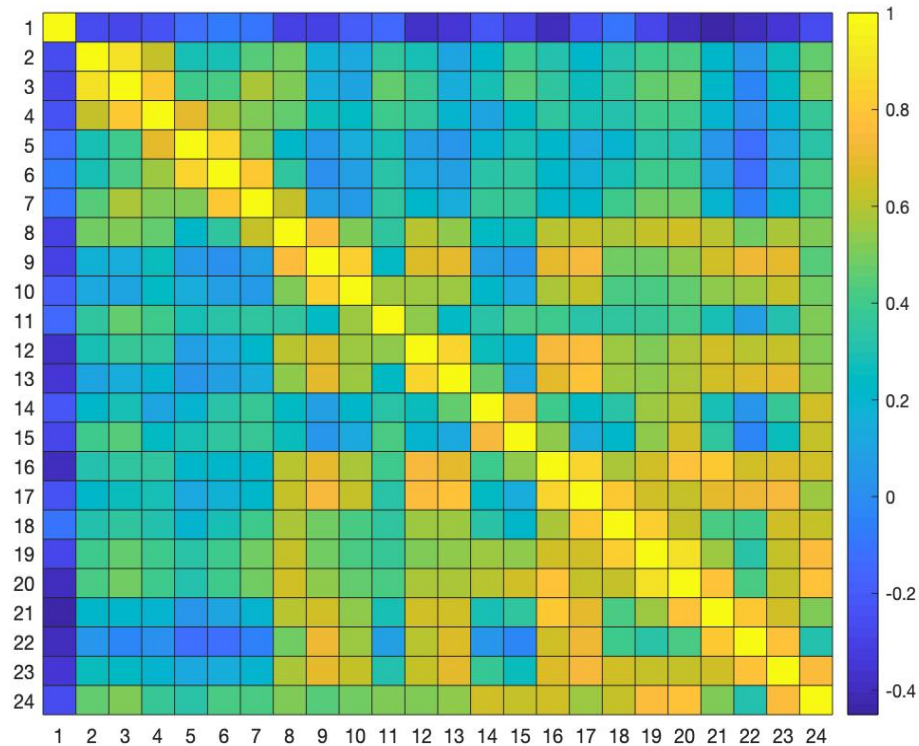


INF2B Report – TASK 1

TASK 1.2

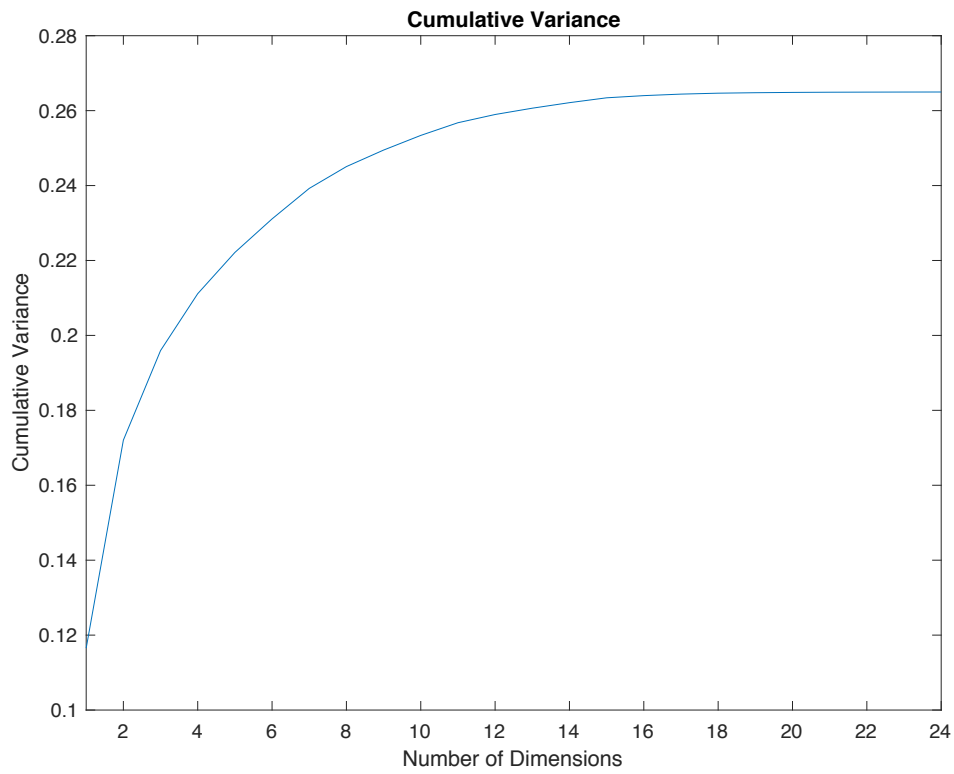
Heatmap indicating Pearson's correlation coefficient between two features



