Inf2B Coursework 1 Report

Task 1 – Anuran-Call analysis and classification

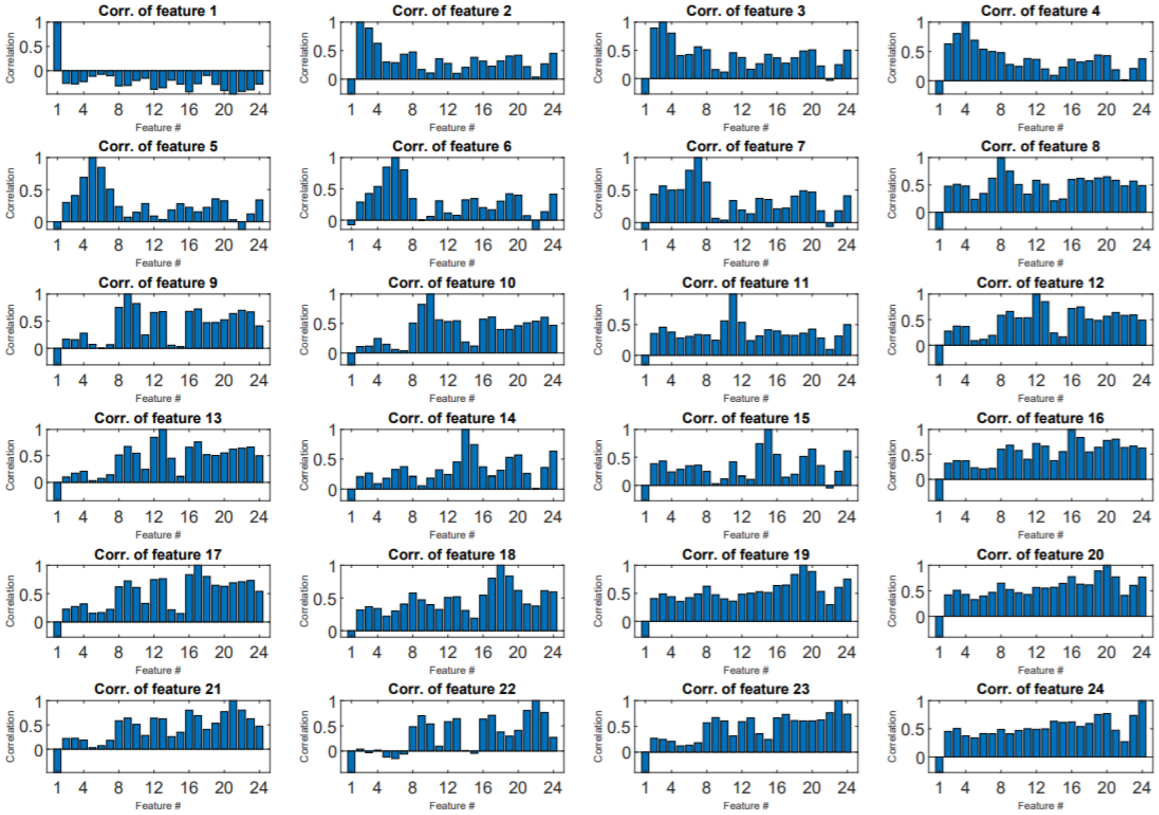
1.2) Findings from the correlation matrix:

Upon analyzing the data within correlation matrix R I found it would be useful to visualize the data in 2 ways:

1. A collection of subplots to show the relationships between all feature vectors
2. A bar graph to show the average correlation value for each feature vector

I used bar graphs to represent both visualizations so it would be easy to recognize the highest/lowest correlations.

