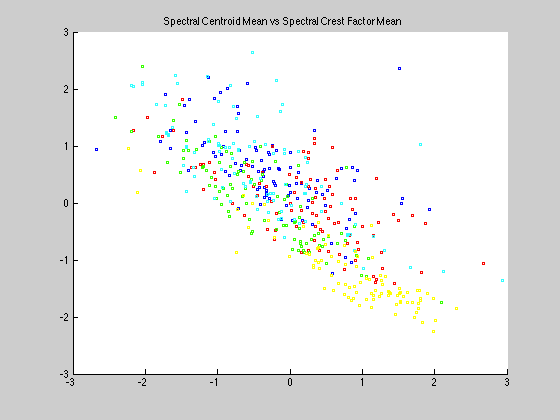
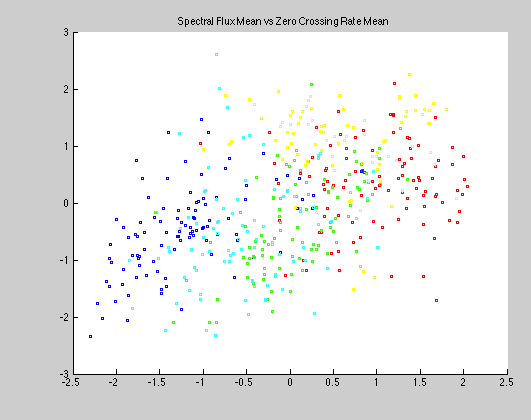
Assignment 2

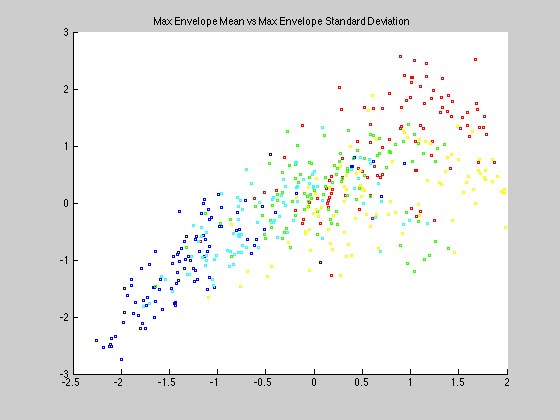
a) Following are the scatter plots



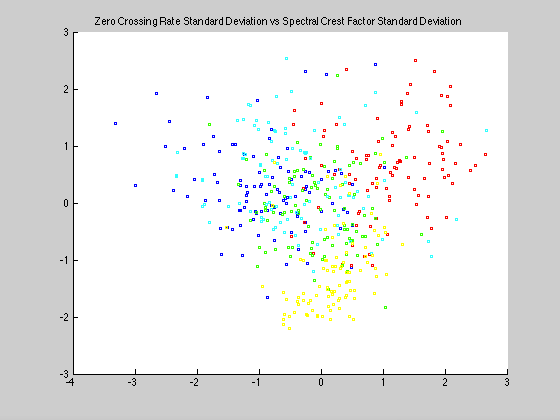
SPECTRAL CENTROID MEAN v/s SPECTRAL CREST FACTOR MEAN



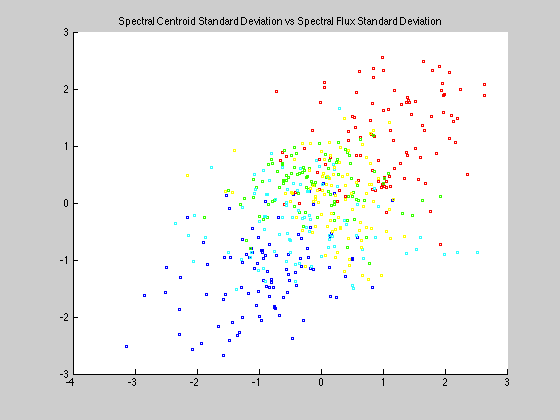
SPECTRAL FLUX MEAN v/s ZERO CROSSING RATE MEAN



MAX ENVELOPE MEAN v/s MAX ENVELOPE STANDARD DEVIATION



ZERO CROSSING RATE STANDARD DEVIATION v/s SPECTRAL CREST FACTOR STANDARD DEVIATION



SPECTRAL CENTROID STANDARD DEVIATION v/s SPECTRAL FLUX STANDARD DEVIATION

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

accuracyMatrix =

28 26 20 32 30 18 34 32 24 30

44 32 8 36 34 24 32 24 32 40

F

O

L

D

S

28 22 18 36 50 42 20 18 32 56

30 28 24 42 38 36 24 20 26 50

32 18 18 32 44 36 28 26 34 48

36 30 22 30 48 32 20 18 20 50

34 32 16 34 44 32 30 18 32 52

28 28 28 36 38 32 30 24 28 54

26 26 26 30 32 42 22 30 34 40

38 26 20 30 28 24 40 28 44 30

FEATURES 🡪

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)

The covariance matrix for the 10 features is given as

covarianceMatrix =

0.9980 0.3179 0.9580 -0.7777 0.3399 0.6566 0.1470 0.5654 -0.5644 0.1335

0.3179 0.9980 0.4516 -0.4880 0.9780 0.4364 0.7487 0.4497 -0.2924 0.8425

0.9580 0.4516 0.9980 -0.8568 0.4667 0.6780 0.2044 0.6187 -0.6316 0.2226

-0.7777 -0.4880 -0.8568 0.9980 -0.4792 -0.4145 -0.1890 -0.3542 0.8765 -0.1771

0.3399 0.9780 0.4667 -0.4792 0.9980 0.5087 0.7651 0.5188 -0.2418 0.8807

0.6566 0.4364 0.6780 -0.4145 0.5087 0.9980 0.4989 0.9737 -0.0541 0.5232

0.1470 0.7487 0.2044 -0.1890 0.7651 0.4989 0.9980 0.5106 0.0683 0.9184

0.5654 0.4497 0.6187 -0.3542 0.5188 0.9737 0.5106 0.9980 0.0138 0.5511

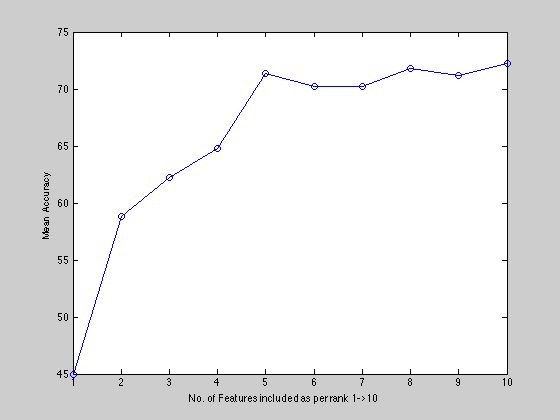
-0.5644 -0.2924 -0.6316 0.8765 -0.2418 -0.0541 0.0683 0.0138 0.9980 0.0889

0.1335 0.8425 0.2226 -0.1771 0.8807 0.5232 0.9184 0.5511 0.0889 0.9980

Discussion:

…

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.

Hence now running kNN for the dataset with 5 features and their corresponding confusion matrices.

i) k = 1

confusionMatrix =

275 11 96 55 40

10 398 35 25 16

113 44 244 126 24

51 26 104 241 67

51 21 21 53 353

ii) k = 3

confusionMatrix =

304 20 138 89 47

18 422 38 41 20

106 23 224 129 26

32 19 79 195 61

40 16 21 46 346

iii) k = 7

confusionMatrix =

292 10 100 68 39

15 433 30 26 11

119 20 254 153 32

41 20 86 202 66

33 17 30 51 352