ML-Assignment 2 report

1. Change all the variable names

2. Write comments

3. Shuffle import statements

4. in task1, we are not getting result properly for kmeans,, so see that.

5. in task 1, didn’t try GMM. Try that also

6. again mention their references in the report

7. in task 2, pca revisit..

8. In task 2, I did kmeans and gmm using inbuilt libraries, keep them like that only and try to do using task1 kmeans and gmm functions,

9. make everything including task1 and 2 in single notebook only.

GMM links:

1. <https://python-course.eu/machine-learning/expectation-maximization-and-gaussian-mixture-models-gmm.php>

2. <http://www.oranlooney.com/post/ml-from-scratch-part-5-gmm/>

3. <https://medium.com/@siddharthvadgama/gaussian-mixture-model-gmm-using-em-algorithm-from-scratch-6b7c764aac9f>

4. <https://towardsdatascience.com/gaussian-mixture-models-implemented-from-scratch-1857e40ea566>

5. <https://github.com/ScienceKot/mysklearn/blob/master/Gaussian%20Mixture%20Models/GMM.py>

6. <https://www.youtube.com/results?search_query=+GMM+using+EM+from+scratch>

7. <https://scikit-learn.org/stable/modules/generated/sklearn.decomposition.PCA.html>

8. <https://towardsdatascience.com/gaussian-mixture-models-and-expectation-maximization-a-full-explanation-50fa94111ddd>

Task1: Kmeans

References

1. <https://www.youtube.com/watch?v=vtuH4VRq1AU>

2. <https://www.javatpoint.com/k-means-clustering-algorithm-in-machine-learning#:~:text=Step%2D1%3A%20Select%20the%20number,form%20the%20predefined%20K%20clusters>.

3. <https://github.com/python-engineer/MLfromscratch/blob/master/mlfromscratch/kmeans.py>

4. <https://towardsdatascience.com/create-your-own-k-means-clustering-algorithm-in-python-d7d4c9077670>

5. <https://github.com/turnerluke/ML-algos/blob/main/k_means/k_means.py>

6. <https://towardsdatascience.com/how-to-split-a-dataset-into-training-and-testing-sets-b146b1649830>

7. <https://www.geeksforgeeks.org/how-to-select-multiple-columns-in-a-pandas-dataframe/>

8. <https://datatofish.com/k-means-clustering-python/>

9. <https://stackoverflow.com/questions/67464023/i-wrote-a-kmeans-class-but-the-results-look-strange-what-am-i-doing-wrong>

10. <https://medium.com/pursuitnotes/k-means-clustering-model-in-6-steps-with-python-35b532cfa8ad>

11. <https://matplotlib.org/stable/api/_as_gen/matplotlib.pyplot.plot.html>

12. <https://www.kdnuggets.com/2020/06/centroid-initialization-k-means-clustering.html>

These are all the references I used to complete kmeans clustering. I hope this doesnot come under plagiarism.

Task2:

PCA to reduce dimensions, k means and GMM application on the pca dataset, and plotting the graphs

I have used inbuilt libaries for all of these tasks

References for the same:

1. <https://www.geeksforgeeks.org/reduce-data-dimentionality-using-pca-python/#:~:text=Steps%20to%20Apply%20PCA%20in%20Python%20for%20Dimensionality%20Reduction&text=In%20this%20example%2C%20we%20will,the%20sklearn%20library%20of%20Python.&text=All%20the%20necessary%20libraries%20required,Python3>

2. <https://stackoverflow.com/questions/56317085/after-choosing-k-components-in-pca-how-do-we-find-out-which-componentsnames-of>

3. <https://www.kaggle.com/code/abheeshthmishra/eda-of-human-activity-recognition>

4. <https://www.geeksforgeeks.org/replacing-strings-with-numbers-in-python-for-data-analysis/>

5. <https://stackoverflow.com/questions/22984335/recovering-features-names-of-explained-variance-ratio-in-pca-with-sklearn>

6. <https://towardsdatascience.com/principal-component-analysis-pca-with-scikit-learn-1e84a0c731b0>

7. <https://www.kaggle.com/code/prashant111/k-means-clustering-with-python/notebook>

8. <https://machinelearningknowledge.ai/tutorial-for-k-means-clustering-in-python-sklearn/#:~:text=Step%2D1%3ATo%20decide%20the,new%20centroid%20for%20every%20cluster>.

9. <https://scikit-learn.org/stable/modules/generated/sklearn.mixture.GaussianMixture.html>

10. <https://towardsdatascience.com/pandas-df-to-numpy-array-c1a9e7d8585f>

11. <https://jakevdp.github.io/PythonDataScienceHandbook/05.12-gaussian-mixtures.html>