Sneha Rajan
sneharajan@mail.usf.edu

Sobiya Manoharan
sobiya@mail.usf.edu

Project Report

Introduction:

This project is based on pattern recognition, a machine learning technique that is used to obtain readable data or patterns from huge datasets. The goal of our project is to extract some continuous-valued features from the scanned digits in our data set, which involves calculation of central moments, covariance, and inverse covariance for different forms of the dataset. The dataset on which these operations are done is an existing dataset called the MNIST digit recognition dataset. This paper will provide a detailed information on the process and the terms involved in the following sections.

MNIST dataset:

The MNIST dataset is an open-source database of handwritten digits, with a training set of 60,000 examples, and a test set of 10,000 examples. From this database, we used Dataset-A for our project which consists of 100 different files. These 100 files comprises of 10 files per digit and there are 10 such digits(0-9). For each digit, 10 different handwritten patterns(files) are provided in the format of a 28X28 image, for more accuracy and less error rate in digit recognition. Each digit is a two dimensional array and we assume i to be the row index increasing from top to bottom and j to be the column index increasing from left to right.

Terms and Formulae Used:

- **Mean(X) :** Mean is the average value of all the values in the dataset.

  Mean= Sum of Values(∑)/ Total number of values(N)

- **Variance:** Variance defines the measure of any point(x) in the dataset from its mean.

  Variance= ∑(X-x)^2 / N

- **Central Moment:** It is defined as the deviation or moment of the probability distribution of any random variable from its mean value.

  The (p,q)th Central Moment of a pattern X is:

  $Mpq = \sum\sum[((i-imean)^{p}) ((j-jmean)^{q}) X(i,j)]$, i=1 to I, j=1 to J.

Sneha Rajan
sneharajan@mail.usf.edu

Sobiya Manoharan
sobiya@mail.usf.edu

- **Root Mean Square Value:** It is the average of the values of varying quantities, especially when there are presence of both positive and negative values. Used for normalization.

- **Normalized Central Moment:** For any dataset, normalized central moment is the value obtained by dividing any dataset's central moment by its root mean square value.

- **Covariance:** Covariance is the measure of the distance travelled by two random variables from ordered sets in the same direction.

  $$\text{Covariance}(x,y)=\sum(x-X)(y-Y)/N$$

  where Y will be the mean of values y_i , i=1, 2,..

- **Inverse Covariance Matrix:** This is the inverse value of the corresponding dataset's covariance matrix, also known as precision matrix.

Implementation:

Language Used: Python Version-2.7.9

Packages Used: NumPy- A package, compatible with Python that allows easy scientific computing with its inbuilt functions.

The project was done using Python and we computed the questions as follows:

1. **Central Moments:** Using the formula mentioned above in the paper and the values of dataset-A, central moments were calculated and normalized using the corresponding root mean square values.

2. **Class-conditional covariance matrix:** Using the output of the previous calculation as input, covariance was computed using the 'cov' inbuilt function that is available in the NumPy package.

3. **Inverse Covariance:** Using the output of the covariance matrix, inverse is calculated using the 'inv()' function available in the NumPy package.

4. **Average Covariance and Inverse:** The average values of the covariance matrix is found and its inverse is calculated.

Sneha Rajan
sneharajan@mail.usf.edu

Sobiya Manoharan
sobiya@mail.usf.edu

Results:

Normalized central moments:

	M_{00}	M_{02}	M_{11}	M_{20}	M_{03}	M_{12}	M_{21}	M_{30}
0- 1	1.16	1.38	-1.42	1.16	0.35	0.01	0.95	-0.48
0- 2	1.30	1.74	-1.66	1.22	0.32	-1.56	0.98	0.18
0- 3	1.20	0.84	-1.19	1.11	-0.03	0.05	0.07	-0.12
0- 4	1.31	1.19	-1.39	1.12	-0.20	-0.20	0.50	-0.15
0- 5	1.54	2.21	-1.49	1.69	0.72	-0.92	1.05	-0.17
0- 6	1.44	2.05	0.06	1.53	0.61	0.91	1.01	0.41
0- 7	1.58	2.18	-1.81	1.39	0.20	-0.43	1.91	-0.77
0- 8	0.91	1.83	0.40	1.14	0.11	0.13	0.90	0.24
0- 9	1.39	1.14	-0.52	1.44	0.26	-0.21	0.50	0.20
0-10	1.22	1.19	-1.40	1.23	-0.03	-0.81	0.93	0.06
mean(0)	0.13	0.16	-0.10	0.13	0.02	-0.03	0.09	-0.01
var (0)	0.14	0.22	0.14	0.14	0.01	0.05	0.08	0.01
1- 1	0.63	0.40	-1.31	0.63	0.06	0.18	-0.20	0.12
1- 2	0.61	0.09	0.40	0.66	-0.02	0.00	0.02	0.04
1- 3	0.45	0.03	0.18	0.46	0.18	-0.00	-0.01	0.06
1- 4	0.49	0.03	0.10	0.48	-0.02	-0.00	-0.02	-0.06
1- 5	0.62	0.39	-1.25	0.60	0.11	0.07	-0.06	-0.01
1- 6	0.91	0.66	-0.49	0.85	-0.66	-0.67	1.19	0.33
1- 7	0.54	0.05	-0.21	0.54	0.12	0.01	-0.03	0.03
1- 8	0.62	0.33	-1.16	0.63	-0.12	0.07	-0.17	0.26
1- 9	0.65	0.24	-0.52	0.61	0.20	0.11	0.15	-0.55
1-10	0.80	0.29	0.48	0.92	-0.68	0.25	0.39	-0.22