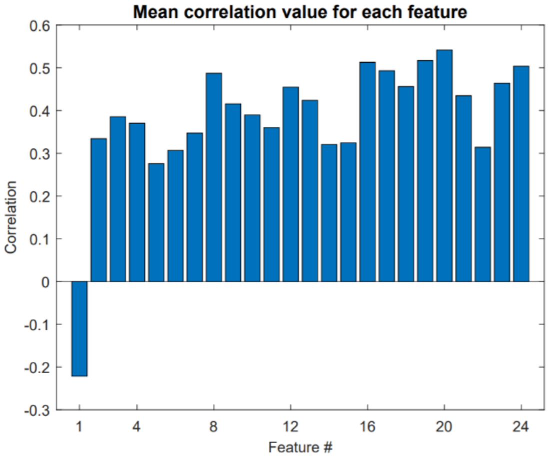
A collection of subplots to show the relationships between all feature vectors in the data set X



Corr. of feature 1, Corr. of feature 2, Corr. of feature 3, Corr. of feature 4, Corr. of feature 5, Corr. of feature 6, Corr. of feature 7, Corr. of feature 8, Corr. of feature 9, Corr. of feature 10, Corr. of feature 11, Corr. of feature 12, Corr. of feature 13, Corr. of feature 14, Corr. of feature 15, Corr. of feature 16, Corr. of feature 17, Corr. of feature 18, Corr. of feature 19, Corr. of feature 20, Corr. of feature 21, Corr. of feature 22, Corr. of feature 23, Corr. of feature 24

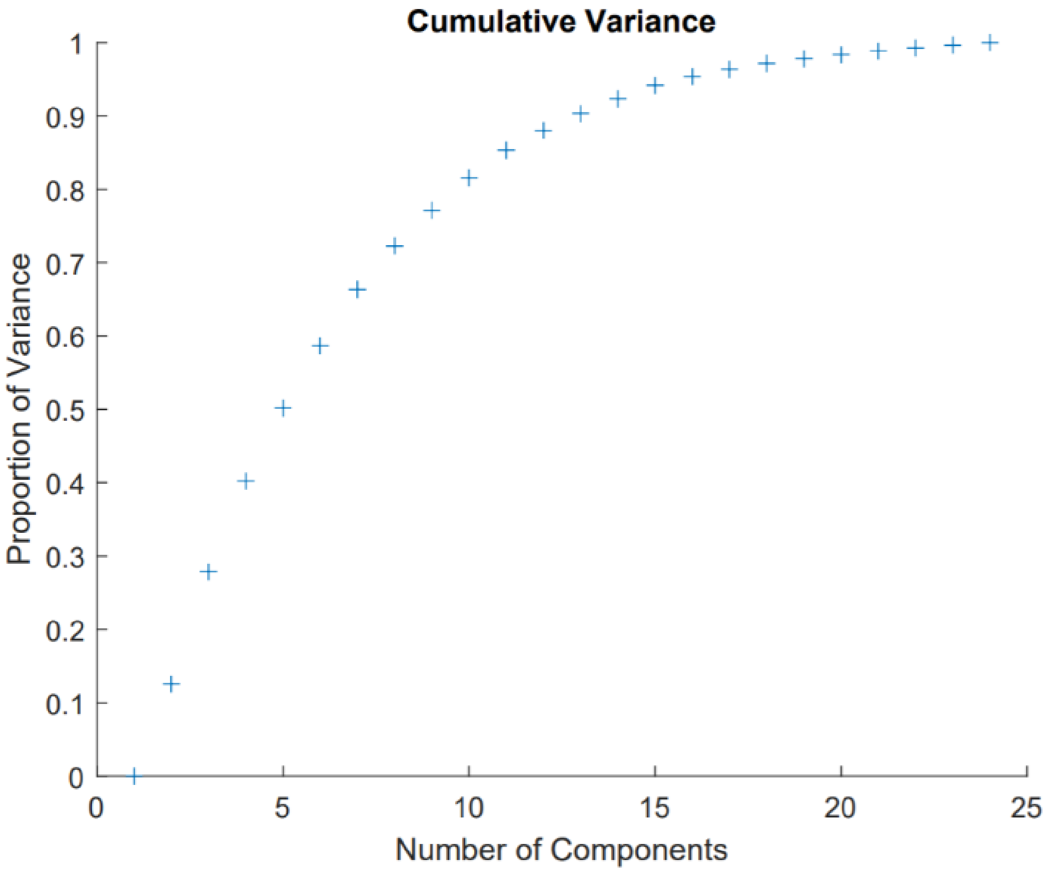
We must note that when analyzing these subplots all bars with correlation 1 represent the correlation between the same feature, and thereby do not represent anything significant.

