EE 657 - Pattern Recognition and Machine Learning

Assignment -1

P. Harsha Vardhan,
Roll no.: 11010281,
ECE Branch,
EEE Department,
IIT Guwahati.

MATLAB CODES:

Assign1CovAll.m: Co-variances of all 5 models are calculated.

Assign1DecAll.m: Decision Statistics of 1(a)&(b).(i) and 1(a)&(b).(iii)

Assign1DecPool.m: Decision Statistics of 1(a)&(b).(ii) and 1(a)&(b).(iv)

Assign1DecSpherical.m: Decision Statistics of 1(a)&(b).(v)

RESULTS:

Question 1(a):

Due to memory issues during simulation in Matlab, every image is resized to 32x32 for feature computation. Regularization factor of the form $\boldsymbol{\lambda I}$ is added to each covariance matrix to overcome the curse of dimensionality. $\boldsymbol{\lambda}$ values for each covariance matrix are selected by trial and error method so that determinant of covariance matrix should not be zero and lies between 10^-15 and 10^15.

i) The samples of a given character class are modelled by an individual covariance matrix Σ_i

Lamda values for the covariance matrices:

$\boldsymbol{\lambda}$ value for $\Sigma_E = 0.37$

$\boldsymbol{\lambda}$ value for $\Sigma_C = 0.37$

$\boldsymbol{\lambda}$ value for $\Sigma_I = 0.5$

Table 1.1: Decision Statistics of Model (i)

Input (100 image files each)	No. of characters detected as 'E'	No. of characters detected as 'C'	No. of characters detected as 'I'
E test characters	94	6	0
C test characters	11	89	0
I test characters	1	0	99

Accuracy of E character = 94%

Accuracy of C character = 89%

Accuracy of I character = 99%

Average accuracy of the model= 94%

ii) The samples across all the character classes are pooled together to generate a common non diagonal covariance matrix Σ.

Lamda values for the covariance matrices:

λ value for Σ_{Pool} = 0.25

Table 1.2: Decision Statistics of Model (ii)

Input (100 image files each)	No. of characters detected as 'E'	No. of characters detected as 'C'	No. of characters detected as 'I'
E test characters	92	7	1
C test characters	3	93	4
I test characters	0	7	93

Accuracy of E character = 92%

Accuracy of C character = 93%

Accuracy of I character = 93%

Average accuracy of the model= 92.6667%

iii) The samples of a given character class are separately modelled by a diagonal covariance matrix Σ_i. The diagonal entries of the matrix correspond to the variances of the individual features. The features are assumed to be independent- hence their cross variances are forced to zero.

Lamda values for the covariance matrices:

λ value for Σ_E = 0.16

λ value for Σ_C = 0.145

λ value for Σ_I = 0.31

Table 1.3: Decision Statistics of Model (iii)

Input (100 image files each)	No. of characters detected as 'e'	No. of characters detected as 'e'	No. of characters detected as 'e'
E test characters	88	12	0
C test characters	18	82	0
I test characters	2	4	94

Accuracy of E character = 88%

Accuracy of C character = 82%

Accuracy of I character = 94%

Average accuracy of the model= 92.6667%

iv) The samples across all the character classes are pooled to generate a common diagonal covariance matrix Σ_{Pool_NDZ}. The diagonal entries correspond to the variances of the individual features, that are considered to be independent.

Lamda values for the covariance matrices:

λ value for Σ_{Pool_NDZ} = 0.11

Table 1.4: Decision Statistics of Model (iv)

Input (100 image files each)	No. of characters detected as 'e'	No. of characters detected as 'e'	No. of characters detected as 'e'
E test characters	83	14	3
C test characters	10	86	4
I test characters	1	0	99

Accuracy of E character = 83%

Accuracy of C character = 86%

Accuracy of I character = 99%

Average accuracy of the model= 89.333%

v) The covariance matrix of each class is forced to be spherical.