Sneha Rajan
sneharajan@mail.usf.edu

Sobiya Manoharan
sobiya@mail.usf.edu

mean(1)	0.06	0.03	-0.04	0.06	-0.01	0.00	0.01	-0.00
var (1)	0.03	0.01	0.05	0.03	0.01	0.01	0.02	0.01
2- 1	1.23	1.39	-0.99	0.90	-0.77	0.04	2.32	-0.59
2- 2	0.98	1.08	-0.50	0.80	-0.94	0.55	0.92	-1.59
2- 3	1.48	1.66	-0.60	1.74	-0.80	0.05	1.18	0.05
2- 4	1.42	1.58	1.44	1.83	1.41	-2.12	-2.49	-1.14
2- 5	1.03	0.88	0.99	1.33	0.26	0.43	0.67	0.69
2- 6	1.38	1.29	-0.25	1.51	-2.23	-0.25	1.49	-1.37
2- 7	0.83	1.12	0.58	0.41	-0.56	3.23	0.94	-0.73
2- 8	1.42	1.69	-0.23	1.26	-0.57	0.82	1.32	-0.08
2- 9	1.39	1.63	1.33	1.51	0.58	-1.55	-1.82	-2.45
2-10	0.90	0.61	-0.46	1.04	-2.51	-0.75	1.48	-0.22
mean(2)	0.12	0.13	0.01	0.12	-0.06	0.00	0.06	-0.07
var (2)	0.12	0.15	0.07	0.14	0.16	0.19	0.24	0.12
3- 1	1.31	1.27	-1.73	1.50	0.28	-0.92	0.88	-1.48
3- 2	1.02	0.50	-0.60	1.14	-0.17	-0.67	1.06	-1.10
3- 3	1.32	1.51	1.30	1.62	-0.06	-1.78	-1.34	-3.30
3- 4	1.58	1.73	-1.15	1.86	0.86	-1.79	0.80	-2.69
3- 5	1.01	1.14	1.57	0.72	0.35	-1.91	-1.12	-0.20
3- 6	0.69	0.33	0.62	0.79	0.81	-0.48	-0.64	-0.67
3- 7	1.21	1.06	-1.10	1.42	1.22	-1.32	0.14	-1.62
3- 8	0.84	0.49	-0.12	1.09	-0.03	-0.37	0.58	-1.14
3- 9	0.97	0.45	0.37	1.12	0.56	-0.21	-0.03	-1.07
3-10	0.91	1.15	2.16	1.21	1.23	-1.88	-1.68	-1.29

Sneha Rajan
sneharajan@mail.usf.edu

Sobiya Manoharan
sobiya@mail.usf.edu

mean(3)	0.11	0.10	0.01	0.12	0.05	-0.11	-0.01	-0.15
var (3)	0.10	0.10	0.15	0.14	0.04	0.15	0.09	0.25

4- 1	0.79	1.64	0.25	0.52	0.30	-4.10	-2.22	1.78
4- 2	0.85	0.83	-1.15	0.55	-0.20	0.45	-1.30	1.10
4- 3	1.26	2.13	1.24	0.89	0.02	-3.37	-1.46	1.93
4- 4	0.61	0.29	-0.01	0.48	0.64	-0.26	-0.66	0.11
4- 5	0.77	0.59	-0.76	0.56	1.06	1.11	0.20	-1.26
4- 6	1.14	1.08	0.68	0.87	0.79	-1.66	-0.81	1.68
4- 7	1.10	2.19	0.39	0.60	-0.47	-1.18	-0.19	0.74
4- 8	0.75	0.50	-0.97	0.53	0.63	0.40	-0.30	-0.28
4- 9	1.04	1.28	-0.74	0.83	-0.32	0.19	-1.24	1.23
4-10	1.03	1.33	-1.20	0.74	0.81	1.37	-2.31	0.30

mean(4)	0.09	0.12	-0.02	0.07	0.03	-0.07	-0.10	0.07
var (4)	0.07	0.15	0.07	0.04	0.03	0.35	0.15	0.14