This visualization of correlation matrix R is convenient for determining the nature of how a given feature is correlated to other features. This can be useful for making predictions about an incomplete sample (does not have data for all features), in which we can predict the value for a given missing feature by using the existing data in the sample with appropriate weightings (weighting for a given feature F is directly proportional to the correlation value between the missing feature and F).

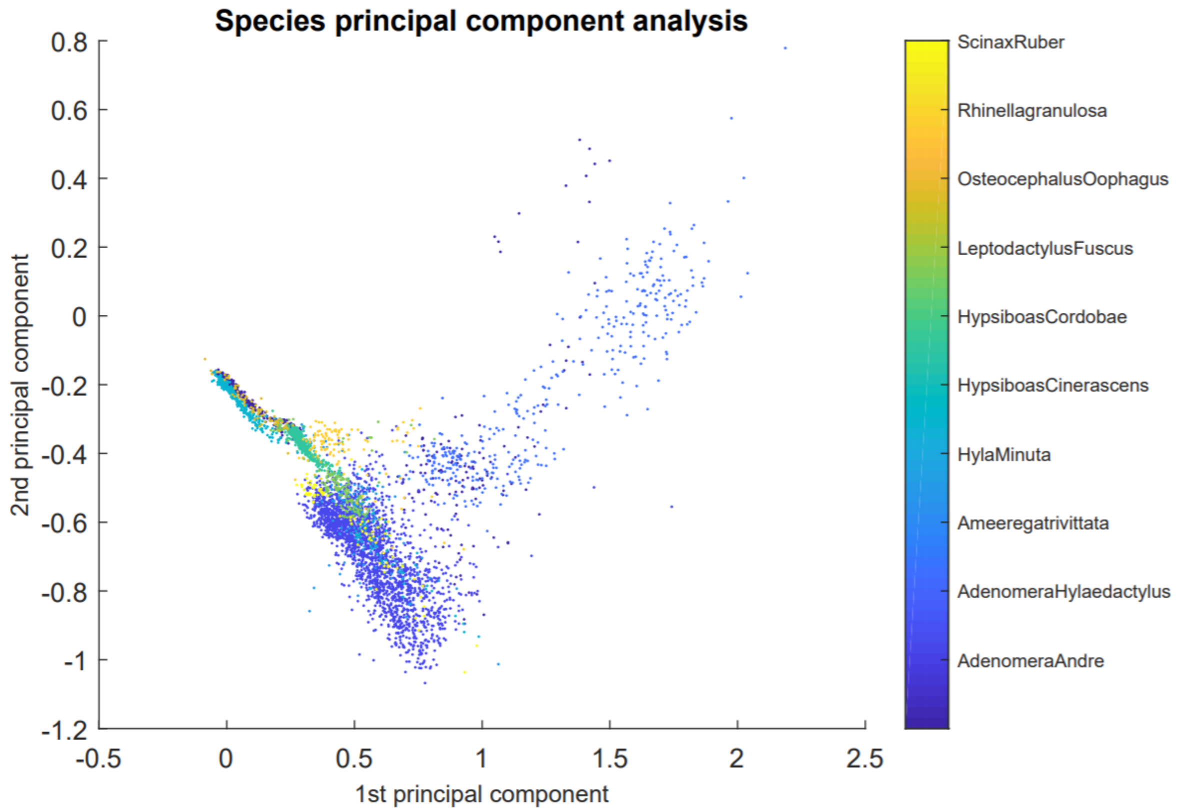


This visualization of correlation matrix R is convenient in order to determine which features are uniquely correlated. This is particularly useful when predicting feature values for an incomplete sample as it indicates the importance of normalization for each of the respective correlation values when equating weights.

