255- Machine Learning LAB-2 Report

By: Shalabh Neema

SJSU ID : 014546259

PART-1

Dataset\_1 S1

K-Nearest-Neighbor

Before.

A screenshot of a cell phone

Description automatically generated

**After**

**A screenshot of a cell phone

Description automatically generated**

Accuracy of both the algorithms implemented using NumPy as well as sk-learn gives the same accuracy that is 99.28 which is quite good. The clusters are evenly divided with no overlapping of boundaries. The graph observed before and after the prediction is quite similar as our model is able to correctly predict most of the values.

K Means Before

A screenshot of a cell phone

Description automatically generated

K means after

A screenshot of a cell phone

Description automatically generated

By visualizing the two graphs we can say that sk-learn correctly identified the clusters and their respective centroids while the NumPy algorithm has some misclassification of centroids’ is able to find out 12 clusters correctly. In remaining three, it placed two centroid in the same cluster and the last centroid in between 2 clusters.

We can improve the performance by taking the initial clusters to be random rather than initializing with zero.

Dataset\_1 S2

Before K-NN

A screenshot of a cell phone

Description automatically generated

After K NN

A screenshot of a cell phone

Description automatically generated

Accuracy of both the algorithms implemented using NumPy as well as sk-learn gives almost the same accuracy across 97.7 which is quite good. The clusters are evenly divided with no overlapping of boundaries. The graph observed before and after the prediction is quite similar as our model is able to correctly predict most of the values.

**K-Means:**

Numpy:

A screenshot of a cell phone

Description automatically generated

Sk-learn

A screenshot of a cell phone

Description automatically generated

By visualizing the two graphs we can say that sk-learn correctly identified the clusters and their respective centroids while the NumPy algorithm has some misclassification of centroids’ is able to find out 12 clusters correctly. In remaining three, it placed two centroid in the same cluster and the last centroid in between 2 clusters.

We can improve the performance by taking the initial clusters to be random rather than initializing with zero.

Dataset\_1 S3

**K-Nearest-Neighbor**

Before:

A screenshot of a cell phone

Description automatically generated

After:

A screenshot of a cell phone

Description automatically generated

Accuracy of both the algorithms implemented using NumPy as well as sk-learn gives almost the same accuracy across 87 % which is somewhat less compared to the prediction made in previous dataset. The clusters are not evenly divided and we can see some overlapping of boundaries. The graph observed before and after the prediction is quite similar as our model is able to correctly predict most of the values.

Dataset\_2

**K-Nearest-Neighbor**

Before:

A screenshot of a cell phone

Description automatically generated

After:

A screenshot of a cell phone

Description automatically generated

Accuracy of both the algorithms implemented using NumPy as well as sk-learn gives the same accuracy 100% which is quite good. The clusters are evenly divided with no overlapping of boundaries. The graph observed before and after the prediction is quite similar as our model is able to correctly all the values .

Dataset\_3

**K-Nearest-Neighbor**

Before:

A screenshot of a cell phone

Description automatically generated

After:

A close up of a logo

Description automatically generated

Accurcay of both the algorithms implented using numpy as well as sklearn gives the same accuracy 100% which is quite good. The clusters are evenly divided with no overlapping of boundaries. The graph observed before and after the prediction is quite similar as our model is able to correctly all the values .

PART -2

Recommendation system

**Recommendations for Toy Story (1995):**

Toy Story (1995)

Toy Story 2 (1999)

Babe (1995)

Bug's Life, A (1998)

Pleasantville (1998)

Babe: Pig in the City (1998)

Aladdin (1992)

Stuart Little (1999)

Secret Garden, The (1993)

Tarzan (1999)

For Toy story, we are seeing that a children genre movie has been recommended which actually make sense. All the movies like Stuart Little, Aladdin, Tarzan follow in the same Category.

**Recommendations for Babe (1995):**

Babe (1995)

Toy Story 2 (1999)

Babe: Pig in the City (1998)

Toy Story (1995)

Secret Garden, The (1993)

Madeline (1998)

Bug's Life, A (1998)

Stuart Little (1999)

Matilda (1996)

101 Dalmatians (1996)

Babe(1995), it is a kind of animal movie which may fall in the children’s category .All the movies like Toy Story, Stuart Little , Aladdin, Tarzan follow in the same Category that has been recommended by the system

**Recommendations for Two Bits (1995):**

Two Bits (1995)

Zero Kelvin (Kj�rlighetens kj�tere) (1995)

Taffin (1988)

It Happened Here (1961)

Broken English (1996)

Enfer, L' (1994)

Hav Plenty (1997)

Shopping (1994)

301, 302 (1995)

Wisdom of Crocodiles, The (a.k.a. Immortality) (2000)

Two Bits is a historical/drama genre category . The recommended movies also falls in the same zone.