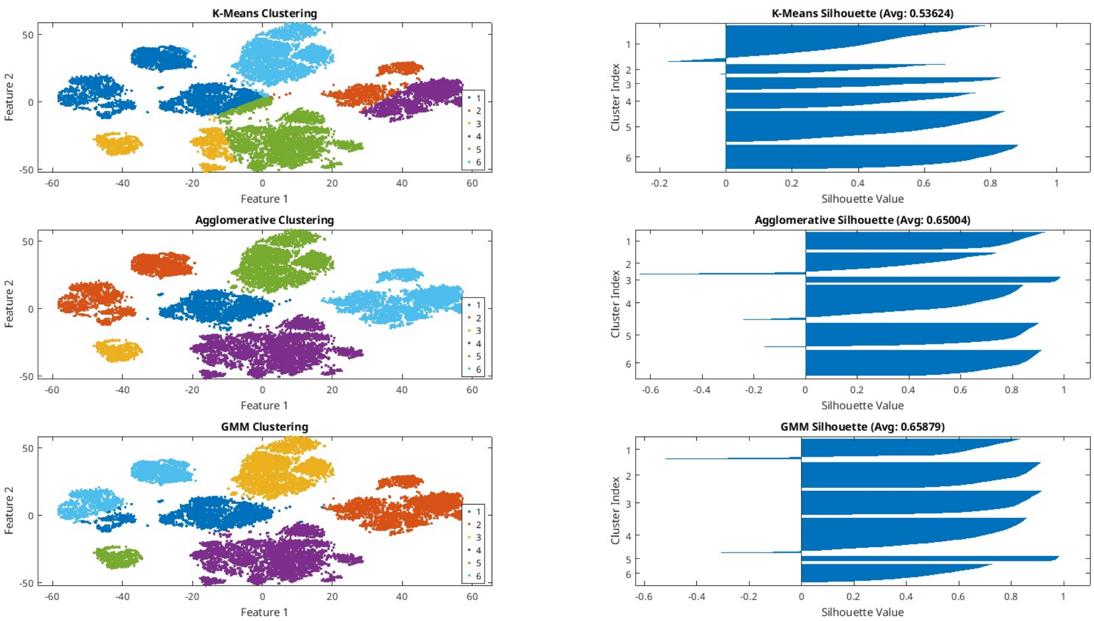

Figure 3.4: All three clustering

Chapter 4

Classification

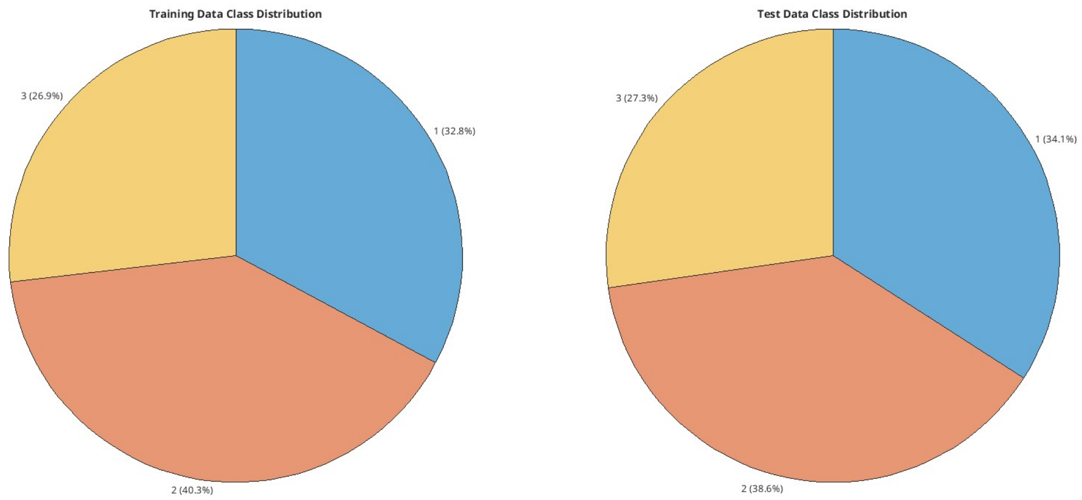

Figure 4.1: Piechart for data split

