Homework3 Question 2

17/11/2023

40/40 Points

Attempt 1

Review Feedback
17/11/2023

Attempt 1 Score:
40/40

Add Comment

Anonymous Grading: **no**

Unlimited Attempts Allowed
21/10/2023

Details

Introduction

This portion of Homework 2 will be done individually, not in a group. Same as in Question 1, your assignment should be submitted by uploading your code (in the form of a **Jupyter Notebook (.ipynb) AND pdf copy of the files** – so we can make comments directly on the file) to Canvas. **Be sure to run the file before committing so that we can directly see your results.** Please mention all the resources that were used to solve the problem (e.g., websites, books, research papers, other people, etc.). To complete the assignment, you can use any Python (or R) package that you want, but we recommend using Scikit-Learn.

Question

To gain a better understanding of the differences across datasets, perform the same tasks as in Question 1 (except for questions 4 and 7D), but on a dataset of your choice (if you worked on a team for Question 1, please do not select the same dataset as your team members). The dataset should contain multiple features (attributes) and you can perform binary or multi-class classification. If your dataset does not have a train/test split you can create it as you find appropriate to complete the assignment.

View Rubric

Select Grader

Md Rysul Kabir (TA)

(https://iu.instructure.com/courses/2165858/modules/items/30397203)

Attempt (https://iu.instructure.com/c

HW3 Q2

Criteria	Ratings		Pts
Question -1	**3 to >0 pts** **Full Marks** PCA (2) Number of Component required to preserve 95 % variance (1)	**0 pts** **No Marks**	3 / 3 pts
Question -2	**3 to >0 pts** **Full Marks** 10 images in original from (1.5) Their reconstruction (1.5)	**0 pts** **No Marks**	3 / 3 pts
Question -3 A	**2 to >0 pts** **Full Marks** PCA with 2 dimensions (1) Amount of Variance preserved with these 2 components (1)	**0 pts** **No Marks**	2 / 2 pts
Question -3 B	**3 to >0 pts** **Full Marks** Scatter Plots of Components with some Rock Images a) t-SNE (1) b) LLE (1) c) MDS (1)	**0 pts** **No Marks**	3 / 3 pts
Question -3 C	**1 to >0 pts** **Full Marks** Discussion on the visualizations (preferred or not) (1)	**0 pts** **No Marks**	1 / 1 pts
Question -5 A	**4 to >0 pts** **Full Marks** K-means is implemented	**0 pts** **No Marks**	4 / 4 pts

(https://iu.instructure.com/courses/2165858/modules/items/30397203) (https://iu.instructure.com/c

HW3 Q2

Criteria	Ratings		Pts
	clusters using one of the techniques (2)		
Question -5 B	**4 to >0 pts** **Full Marks** Visualization: Boundaries can be inferred with centroids of each cluster (3), Dots are color mapped according to labels (1)	**0 pts** **No Marks**	4 / 4 pts
Question -6 A	**4 to >0 pts** **Full Marks** EM has been implemented correctly (2), Evaluation for the number of clusters (2)	**0 pts** **No Marks**	4 / 4 pts
Question -6 B	**4 to >0 pts** **Full Marks** Visualization: Boundaries can be inferred with centroids of each cluster (3), Dots are color mapped according to labels (1)	**0 pts** **No Marks**	4 / 4 pts
Question -6 C	**4 to >0 pts** **Full Marks** Generation of 20 rock images using sample method with visualization - (4)	**0 pts** **No Marks**	4 / 4 pts
Question -7 A	**1 to >0 pts** **Full Marks** Training Time has been Reported	**0 pts** **No Marks**	1 / 1 pts

(https://iu.instructure.com/courses/2165858/modules/items/30397203) (https://iu.instructure.com/c

HW3 Q2

Criteria	Ratings		Pts
	Full Marks Sequential Model has been implemented correctly with right number of neurons (-2) Validation Data has been incorporated (-1) Accuracy is increasing with epochs (-1). Plots of val and training loss via training epochs (-2)	**No Marks**	
Question -7 C	**1 to >0 pts** **Full Marks** Total Number of parameters (0.75) Number of Bias Parameters (0.25)	**0 pts** **No Marks**	1 / 1 pts
			Total Points: 40

	File Name	Size	
	Fashion.pdf	3.81 MB	
	Fashion.ipynb	5.54 MB	

(https://iu.instructure.com/courses/2165858/modules/items/30397203) (https://iu.instructure.com/c

v9c2ugjeb

November 18, 2023

1 Fashion Dataset

https://www.kaggle.com/datasets/paramaggarwal/fashion-product-images-small/data

The growing e-commerce industry presents us with a large dataset waiting to be scraped and researched upon. In addition to professionally shot high resolution product images, the dataset includes multiple label attributes describing the product which was manually entered while cataloging.

Each product is identified by an ID like 42431. A map to all the products in styles.csv. From here, you can fetch the image for this product from images/42431.jpg. This will serve as the label for ourusecase.

Our task here would be to classify the products into their masterCategory which are Accessories', 'Apparel', 'Footwear', 'Personal Care', 'Free Items', 'Sporting Goods', 'Home' and 'Cosmetics'

1.1 Data Preprocessing

[9]:
```
# Loading the CSV file
import pandas as pd
import os
import cv2

csv_file_path = 'archive/styles.csv'
csv_data = pd.read_csv(csv_file_path)
csv_data.head()
```

[9]:

	id	gender	masterCategory	subCategory	articleType	baseColour	season	\
0	15970	Men	Apparel	Topwear	Shirts	Navy Blue	Fall	
1	39386	Men	Apparel	Bottomwear	Jeans	Blue	Summer	
2	59263	Women	Accessories	Watches	Watches	Silver	Winter	
3	21379	Men	Apparel	Bottomwear	Track Pants	Black	Fall	

(https://iu.instructure.com/courses/2165858/modules/items/30397203) (https://iu.instructure.com/c