1.3) b) Graph of cumulative variance



1.3) c) Plotting of data on 2D-PCA plane



1.4) b) Accuracy for CovKind = 1,2,3

	Overall accuracy for a given class										
CovKind	**Class 1**	**Class 2**	**Class 3**	**Class 4**	**Class 5**	**Class 6**	**Class 7**	**Class 8**	**Class 9**	**Class 10**	**Average**
1	0.3971	0.9972	0.9787	0.6035	0.8276	0.9919	0.8661	1	0.9920	0.9867	**0.8641**
2	0.2853	0.9823	0.7184	0.4000	0.0269	0.9899	0.8596	1	0.9444	0.5000	**0.6707**
3	0.3765	0.9963	0.9230	0.7139	0.8785	0.9879	0.8411	0.9765	0.9634	0.9267	**0.8584**
Average	**0.3529**	**0.9919**	**0.8733**	**0.5725**	**0.5777**	**0.9899**	**0.8556**	**0.9922**	**0.9666**	**0.8044**	**0.7977**

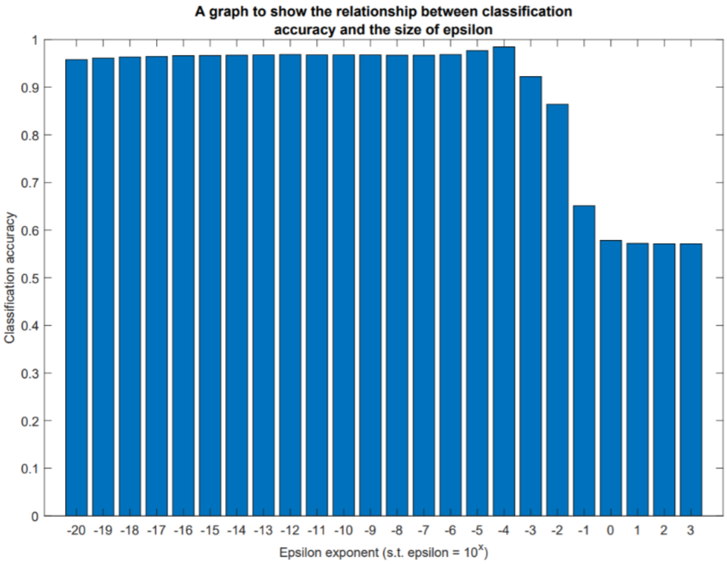
	Overall accuracy for a given partition					
CovKind	**Partition 1**	**Partition 2**	**Partition 3**	**Partition 4**	**Partition 5**	**Average**
1	0.9122	0.9170	0.9063	0.9229	0.9146	**0.9146**
2	0.7960	0.7995	0.7900	0.8233	0.7947	**0.8007**
3	0.9146	0.9110	0.8992	0.9336	0.8997	**0.9116**
Average	**0.8743**	**0.8758**	**0.8652**	**0.8932**	**0.8697**	**0.8756**

1.5) Classification accuracy VS epsilon

Since we are making predictions on our test samples using a multivariate Gaussian classifier, we must be wary about the stability of our statistical measures. They may become unstable due to anomalies in the data set, particularly in the implementation of the log likelihood equation due to its use of inverse covariance matrices: in which $\sum^{-1}$ can become unstable when $|\sum|$ is small.

In order to minimize test error and make more accurate predictions we need a statistical learning method that simultaneously achieves low variance and low bias. In this case we used regularisation by adding a small positive number (epsilon) to the diagonal elements (which represent the variances) of the covariance matrix.

Upon this regularisation we must choose a value of epsilon that promotes optimal accuracy by balancing the bias-variance trade-off. In order to do this, we can test how different sizes of epsilon affect the overall classification accuracy.



We can see that for values of epsilon greater than or equal to 1 ($10^0 \leq \epsilon \leq 10^3$) that the classification accuracy lies below 60%. We can attribute this to prediction modelling with a high variance.

However, this classification accuracy exponentially increases when values of epsilon are less than 1. This is particularly evident in the range $10^{-17} \leq \epsilon \leq 10^{-3}$ in which the classification accuracy is greater than 90%.

The optimal value of epsilon for this data model is that of $\epsilon = 10^{-4}$ with a classification accuracy of 98.45%. This suggests an optimal bias-variance ratio.

Overall, we can see that these classification accuracies continue to decrease on either side of the optimal value $\epsilon = 10^{-4}$ which tells us the nature of the associated bias-variance tradeoff.