**Self-Study Knowledge Check 6.1: SVD and PCA**

**Question 11 pts**

PCA is used for clustering.

Group of answer choices



True



False

[Flag question: Question 2](https://classroom.emeritus.org/courses/1960/quizzes/95384/take)

**Question 21 pts**

PCA looks for linear combinations of existing (blank) that capture the bulk of the variance.

Group of answer choices



Columns



Rows



Tuples



Sets

[Flag question: Question 3](https://classroom.emeritus.org/courses/1960/quizzes/95384/take)

**Question 31 pts**

The “curse of dimensionality” states that the amount of data you need to train a model increases exponentially with the number of inputs.

Group of answer choices



True



False

[Flag question: Question 4](https://classroom.emeritus.org/courses/1960/quizzes/95384/take)

**Question 41 pts**

What does running SVD on X decompose X into?

Group of answer choices



Σ V



U



U Σ



U Σ V

[Flag question: Question 5](https://classroom.emeritus.org/courses/1960/quizzes/95384/take)

**Question 51 pts**

What is the formula to normalize a dataset X?

Group of answer choices



Xnorm = (X−μ)



Xnorm = (X−μ)/σ



Xnorm = X/σ

[Flag question: Question 6](https://classroom.emeritus.org/courses/1960/quizzes/95384/take)

**Question 61 pts**

What is the function in Python library “scipy.linalg” to compute the singular value decomposition?

Group of answer choices



Svd()



single\_value\_decomp()



s\_v\_d()



svd()

[Flag question: Question 7](https://classroom.emeritus.org/courses/1960/quizzes/95384/take)

**Question 71 pts**

The Python function “numpy.allclose()” is used to find whether two arrays are element-wise equal.

Group of answer choices



True



False

**Self-Study Knowledge Check 6.2: Interpretation of PCA**

**Question 11 pts**

In SVD, the matrix sigma has all the diagonal values as zero.

Group of answer choices



True



False

[Flag question: Question 2](https://classroom.emeritus.org/courses/1960/quizzes/95400/take)

**Question 21 pts**

Below is the equation to represent the multiplication of the matrices for SVD:  
  
∑Di=1 σiuivit

What does the parameter *D*represent?

Group of answer choices



Columns



Matrices



Principal components



Rows

[Flag question: Question 3](https://classroom.emeritus.org/courses/1960/quizzes/95400/take)

**Question 31 pts**

The Matrix Σ in SVD has values “σi” in the diagonal of the matrix, which represent the importance of the i’th component for the dataset.

Group of answer choices



True



False

[Flag question: Question 4](https://classroom.emeritus.org/courses/1960/quizzes/95400/take)

**Question 41 pts**

The formula used to project the data into desired dimensions is  
“ x̄rr  = Ur∑r ”  
The parameter “r” is defined as the total number of columns.

Group of answer choices



True



False

[Flag question: Question 5](https://classroom.emeritus.org/courses/1960/quizzes/95400/take)

**Question 51 pts**

What is the symbol used for matrix multiplications?

Group of answer choices



@



\*



~



x

# Self-Study Knowledge Check 6.3: Clustering and K-Means

**Question 1**

**1 / 1 pts**

Clustering is a method for creating groups out of the columns of a dataset.



True

**Correct!**



False

*You are correct! The answer “False” is correct because clustering is a method for creating groups out of the rows of a dataset.*

**Question 2**

**1 / 1 pts**

Clustering is an unsupervised machine learning model.

**Correct!**



True

*You are correct! The answer “True” is correct because clustering has no labeled datasets.*



False

**Question 3**

**1 / 1 pts**

How is the centroid of a cluster “k” in k-means clustering represented?

**Correct!**



μk

*You are correct! The answer “*μk*” is correct because the mean of each cluster is declared as the centroid of that cluster, which, for cluster k, is “*μk*”.*



Σk



σk

**Question 4**

**1 / 1 pts**

In the K-means clustering algorithm, how is inertia defined?



Integration of the squared distances from points to their centroids



Sum of the squared distances from points to points



Squared distances from points to their centroids

**Correct!**



Sum of the squared distances from points to their centroids

*You are correct! The answer “Sum of the squared distances from points to their centroids” is correct because inertia in K-means is the summation of the squared distances of data points from their respective centroids.*

**Question 5**

**1 / 1 pts**

The stepwise sequence for the K-means algorithm is as follows:

1. Assignment
2. Updation



True

**Correct!**



False

*You are correct! The answer “False” is correct because K-means starts with updating of centroids initially and then assigning data points to the nearest centroids.*

**Question 6**

**1 / 1 pts**

What is the stopping criteria for K-means clustering?