Covariance Matrix = Identity Matrix of size 1024x1024

Table 1.5: Decision Statistics of Model (v)

Input (100 image files each)	No. of characters detected as 'e'	No. of characters detected as 'e'	No. of characters detected as 'e'
E test characters	86	10	4
C test characters	10	85	5
I test characters	0	0	100

Accuracy of E character = 86%

Accuracy of C character = 85%

Accuracy of I character = 100%

Average accuracy of the model= 90.333%

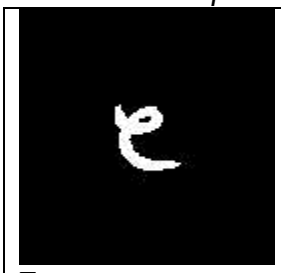
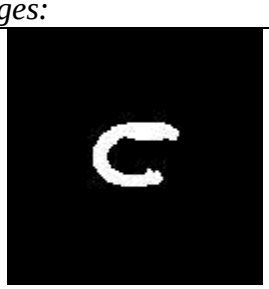
Question 1(b):

Images from the test set that are misclassified by each of the classifiers designed in Task 1(a).

Model (i)

Image files that are miss classified	*No. of mis-classifications*	*Image file numbers*
'E' misclassified as 'C'	6	201, 222, 230, 234, 245,247
'E' misclassified as 'I'	0	Nil
'C' misclassified as 'E'	11	203,205,221,222,223,238,274, 288,293,295,299
'C' misclassified as 'I'	0	Nil
'I' misclassified as 'E'	1	226
'I' misclassified as 'C'	0	Nil

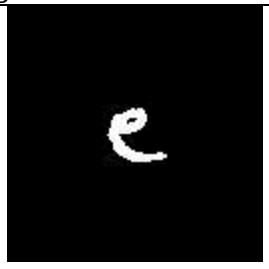
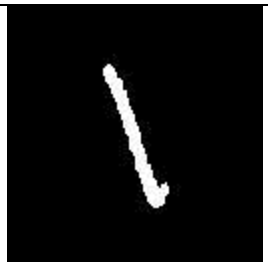
Four Misclassified images:

			
E test characters/247.jpg *-Misclassified as C*	C test characters/288.jpg *-Misclassified as C*	C test characters/299.jpg *-Misclassified as E*	I test characters/226.jpg *-Misclassified as E*

Model (ii)

Image files that are miss classified	*No. of mis-classifications*	*Image file numbers*
'E' misclassified as 'C'	7	215, 218,222,245,254, 257,264
'E' misclassified as 'I'	1	236
'C' misclassified as 'E'	3	245,279,292
'C' misclassified as 'I'	4	225,233,247,281
'I' misclassified as 'E'	0	Nil
'I' misclassified as 'C'	7	209,213,239,244,245,246,247

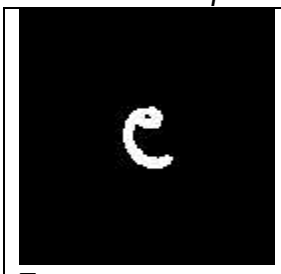
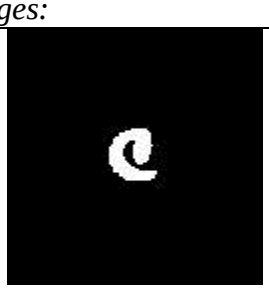
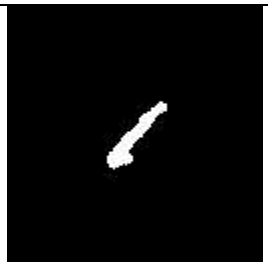
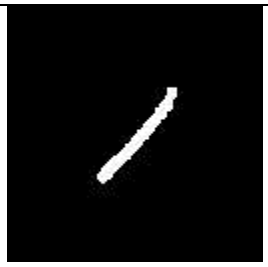
Four Misclassified images:

			
E test characters/236.jpg *-Misclassified as I*	E test characters/254.jpg *-Misclassified as C*	C test characters/233.jpg *-Misclassified as I*	I test characters/226.jpg *-Misclassified as C*

Model (iii)