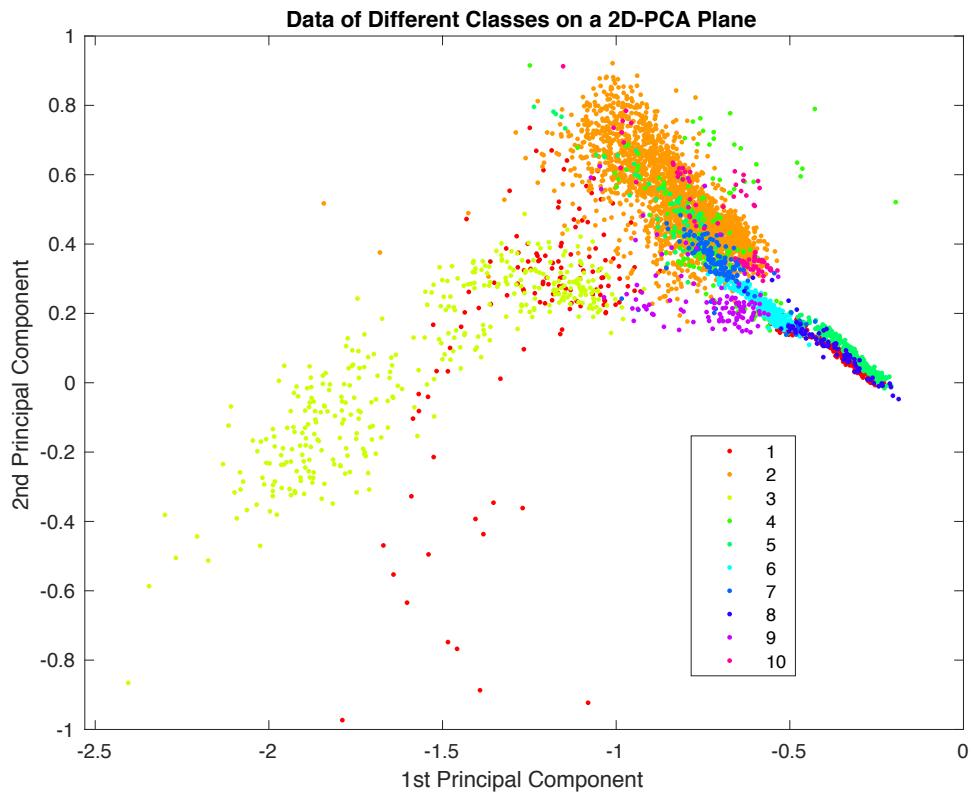
From the graph above, it can be seen that feature 1 is negatively correlated to other features. This means that as value for feature 1 increases, value for features 2 to 24 decrease and vice versa. However, the correlation is relatively weak as the lowest correlation coefficient is around -0.4. This means that as the value for feature 1 increases/decreases, there is a lower likelihood of there being a relationship with features 2 to 24.