Updation step has a new centroid position

**Correct!**



Assignment step has no change of data points

*You are correct! The answer “*Assignment step has no change of data points*” is correct because the stopping criteria for K-means clustering is when assignment stops changing.*



Updation step has a new centroid



Assignment step has a change of data points

**Self-Study Knowledge Check 6.4: Clustering and K-Means**

**Question 1**

**1 / 1 pts**

Consider this dataframe:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Dataframe | | | | |
| **CustomerID** | **Gender** | **Age** | **Annual Income (k$)** | **Spending Score (1-100)** |
| 1 | Male | 19 | 15 | 39 |
| 2 | Male | 21 | 15 | 81 |
| 3 | Female | 20 | 16 | 6 |
| 4 | Female | 23 | 16 | 77 |
| 5 | Female | 31 | 17 | 40 |
| ... | ... | ... | ... | ... |
| 196 | Female | 35 | 120 | 79 |
| 197 | Female | 45 | 126 | 28 |
| 198 | Male | 32 | 126 | 74 |
| 199 | Male | 32 | 137 | 18 |
| 200 | Male | 30 | 137 | 83 |

What should the index of this dataframe be set to?



Spending Score (1–100)



Age

**Correct!**



CustomerID

*You are correct! The answer “*CustomerID*” is correct because the index of a dataframe should be a unique value for every row.*



Gender

**Question 2**

**1 / 1 pts**

In the Python function “KMeans()”, the constructor ‘‘init’’ is used to select the criteria for the initialization of data points.



True

**Correct!**



False

*You are correct! The answer “False” is correct because the constructor ‘init’ is used to select criteria for the initialization of the centroids of the clusters.*

**Question 3**

**1 / 1 pts**

In “KMeans()”, how do you generate the array that tells which cluster the data point belongs to?

**Correct!**



kmeans.labels\_

*You are correct! The answer “*kmeans.labels\_*” is correct because the statement is used to get an array that tells which data point belongs to which cluster.*



kmeans(labels\_)



kmeans(labels)



kmeans.labels

**Question 4**

**1 / 1 pts**

The default initialization in K-means is random initialization.



True

**Correct!**



False

*You are correct! The answer “False” is correct because the default initialization in K-means is improved initialization, which is K-means++.*

**Question 5**

**1 / 1 pts**

K-means++ only finds centroids once each time you run the function.



True

**Correct!**



False

*You are correct! The answer “False” is correct because K-means++ finds the initial centroids and then searches again in an attempt to lower the inertia of the dataset.*

**Self-Study Knowledge Check 6.5: DBSCAN**

**Question 11 pts**

The number of clusters for DBSCAN are declared beforehand.

Group of answer choices



True



False

[Flag question: Question 2](https://classroom.emeritus.org/courses/1960/quizzes/95323/take)

**Question 21 pts**

What is the clustering algorithm which has the ability to create curved boundaries between clusters?

Group of answer choices



DBSCAN



K-means

[Flag question: Question 3](https://classroom.emeritus.org/courses/1960/quizzes/95323/take)

**Question 31 pts**

Points that are sufficiently removed from other points are designated by DBSCAN as (blank).

Group of answer choices



Outliers



Clusters



Inliers

[Flag question: Question 4](https://classroom.emeritus.org/courses/1960/quizzes/95323/take)

**Question 41 pts**

In the Python function “cluster.DBSCAN()”, the constructors of the function are (blank). *(Check all that apply.)*

Group of answer choices



min\_samples



init



eps



k

[Flag question: Question 5](https://classroom.emeritus.org/courses/1960/quizzes/95323/take)

**Question 51 pts**

If the ball of radius epsilon captures less than min\_sample points, then that point is designated as a core point.

Group of answer choices



True



False

[Flag question: Question 6](https://classroom.emeritus.org/courses/1960/quizzes/95323/take)

**Question 61 pts**

How does DBSCAN declare a point as an outlier?

Group of answer choices



Points with at least min\_samples points in their epsilon ball radius



Points with no core or boundary points in their epsilon ball radius



Points that are not core points but that contain at least one other core point in their epsilon