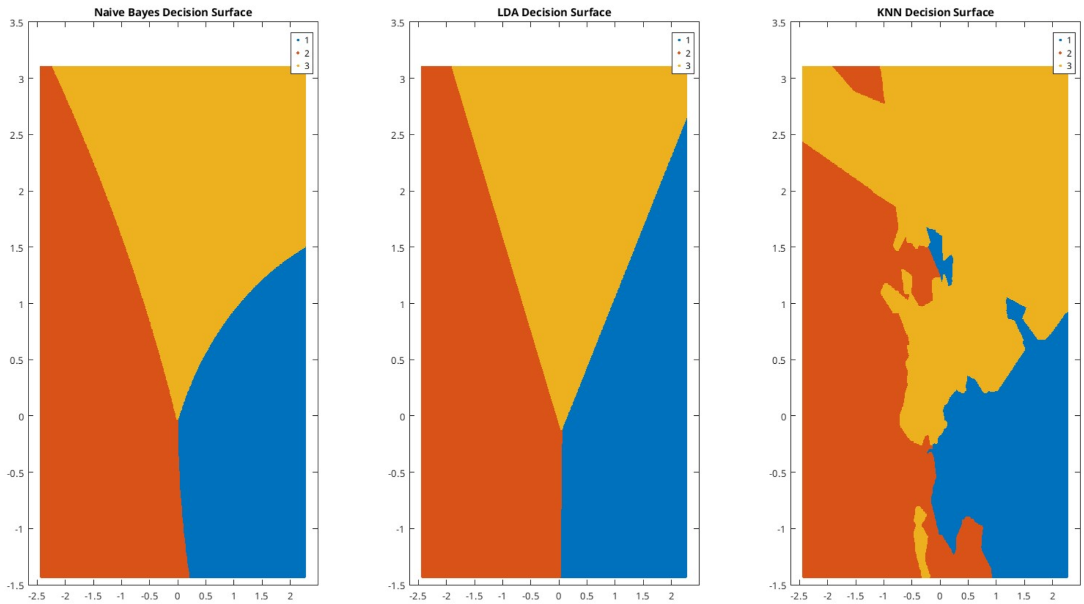

Figure 4.2: Decision surface for all three models

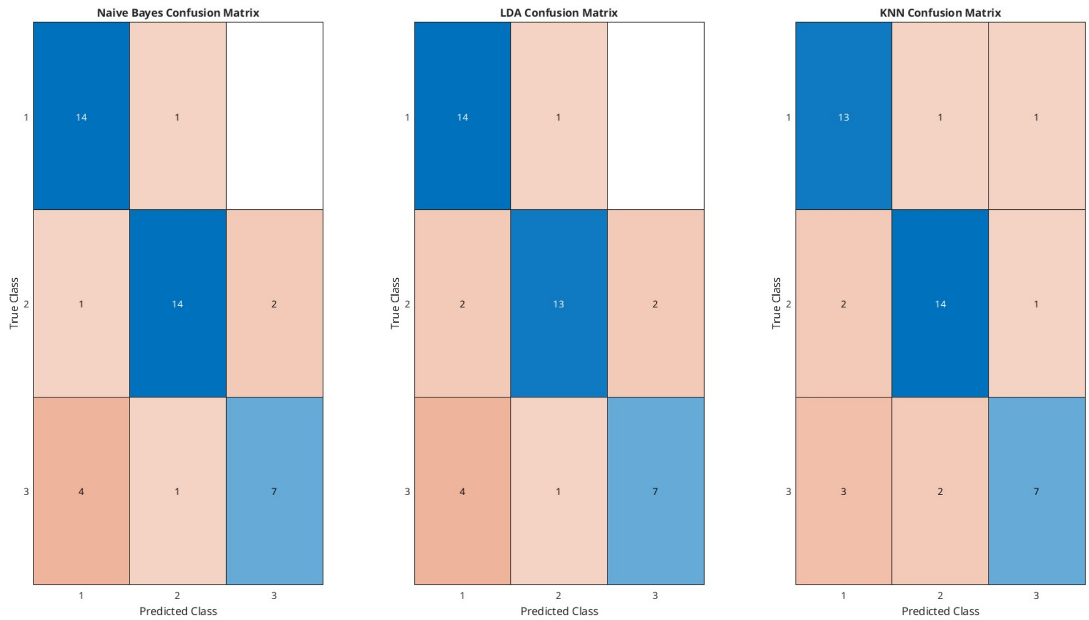

Figure 4.3: Confusion matrix for all three models