5- 1	1.09	0.96	-1.12	1.44	0.69	-0.14	0.25	-1.60
5- 2	0.60	0.76	-1.23	0.35	-0.02	0.97	-0.81	0.47
5- 3	0.64	0.68	-1.27	0.47	0.02	-0.01	-0.14	0.18
5- 4	0.73	0.87	-1.71	0.76	-0.06	1.51	-1.50	0.89
5- 5	0.68	0.91	-1.41	0.43	0.36	1.79	-1.03	0.07
5- 6	0.70	0.32	0.28	0.77	0.34	0.09	-0.17	1.47
5- 7	0.67	0.17	-0.29	0.83	-0.19	-0.23	0.47	-0.82
5- 8	1.12	0.88	-0.59	1.22	-1.13	-0.58	0.97	0.54

Sneha Rajan
sneharajan@mail.usf.edu

Sobiya Manoharan
sobiya@mail.usf.edu

5- 9	0.83	1.06	-1.38	0.55	0.18	2.38	-2.02	0.78
5-10	0.68	0.80	-1.23	0.37	-0.01	0.72	-0.56	0.29
mean(5)	0.08	0.07	-0.10	0.07	0.00	0.06	-0.05	0.02
var (5)	0.05	0.05	0.11	0.05	0.02	0.12	0.09	0.07
6- 1	1.10	0.81	-0.50	0.89	-1.45	0.26	0.62	0.21
6- 2	0.70	0.34	-0.21	0.72	-0.94	0.03	0.31	0.76
6- 3	0.79	0.47	0.16	0.78	-1.93	0.34	0.49	0.22
6- 4	1.08	0.52	0.22	0.97	-1.61	0.19	0.67	0.66
6- 5	1.05	0.65	0.07	0.80	-0.88	0.30	0.32	-0.44
6- 6	1.06	0.53	-0.33	0.96	-0.92	0.07	0.18	0.73
6- 7	0.87	0.56	-0.46	0.68	-0.98	0.03	0.48	0.50
6- 8	1.07	0.64	-0.46	0.84	-1.44	0.09	0.71	0.09
6- 9	0.93	0.75	-0.45	0.72	-1.12	0.18	0.82	-0.20
6-10	1.16	0.98	-0.43	1.03	-1.84	0.18	0.37	1.30
mean(6)	0.10	0.06	-0.02	0.08	-0.13	0.02	0.05	0.04
var (6)	0.08	0.03	0.01	0.06	0.15	0.00	0.02	0.04
7- 1	0.95	0.77	-0.88	0.86	1.66	0.07	-0.82	-1.14
7- 2	0.67	0.49	-0.54	0.37	0.50	-0.65	-0.48	-0.64
7- 3	0.93	0.97	0.49	0.65	0.20	-0.53	-0.60	-0.63
7- 4	0.54	0.17	0.03	0.59	1.34	-0.28	-0.47	-0.11
7- 5	1.03	1.02	0.53	1.11	2.68	-1.17	-2.41	-0.46
7- 6	0.85	0.75	-1.25	0.83	2.05	0.20	0.04	-2.15

