**MUSI 6201**

**Assignment 2 – Genre Classification**

Imankalyan Mukherjee, Govinda Ram Pingali

**I) Scatter plots for pairs of audio features**

Color Legend:

Blue: Classical

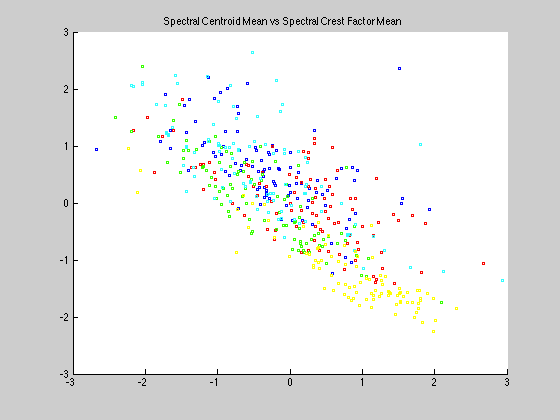
Red: Hip Hop

Green: Country

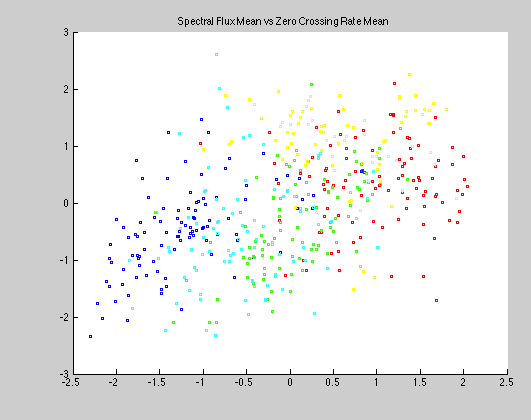
Yellow: Jazz

Cyan: Metal

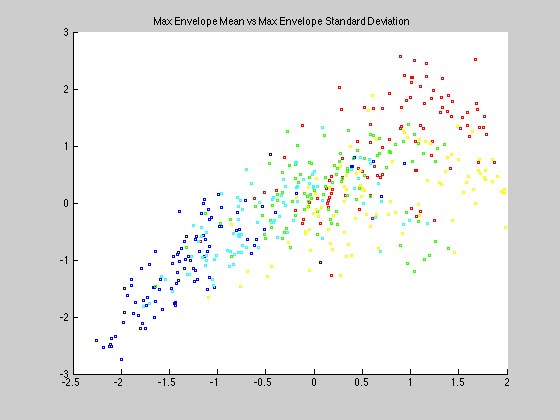
i) Spectral Centroid Mean v/s Spectral Crest Factor Mean



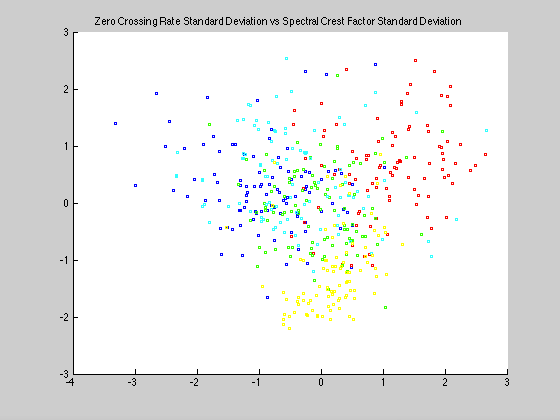
iI) Spectral Flux Mean v/s Zero Crossing Rate Mean