Moreover, the boxes colored around the shade of orange/yellow has a strong positive correlation. For instance, as feature 2 increases, there is a high likelihood that feature 3 also increases and vice versa.

TASK 1.3b



TASK 1.3c



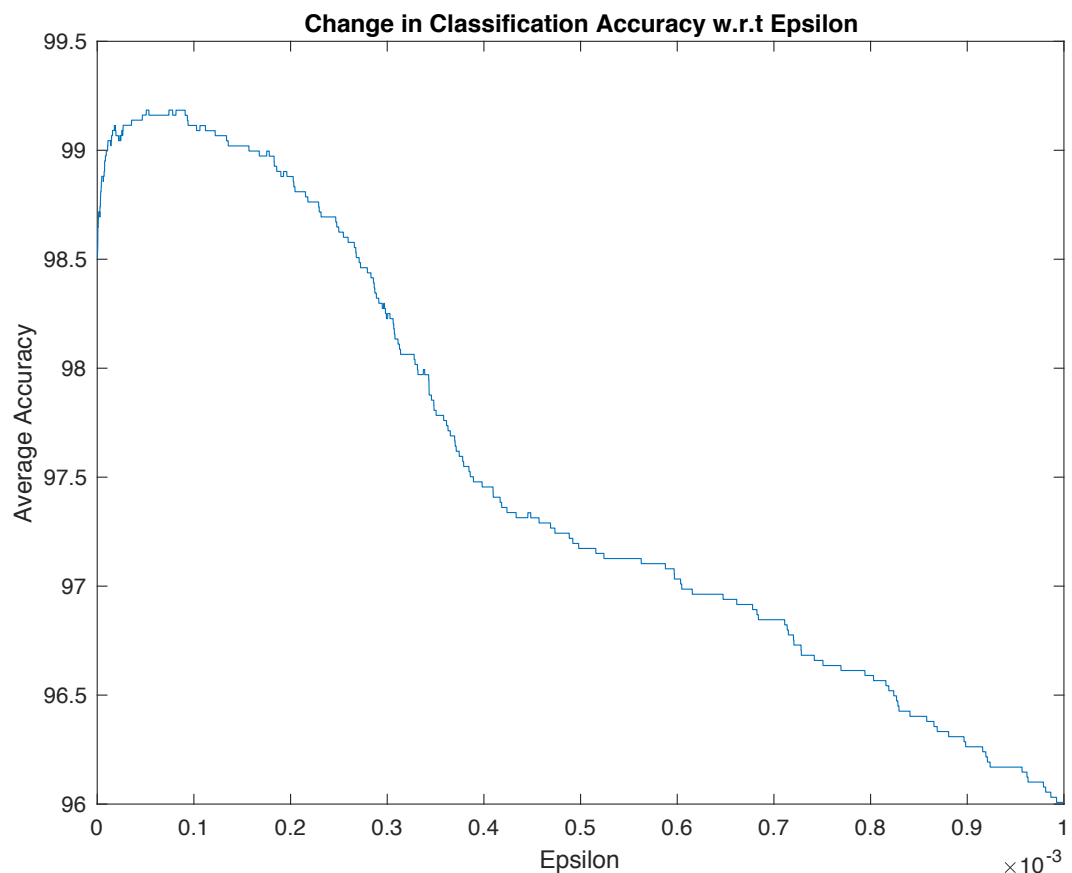
TASK 1.4b

With epsilon = 0.01 and Kfolds = 5, the accuracy is as follows:

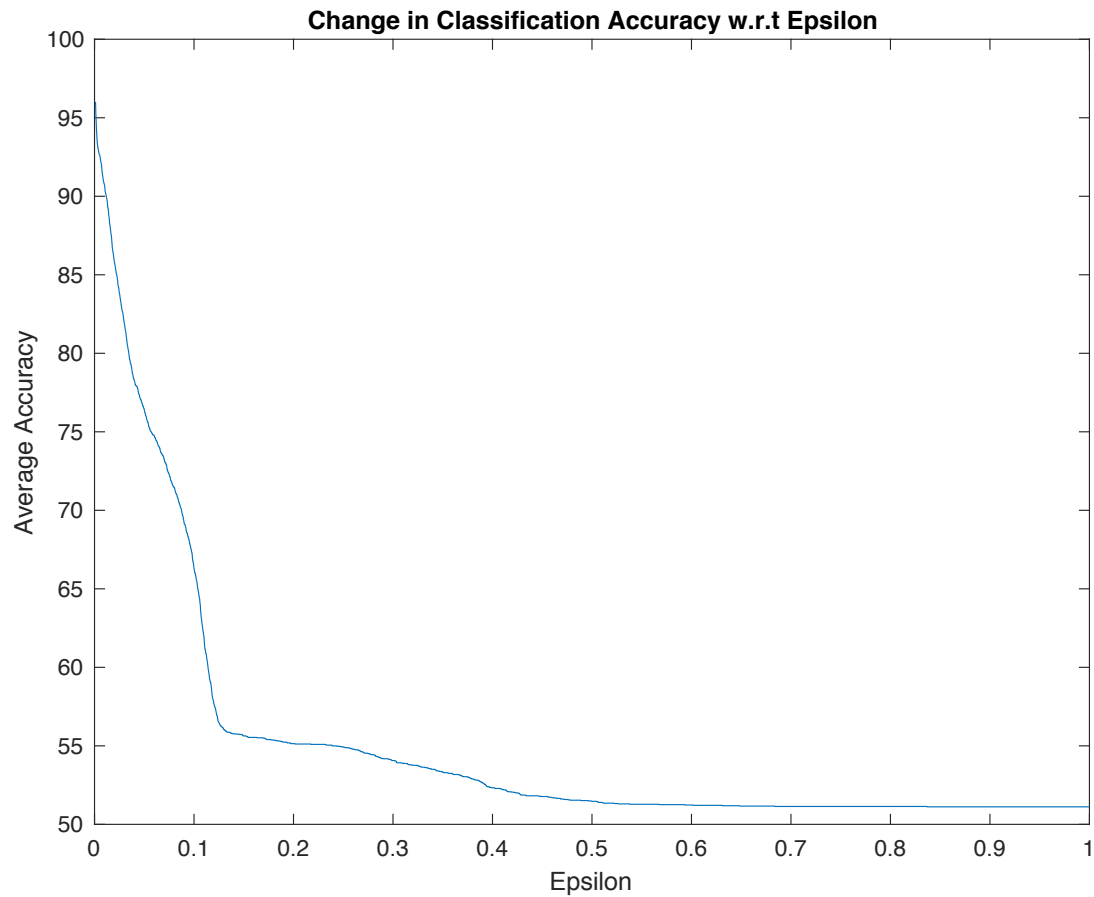
CovKind Fold	1	2	3
1	91.54%	84.02%	89.89%
2	90.25%	83.55%	87.78%
3	91.07%	83.67%	88.48%
4	90.36%	83.78%	88.72%
5	90.40%	80.23%	88.00%
Average Accuracy	90.72%	83.05%	88.57%

TASK 1.5

- With a full covariance matrix and Kfolds equal to 5, the findings for the change in classification accuracy with respect to epsilon is split into two graphs.
- We first observe the change when the interval for epsilon is [0.0000001, 0.001].



- It is found that the epsilon that gives the best accuracy is a number around 0.1×10^{-3} , which is equal to 0.0001; Values of epsilon smaller than that actually reduce the accuracy.
- Next, we observe the change when the interval for epsilon is [0.001, 1].



- It is found that the accuracy decreases relatively drastic from when epsilon is equal to 0.001 up to around 0.1.
- It then decreases more gradually before reaching an accuracy of 51% when epsilon is equal to 1.