Sneha Rajan
sneharajan@mail.usf.edu

Sobiya Manoharan
sobiya@mail.usf.edu

7- 7	0.89	0.75	-1.32	0.82	1.83	0.19	-0.16	-1.66
7- 8	0.72	0.68	0.53	0.96	1.71	-1.86	-2.63	-1.06
7- 9	0.97	1.03	-1.22	0.91	2.12	-0.28	-0.09	-2.38
7-10	0.70	0.25	-0.22	0.80	1.45	-0.12	-0.65	-0.44
mean(7)	0.08	0.07	-0.04	0.08	0.16	-0.04	-0.08	-0.11
var (7)	0.06	0.05	0.06	0.05	0.25	0.05	0.13	0.15
8- 1	1.06	0.76	-1.76	1.03	0.60	0.73	-0.65	-0.19
8- 2	1.15	1.18	-2.47	1.19	0.15	0.21	0.27	-0.59
8- 3	1.06	0.58	-1.12	1.07	0.47	0.14	-0.26	-0.02
8- 4	1.02	0.53	-0.27	1.09	0.85	-0.10	-0.33	0.47
8- 5	1.23	1.00	-1.91	1.39	1.39	1.10	-1.05	-0.66
8- 6	1.06	0.84	-1.74	1.24	0.68	0.26	-0.07	-0.65
8- 7	1.04	0.56	-0.98	1.08	0.72	0.44	-0.58	-0.51
8- 8	1.15	0.81	-1.58	1.27	0.67	0.60	-0.56	-0.56
8- 9	1.06	0.64	-1.22	1.17	0.25	0.43	-0.45	0.20
8-10	0.85	0.35	-0.97	0.89	0.72	0.11	0.09	-0.94
mean(8)	0.11	0.07	-0.14	0.11	0.07	0.04	-0.04	-0.03
var (8)	0.09	0.05	0.19	0.11	0.04	0.02	0.02	0.03
9- 1	0.93	0.56	0.09	0.83	1.66	-0.58	-0.68	1.63
9- 2	0.72	0.32	-0.90	0.73	1.18	0.07	0.29	-1.31
9- 3	0.71	0.31	-0.66	0.58	0.63	0.19	-0.27	-0.26
9- 4	0.85	0.58	-1.46	0.86	1.52	-0.30	1.02	-2.21
9- 5	0.70	0.23	-0.58	0.71	1.27	0.05	-0.05	-0.84
9- 6	0.95	0.57	-0.38	0.82	1.74	-0.14	-0.84	-0.12
9- 7	0.74	0.27	0.59	0.85	0.53	-0.04	0.08	-0.31
9- 8	0.96	0.90	-0.18	0.85	2.09	0.19	-2.31	-0.00

Sneha Rajan
sneharajan@mail.usf.edu

Sobiya Manoharan
sobiya@mail.usf.edu

9- 9	0.77	0.38	-0.67	0.69	1.23	0.08	-0.26	-0.78
9-10	1.07	0.81	-0.00	1.18	0.12	0.03	0.47	-0.17
mean(9)	0.08	0.05	-0.04	0.08	0.12	-0.00	-0.03	-0.04
var (9)	0.06	0.02	0.04	0.06	0.15	0.01	0.08	0.10

RMS	152.35	2996.81	1341.31	4855.60	7851.78	4009.52	3677.05	3902.26

COVARIANCE:

Class 1 : [5, 7] , [3, 7] : 4.7, -7.6

Class 2 : [6, 7] , [5, 1] : -7.4, -3.4

Class 3 : [4, 1] , [4, 2] : 8.1, 7.5

Class 4 : [3, 8] , [2, 5] : 19.6,4.2

Class 5 : [3, 5] , [3, 7] : -7.4,-0.1

Class 6 : [7, 8] , [4, 6] : -30.3, -18.6

Class 7 : [3, 5] , [2, 4] : -2.3, 0.9

Sneha Rajan
sneharajan@mail.usf.edu

Sobiya Manoharan
sobiya@mail.usf.edu

Class 8 : [8, 5] , [8, 6] : -20.7, -15.3

Class 9 : [3, 7] , [2, 5] : -1.3, -0.3

Class 10 : [3, 2] , [1, 5] : 1.6, 1.1

INVERSE COVARIANCE MATRIX

(Multiplier: 100.000000)

Class 1 : [5, 7] , [3, 7] : 5.903 , -3.105

Class 2 : [6, 7] , [5, 1] : 1006.691 , 470.876

Class 3 : [4, 1] , [4, 2] : -1483.034 , 714.618

Class 4 : [3, 8] , [2, 5] : -4.559 , 27.155

Class 5 : [3, 5] , [3, 7] : -6.453 , -7.788

Class 6 : [7, 8] , [4, 6] : 38.321 , -45.738

Sneha Rajan
sneharajan@mail.usf.edu

Sobiya Manoharan
sobiya@mail.usf.edu

Class 7 : [3, 5] , [2, 4] : -16.397 , 696.466

Class 8 : [8, 5] , [8, 6] : -38.288 , -35.715

Class 9 : [3, 7] , [2, 5] : 306.100 , -128.880

Class 10 : [3, 2] ,[1, 5] : 117.314, 659.715

AVERAGE COVARIANCE MATRIX:

(Multiplier 100.000)

[3,5] ,[6,3] :4.2 , -22.3

INVERSE AVERAGE COVARIANCE MATRIX:

(Multiplier 100.000)

[3,5] ,[6,3] : 16.1 , 69.9