MNIST PIPELINE

1. Height, Radial and Density Data: Total_data : (60000 , 42) entries

Every filtration except grayscale has been performed on binarized image with threshold 40%.

Filtration Type	Parameter	Vectorization	Column
Height	1,0	Persistent Entropy	1 (dim 0), 2 dim(1)
Height	0,1	Persistent Entropy	3,4
Height	1,1	Persistent Entropy	5,6
Height	-1,1	Persistent Entropy	7,8
Height	1,-1	Persistent Entropy	9,10
Height	-1,-1	Persistent Entropy	11,12
Height	-1,0	Persistent Entropy	13,14
Height	0,-1	Persistent Entropy	15,16
Radial	6,6	Persistent Entropy	17,18
Radial	13,6	Persistent Entropy	19,20
Radial	20,6	Persistent Entropy	21,22
Radial	20,13	Persistent Entropy	23,24
Radial	13,13	Persistent Entropy	25,26
Radial	6,13	Persistent Entropy	27,28
Radial	6,20	Persistent Entropy	29,30
Radial	13,20	Persistent Entropy	31,32
Radial	20,20	Persistent Entropy	33,34
Grayscale	NIL	Persistent Entropy	35,36
Density	2	Persistent Entropy	37,38
Density	4	Persistent Entropy	39,40
Density	6	Persistent Entropy	41,42

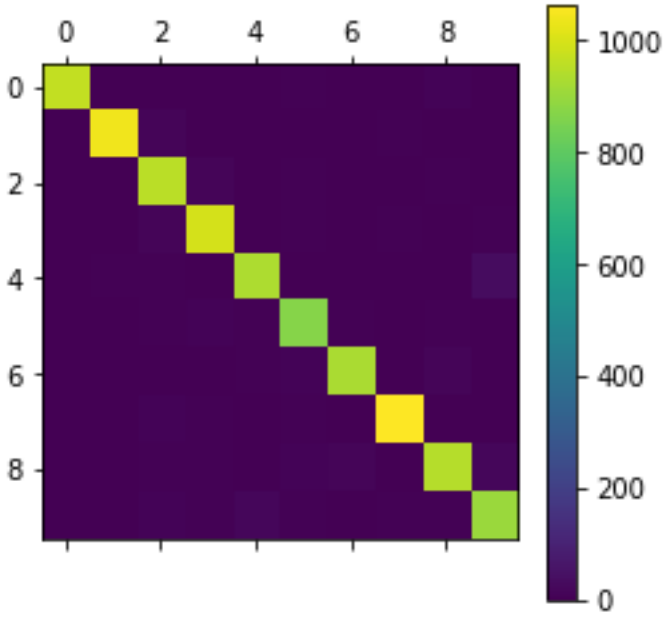
Height_1 denotes the first 8 columns of the height data and height 2 the remaining 8.

2. Height,Radial and Density Data: split into test and train: 50,000 and 10,000 rows respectively

Random Forest classifier with max depth 1000 applied to train data

Confusion Matrix for prediction: Labels = [0,1,2,3,4,5,6,7,8,9]

967	0	3	1	0	6	3	1	9	1
0	1042	13	0	1	0	0	7	0	1
2	1	953	13	0	6	3	4	5	3
0	0	15	995	0	7	0	7	0	6
1	5	5	0	931	1	4	1	1	34
4	0	7	10	5	870	8	0	7	4
0	3	3	0	7	9	929	0	13	3
0	3	10	6	1	5	0	1064	0	1
3	2	6	5	3	10	16	0	947	17
4	1	11	6	18	5	2	7	5	902



3. Other Data used:

Dilation data: Consists of persistent entropy applied to cubical homology persistence corresponding to dilation, erosion and signed distance filtration. Data consists of 6 columns.

Vietoris Rips: We consider the binary image as a point cloud in R2. Vietoris Rips homology and then persistent entropy is applied. Data consists of 2 columns.

Line Data: Given any line of the form ax+by+c=0 and a binary image, each pixel with value 1 is assigned a value equal to either distance or negative of distance from the line depending on whether parameter par value is true or false respectively. For pixels with grayscale value 0, the maximum of the distance from line of 1 pixels + 5 is considered.

Row is considered to represent x axis while column, the y axis.

The following lines are considered:

Line 1: x – y = 20 , par = True ; Line 2: y = 27, par = True ; Line 3: y = 13 , par = False ;

Line 4 : x = 13, par = False.

4. Data and accuracy: Random forest classifier with max no. of estimators: 1000, random_State =31415

S. No	Size	height_1	radial	grayscale	density	height_2	dilation	line	Vietoris	accuracy
1	34	yes	yes	yes	yes	no	no	no	no	95.05
2	42	yes	yes	yes	yes	yes	no	no	no	96
3	48	yes	yes	yes	yes	yes	yes	no	no	96
4	56	yes	yes	yes	yes	yes	yes	yes	no	96.11
5	52	yes	yes	yes	yes	yes	yes	1:4 col	no	96.05
6	46	yes	yes	yes	yes	yes	no	1:4 col	no	96.01
7	38	yes	yes	yes	yes	no	no	1:4 col	no	95.64
8	50	yes	yes	yes	yes	yes	no	yes	no	96.13
9	52	yes	yes	yes	yes	yes	no	yes	yes	96.16

5. Confusion matrix corresponding to data 9. (52 columns)