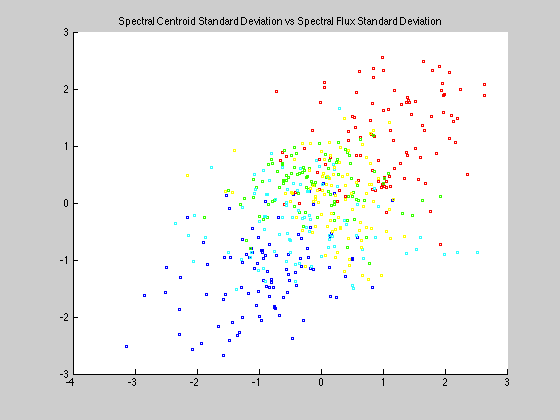
iiI) Max Envelope Mean v/s Max Envelope Standard Deviation



iv) Zero Crossing Rate Standard Deviation v/s Spectral Crest Factor Standard Deviation



v) Spectral Centroid Standard Deviation v/s Spectral Flux Standard Deviation



**II) Ranking Features Based on Accuracy**

The Accuracy Matrix for individual features after 10-fold validation looks like as follows

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **Features** | | | | | | | | | | |
| **Folds** |  | **F1** | **F2** | **F3** | **F4** | **F5** | **F6** | **F7** | **F8** | **F9** | **F10** |
| **1** | 28 | 26 | 20 | 32 | 30 | 18 | 34 | 32 | 24 | 30 |
| **2** | 44 | 32 | 8 | 36 | 34 | 24 | 32 | 24 | 32 | 40 |
| **3** | 28 | 22 | 18 | 36 | 50 | 42 | 20 | 18 | 32 | 56 |
| **4** | 30 | 28 | 24 | 42 | 38 | 36 | 24 | 20 | 26 | 50 |
| **5** | 32 | 18 | 18 | 32 | 44 | 36 | 28 | 26 | 34 | 48 |
| **6** | 36 | 30 | 22 | 30 | 48 | 32 | 20 | 18 | 20 | 50 |
| **7** | 34 | 32 | 16 | 34 | 44 | 32 | 30 | 18 | 32 | 52 |
| **8** | 28 | 28 | 28 | 36 | 38 | 32 | 30 | 24 | 28 | 54 |
| **9** | 26 | 26 | 26 | 30 | 32 | 42 | 22 | 30 | 34 | 40 |
| **10** | 38 | 26 | 20 | 30 | 28 | 24 | 40 | 28 | 44 | 30 |

The features are:

F1: Spectral Centroid Mean

F2: Max Envelope Mean

F3: Zero Crossing Rate Mean

F4: Spectral Crest Mean

F5: Spectral Flux Mean

F6: Spectral Centroid Standard Deviation

F7: Max Envelope Standard Deviation

F8: Zero Crossing Rate Standard Deviation

F9: Spectral Crest Standard Deviation

F10: Spectral Flux Standard Deviation

Hence, the features after individual kNN classification performance with k=1 are ranked as below

1) Spectral Flux Standard Deviation (10)

2) Mean Spectral Flux (5)

3) Mean Spectral Crest (4)

4) Mean Spectral Centroid (1)

5) Spectral Centroid Standard Deviation (6)

6) Spectral Crest Standard Deviation (9)

7) Max Envelope Standard Deviation (7)

8) Mean Max Envelope (2)

9) Zero Crossing Rate Standard Deviation (8)

10) Mean Zero Crossing Rate (3)

**III) Computing Covariance**

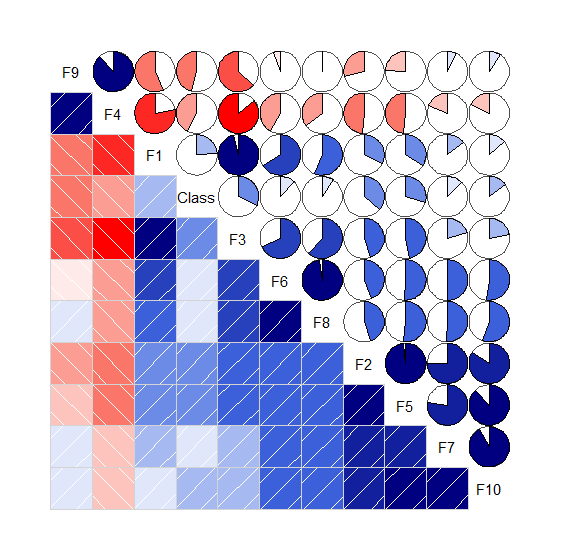
The covariance matrix for the 10 features is given as

