

CS305 Exercise 8

Task 1: Clustering

For some machine learning methods feature scaling is important and for other machine learning methods it is not. True or False: feature scaling is important prior to clustering.

- True
- False

True or False: if the hierarchical clustering algorithm is re-run, it will necessarily yield the same clustering result.

- True
- False

True or False: if the k -means clustering algorithm is re-run, it will necessarily yield the same clustering result.

- True
- False

Which one of the following might be an appropriate application of clustering?

- We wish to classify labeled data where each example corresponds to one of k different classes. We have reason to believe that our data are normally distributed rather than linear.
- As a music company, we want to recommend new songs to our users. For a given user, we plan to find the k most similar other users and make song recommendations based on the music preferences of these k similar customers.
- We have data that are labeled not with a discrete number of classes but with continuous values and we wish to predict labels for new data as they become available to us.
- We have a classification problem but there are a large number of similar features and our classification algorithm is taking a long time to execute. To speed up the classification algorithm, we combine similar features into meta-feature groups and execute our classification algorithm on the new smaller set of meta-features where each meta-feature corresponds to one of the groups.

Task 2: Ethics of Artificial Intelligence

Please read [this article](#) about ethical considerations in artificial intelligence.

Please address (minimum 100 words) the following questions. As the article describes, "many machine-learning systems are now essentially black boxes; their creators know they work, but they can't explain exactly why they make particular decisions." In other words, they are not *interpretable*. Consider many of the machine learning applications mentioned by the article, e.g., image processing, criminal justice, hiring, healthcare, military. Are there machine learning applications where you would sacrifice accuracy for interpretability? Are there machine learning applications where you would sacrifice interpretability for accuracy? The article concludes with the quote "Are there some things that we just shouldn't build?" What is your response to this question?

Download the Jupyter Notebook for Exercise 8 from the course website. Open the Notebook in your web browser and work through it. As you work through the Notebook, answer the following questions.

Task 3: Image Compression

Exactly how many unique colors are in the panda image?

Do you observe a significant difference in the quality of the image when 256 colors are used rather than ~100,000?

In the new image that uses a reduced number of colors, calculate exactly how many different colors occur in the image, as you did for the original image. Exactly how many unique colors are in the “reduced colors” panda image?

Do you observe a significant difference in the quality of the image when 10 colors are used rather than ~100,000?

Task 4: Choosing Clustering Parameter k

Looking at the elbow plot you created, what is the elbow point, i.e., what might be a good value to try for parameter k when clustering the data?

Task 5: Iris Species

Looking at the elbow plot you created, what value(s) for the number of clusters would you consider to be the elbow point?

Does your clustering, based on the elbow point, correspond poorly or well or exactly with the known species of iris?

CS305 Exercise 8 Final Page

Name(s): _____

In the *TIME* column, please estimate the time you spent on this exercise. Please try to be as accurate as possible; this information will help us to design future exercises.

PART	TIME	SCORE
Exercise		