Image files that are miss classified	*No. of mis-classifications*	*Image file numbers*
‘E’ misclassified as ‘C’	12	215,216,218,219,220, 222, 223,224,228,229,230, 234, 245
‘E’ misclassified as ‘I’	0	Nil
‘C’ misclassified as ‘E’	18	212,214,221,222,233,238,242, 245,250,266,273,281,283,285, 288,295,299
‘C’ misclassified as ‘I’	0	Nil
‘I’ misclassified as ‘E’	2	203,226
‘I’ misclassified as ‘C’	4	214,239,241,246

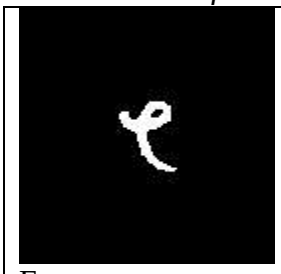
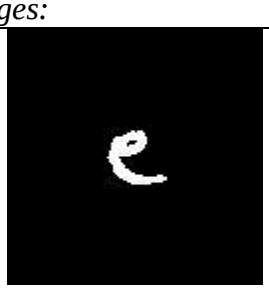
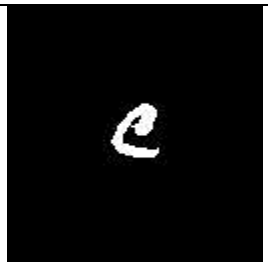
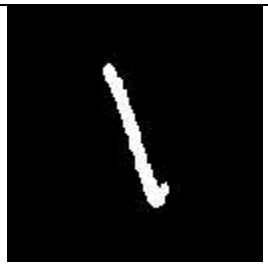
Four Misclassified images:

			
E test characters/245.jpg *-Misclassified as C*	C test characters/285.jpg *-Misclassified as E*	I test characters/203.jpg *-Misclassified as E*	I test characters/214.jpg *-Misclassified as C*

Model (iv)

Image files that are miss classified	*No. of mis-classifications*	*Image file numbers*
‘E’ misclassified as ‘C’	14	215,216,218,219,220, 222, 223,228,229,230,245,247,252,254
‘E’ misclassified as ‘I’	3	236,255,257
‘C’ misclassified as ‘E’	10	221,223,238,242,245,250, 273,274,295,299
‘C’ misclassified as ‘I’	4	212,234,256,283
‘I’ misclassified as ‘E’	1	226
‘I’ misclassified as ‘C’	0	Nil

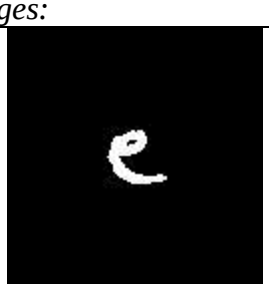
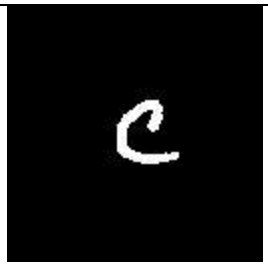
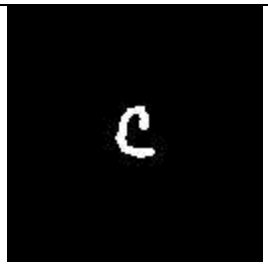
Four Misclassified images:

			
E test characters/236.jpg *-Misclassified as I*	E test characters/254.jpg *-Misclassified as C*	C test characters/283.jpg *-Misclassified as I*	I test characters/226.jpg *-Misclassified as E*

Model (v)

Image files that are miss classified	*No. of mis-classifications*	*Image file numbers*
‘E’ misclassified as ‘C’	10	215,216,218,219,222, 223,229,230,245,254
‘E’ misclassified as ‘I’	4	236,241,255,257
‘C’ misclassified as ‘E’	10	214,221,223,238,242,245, 250,273,274,285
‘C’ misclassified as ‘I’	5	210,212,234,256,283
‘I’ misclassified as ‘E’	0	Nil
‘I’ misclassified as ‘C’	0	Nil

Four Misclassified images:

			
E test characters/236.jpg *-Misclassified as I*	E test characters/254.jpg *-Misclassified as C*	C test characters/238.jpg *-Misclassified as E*	C test characters/256.jpg *-Misclassified as I*