

DMML ASSIGNMENT-3

Jerin Biju(MDS202329)

Aparna C(MDS202310)

We have done the assignment using kmeans clustering with number of clusters varying from k=20 to 100 with step size of 5 for second data and k=20 to 70 with step size 5 for first data. We have tried the following approaches,

- 1) Randomly selected k number of points, trained MLP using these selected points and then checked accuracy.
- 2) Clustered the data into k clusters, took the geometrical centroids, chose the nearest points to centroids to train the MLP and checked the accuracy on the test data.
- 3) Instead of taking just the centroids, took the whole clusters with the labels same as the centroids and chosen as the training data. Again tested the accuracy on the test data.
- 4) Centroid labels are extrapolated to the near points (sorted the data points within the cluster in the ascending order of distance from the centroid), chosen as training data, and verified the accuracy on the test data.

The accuracy we obtained is as follows.

Q1)

no_of_points		accuracy with random points	accuracy with centroids
0	20	0.4611	0.5677
1	30	0.5913	0.6310
2	40	0.6099	0.5895
3	50	0.6288	0.6548
4	60	0.6541	0.6388
5	70	0.7072	0.6688

accuracy_with_extrapolated_to_fulldata \

0	0.5588
1	0.5425
2	0.5507
3	0.6264
4	0.5855
5	0.6135

accuracy_with_extrapolated_to_near_points

0	0.5661
1	0.5542
2	0.5290
3	0.6266
4	0.5883
5	0.6310

Q2)

no_of_points		accuracy with random points	accuracy_with centroids
0	20	0.150235	0.218779
1	25	0.157746	0.209390
2	30	0.145540	0.211268
3	35	0.150235	0.260094
4	40	0.156808	0.276995

5	45	0.159624	0.244131
6	50	0.173709	0.293897
7	55	0.170892	0.299531
8	60	0.201878	0.299531
9	65	0.197183	0.264789
10	70	0.197183	0.284507
11	75	0.213146	0.271362
12	80	0.227230	0.300469
13	85	0.232864	0.297653
14	90	0.240376	0.335211
15	95	0.231925	0.310798
16	100	0.262911	0.295775

accuracy_with_extrapolated_to_fulldata \

0	0.353991
1	0.370892
2	0.359624
3	0.430047
4	0.347418
5	0.372770
6	0.372770
7	0.388732
8	0.365258
9	0.392488
10	0.351174
11	0.350235

12	0.362441
13	0.344601
14	0.392488
15	0.364319
16	0.388732

accuracy_with_extrapolated_to_near_points

0	0.350235
1	0.339906
2	0.351174
3	0.397183
4	0.333333
5	0.327700
6	0.359624
7	0.329577
8	0.364319
9	0.359624
10	0.327700
11	0.354930
12	0.344601
13	0.332394
14	0.394366
15	0.388732
16	0.373709