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **Features** | | | | | | | | | | |
| **Features** |  | **F1** | **F2** | **F3** | **F4** | **F5** | **F6** | **F7** | **F8** | **F9** | **F10** |
| **F1** | 0.998 | 0.318 | 0.958 | -0.778 | 0.340 | 0.657 | 0.147 | 0.565 | -0.564 | 0.133 |
| **F2** | 0.318 | 0.998 | 0.452 | -0.488 | 0.978 | 0.436 | 0.749 | 0.450 | -0.292 | 0.842 |
| **F3** | 0.958 | 0.452 | 0.998 | -0.857 | 0.467 | 0.678 | 0.204 | 0.619 | -0.632 | 0.223 |
| **F4** | -0.778 | -0.488 | -0.857 | 0.998 | -0.479 | -0.415 | -0.189 | -0.354 | 0.876 | -0.177 |
| **F5** | 0.340 | 0.978 | 0.467 | -0.479 | 0.998 | 0.509 | 0.765 | 0.519 | -0.242 | 0.881 |
| **F6** | 0.657 | 0.436 | 0.678 | -0.415 | 0.509 | 0.998 | 0.499 | 0.974 | -0.054 | 0.523 |
| **F7** | 0.147 | 0.749 | 0.204 | -0.189 | 0.765 | 0.499 | 0.998 | 0.511 | 0.068 | 0.918 |
| **F8** | 0.565 | 0.450 | 0.619 | -0.354 | 0.519 | 0.974 | 0.511 | 0.998 | 0.014 | 0.551 |
| **F9** | -0.564 | -0.292 | -0.632 | 0.876 | -0.242 | -0.054 | 0.068 | 0.014 | 0.998 | 0.089 |
| **F10** | 0.133 | 0.842 | 0.223 | -0.177 | 0.881 | 0.523 | 0.918 | 0.551 | 0.089 | 0.998 |

**Discussion:**

A covariance matrix shows the linear dependence between pairs of features.

Plotting the above matrix as pie charts,



It can be clearly observed from the covariance matrix and the above plot that the following pairs of features are highly correlated and hence causing the KNN classifier to be less accurate:

F1 and F3

F9 and F4

F1 and F3

F6 and F8

F2 and F5

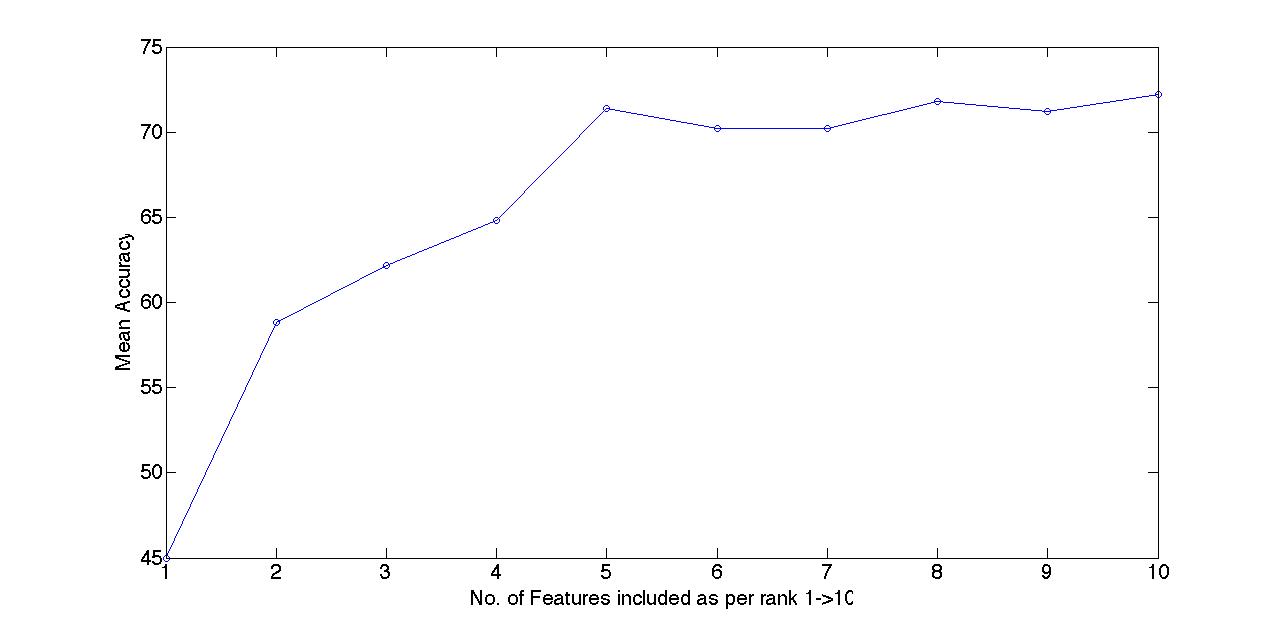
F7 and F10

F4 and F3

Therefore, ideally one feature from each pair should be used in the final KNN Classifier.

**IV) Forward Feature Selection**

The features are sorted as per accuracy and the forward selection leaves us with the following trend in increasing accuracy



This suggests that 5 features are reasonable enough to be used for ~70% accurate classification.

Therefore, the final features used in the KNN Classifier are:

1) Spectral Flux Standard Deviation

2) Mean Spectral Flux

3) Mean Spectral Crest

4) Mean Spectral Centroid

5) Spectral Centroid Standard Deviation

From the covariance matrix, it can be confirmed these features are reasonably independent and form a good subset for classification.

**V) nFold Cross Validation and Confusion Matrices**

Running the kNN classifier for the dataset with the above five features and their corresponding confusion matrices:

**i) k = 1**

Average Accuracy: 64%

Confusion Matrix:

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | Correct Class | | | | | |
| Predicted Class |  | Classical | Hip Hop | Country | Jazz | Metal |
| Classical | 69 | 2 | 14 | 12 | 7 |
| Hip Hop | 1 | 92 | 2 | 3 | 0 |
| Country | 11 | 2 | 60 | 28 | 1 |
| Jazz | 10 | 3 | 24 | 50 | 6 |
| Metal | 9 | 1 | 0 | 7 | 86 |

**ii) k = 3**

Average Accuracy: 62%

Confusion Matrix:

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | Correct Class | | | | | |
| Predicted Class |  | Classical | Hip Hop | Country | Jazz | Metal |
| Classical | 73 | 3 | 19 | 15 | 5 |
| Hip Hop | 1 | 92 | 4 | 5 | 0 |
| Country | 15 | 3 | 58 | 26 | 1 |
| Jazz | 2 | 1 | 19 | 48 | 6 |
| Metal | 9 | 1 | 0 | 6 | 88 |

**iii) k = 7**

Average Accuracy: 72%

Confusion Matrix:

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | Correct Class | | | | | |
| Predicted Class |  | Classical | Hip Hop | Country | Jazz | Metal |
| Classical | 76 | 1 | 12 | 13 | 4 |
| Hip Hop | 2 | 94 | 4 | 2 | 0 |
| Country | 13 | 1 | 69 | 30 | 0 |
| Jazz | 2 | 1 | 15 | 52 | 7 |
| Metal | 7 | 3 | 0 | 3 | 89 |

**Discussion:**

With the increase in the number of neighbors (k), the accuracy generally increases.

This was expected as, in the regions of overlap, more votes from the nearest neighbors increases the chances of identifying the right class.