Reading & Discussion: Chapter 5 - Machine Learning (PCA Pages 433-445)

TASK

Your task for this discussion is to read [Chapter 5 (PCA Pages 433-445)Links to an external site.](https://jakevdp.github.io/PythonDataScienceHandbook/05.09-principal-component-analysis.html). This can be found in the link provided or the [pdf version of the textbook](https://elearning.mines.edu/courses/52392/files/5459013/download?wrap=1)[Download pdf version of the textbook](https://elearning.mines.edu/courses/52392/files/5459013/download?download_frd=1)for the class that we have provided. You will then post a discussion and a reply below!

DISCUSSION REQUIREMENTS

Once you have completed this weeks reading, you are asked to complete the following tasks:

1. Post a 1-2 sentence response from the reading for 2 of the three below prompts:

a. How does PCA differ from Linear Regression?

b. What does using PCA for dimensionality reduction do?

c. Provide a free-response: this can be anything from clarification questions, something that piqued your interest, or maybe a personal experience you have with what was discussed in the book section.

2. Post a meaningful reply to another student's question/post.

Our goal is to help you get a deeper understanding of data science and discussions are a wonderful way to do this! Think of this as a learning opportunity for both you, your classmates, and your professor! Be sure to use academic language and question ideas from the readings for clarity or to push your own thinking.  You may of course respond to more than two of the questions and more of your classmates if you would like.

Reading & Discussion: Chapter 5 - Machine Learning (k-Means Pages 462-476)

TASK

Your task for this discussion is to read [Chapter 5 (K-Means Pages 462-476)Links to an external site.](https://jakevdp.github.io/PythonDataScienceHandbook/05.11-k-means.html). This can be found in the link provided or the [pdf version of the textbook](https://elearning.mines.edu/courses/52392/files/5459013/download?wrap=1)[Download pdf version of the textbook](https://elearning.mines.edu/courses/52392/files/5459013/download?download_frd=1)for the class that we have provided. You will then post a discussion and a reply below!

DISCUSSION REQUIREMENTS

Once you have completed this weeks reading, you are asked to complete the following tasks:

1. Post a 1-2 sentence response from the reading for 2 of the three below prompts:

a. What is the cluster center and what does it do (in relation to the data points)?

b. How do M and E steps improve the K-Means algorithm?

c. Provide a free-response: this can be anything from clarification questions, something that piqued your interest, or maybe a personal experience you have with what was discussed in the book section.

2. Post a meaningful reply to another student's question/post.

Our goal is to help you get a deeper understanding of data science and discussions are a wonderful way to do this! Think of this as a learning opportunity for both you, your classmates, and your professor! Be sure to use academic language and question ideas from the readings for clarity or to push your own thinking.  You may of course respond to more than two of the questions and more of your classmates if you would like.

# **Instructional Videos & Learning Materials: Unsupervised Learning & Clustering**

## WATCH

Watch this week's videos, download the working files so you can follow along and experiment with the platform we'll be using for the course.

**Unsupervised Learning Video**

**Clustering Video**

WORKING FILES

You will need the following files to get set up for this week. They include the following:

* Jupyterhub notebooks (download to a personal device then upload to your Jupyterhub). Also, download the images and place them in the same directory in your Jupyterhub notebooks.
  + Unsupervised Learning
    - [Unsupervised Learning Jupyter Notebook](https://elearning.mines.edu/courses/52392/files/5459065/download?wrap=1)[Download Unsupervised Learning Jupyter Notebook](https://elearning.mines.edu/courses/52392/files/5459065/download?download_frd=1)
    - [Supporting Image](https://elearning.mines.edu/courses/52392/files/5458994/download?wrap=1)[Download Supporting Image](https://elearning.mines.edu/courses/52392/files/5458994/download?download_frd=1)
    - [Supporting File](https://elearning.mines.edu/courses/52392/files/5459002/download?wrap=1)[Download Supporting File](https://elearning.mines.edu/courses/52392/files/5459002/download?download_frd=1)
  + Clustering
    - [Clustering Jupyter Notebook](https://elearning.mines.edu/courses/52392/files/5459035/download?wrap=1)[Download Clustering Jupyter Notebook](https://elearning.mines.edu/courses/52392/files/5459035/download?download_frd=1)
    - [Supporting Image](https://elearning.mines.edu/courses/52392/files/5459170/download?wrap=1)[Download Supporting Image](https://elearning.mines.edu/courses/52392/files/5459170/download?download_frd=1)
  + PDF files that contain the same information as the Jupyterhub notebooks, in case your Juypterhub is not yet working.
    - [Unsupervised Learning PDF](https://elearning.mines.edu/courses/52392/files/5459160/download?wrap=1)[Download Unsupervised Learning PDF](https://elearning.mines.edu/courses/52392/files/5459160/download?download_frd=1)
    - [Clustering PDF](https://elearning.mines.edu/courses/52392/files/5459237/download?wrap=1)[Download Clustering PDF](https://elearning.mines.edu/courses/52392/files/5459237/download?download_frd=1)
  + Additional Information:
    - This page on [PCA Additional Information](https://elearning.mines.edu/courses/52392/pages/pca-additional-information) has some key points, definitions, and samples with graphs to reinforce the use of principal component analysis.

ADDITIONAL RESOURCES

The links below are supplementary resources to help you review or strengthen the topics we have discussed:

* [PCA Quick ExplanationLinks to an external site.](https://setosa.io/ev/principal-component-analysis/)
* [PCA More In-Depth ExplanationLinks to an external site.](https://georgemdallas.wordpress.com/2013/10/30/principal-component-analysis-4-dummies-eigenvectors-eigenvalues-and-dimension-reduction/)
* [PCA in SciKit-learnLinks to an external site.](https://stackabuse.com/implementing-pca-in-python-with-scikit-learn)
* [NormalizationLinks to an external site.](https://developers.google.com/machine-learning/data-prep/transform/normalization)
* [ClusteringLinks to an external site.](https://www.analyticsvidhya.com/blog/2019/08/comprehensive-guide-k-means-clustering/)
* [K-Means ClusteringLinks to an external site.](https://www.analyticsvidhya.com/blog/2020/10/a-simple-explanation-of-k-means-clustering/)
* [Quick Hierarchical ClusteringLinks to an external site.](https://www.displayr.com/what-is-hierarchical-clustering/#:~:text=Hierarchical%20clustering%2C%20also%20known%20as,broadly%20similar%20to%20each%20other.)
* [Detailed Hierarchical ClusteringLinks to an external site.](https://www.analyticsvidhya.com/blog/2019/05/beginners-guide-hierarchical-clustering/)