# More about inertia and silhouette coefficient metrics

You know that the evaluation metrics you used for supervised learning models don't apply to unsupervised learning models. This is because unsupervised learning model results cannot be categorized as "correct" or "incorrect." While supervised learning models use predictor variables to predict a defined target variable, unsupervised learning methods have metrics that seek an underlying structure within the data.

Clustering models are a type of unsupervised learning that do this by grouping observations together. Data professionals often use inertia and silhouette scores to evaluate their clustering models and help them determine which groupings make sense. This reading reviews these concepts and examines them in greater detail.

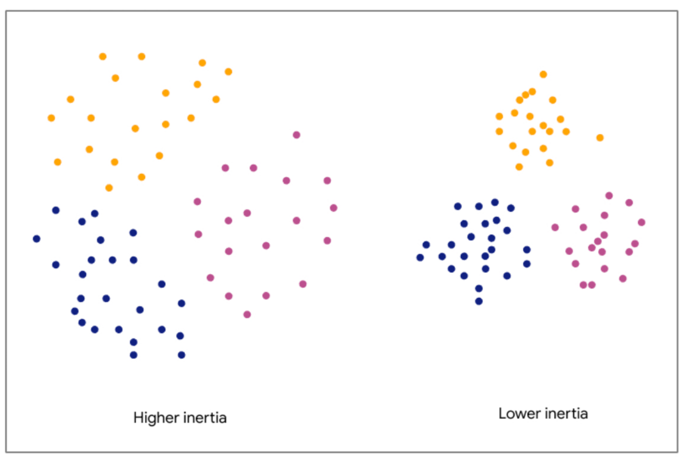
## Inertia

**Inertia** is a measurement of intracluster distance. It indicates how compact the clusters are in a model. Specifically, inertia is the sum of the squared distance between each point and the centroid of the cluster that it's assigned to. It can be represented by this formula, where:

- $n$ = the number of observations in the data,
- $x_i$= the location of a particular observation,
- $C_k$ = the location of the centroid of cluster $k$, which is the cluster to which point $x_i$is assigned.

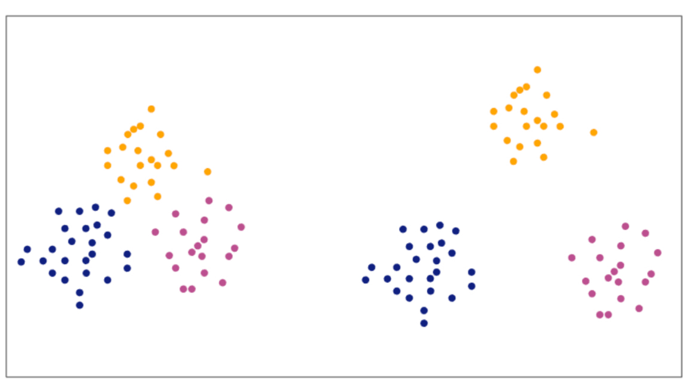
$$\text{Inertia} = \sum_{i=1}^{n}(x_i - C_k)^2$$

The greater the inertia, the greater the distances between points and their centroids, which means the points within each cluster are farther apart from each other. In the following figure, the three clusters on the left have higher inertia than the three clusters on the right, because they are less compactly positioned around their respective centroids.




**The three clusters on the left have higher inertia than the three on the right.**

Note, however, that inertia only measures intracluster distance. Therefore, both of the clusterings in the figure below have the same inertia.



**The three clusters on the left have the same inertia as the three on the right.**

For the same dataset and the same number of clusters, lower inertia values are typically better than higher values, because low values indicate that points are closer together within their clusters.

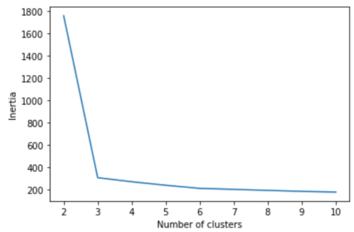
Why is it more meaningful for points in a dataset to be close together within their clusters? Well, notice that you're trying to make sense of data by identifying observations that are related to each other somehow. Clustering models do this by grouping points together based on their similarity. For some techniques, "similarity" is about the distance between points themselves (and so points that are close to each other belong to the same cluster). With others, it's about the distance between points and a cluster center (such that points that are close to the same center belong to the same cluster)—like K-means. In both cases, observations that are closer together are assumed to be more similar to each other. In other words, a tighter cluster could be indicative of greater similarity between the real-world observations that are represented by those data points.

**Evaluating inertia**

Inertia is a useful metric to determine how well your clustering model identifies meaningful patterns in the data. But it's generally not very useful *by itself.* If your model has an inertia of 53.25, is that good? It depends. The measurement becomes meaningful when it's compared to the inertia values and *k* values of other models on the same data. As you increase the number of clusters *(k)*, the inertia value will drop, but there comes a point where adding more clusters will have only small changes in inertia. And it's this transition that we need to detect.

**The elbow method**

The elbow method is a great way to find this point of transition. It's a way to help decide which clustering gives the most meaningful model of your data. It uses a line plot to visually compare the inertias of different models. With K-means models, this is done as a comparison between different values of *k*. Here's an example:



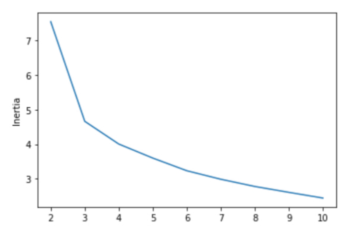

This plot compares the inertias of nine different K-means models—one for each value of *k* from two through 10. It's clear that inertia begins very high when the data is grouped into two clusters. The three-cluster model, however, has much lower inertia, creating a steep negative slope between two and three clusters. After that, the rate of inertial decline slows down dramatically, as indicated by the much flatter line in the plot.

To use the elbow method to evaluate a model, you want to find the part of the curve that looks like an elbow—the sharpest bend in the curve. That's usually the model that will give you the most meaningful clustering of your data, because if inertia is dropping significantly with added clusters, it means distance between points and their centroids is shortening significantly. This implies denser clusters and thus more similarity between points in a given cluster. But when the drop in inertia is very minor, the distance between points and their centroids isn't changing much, while your model is becoming more complex due to the additional cluster. Adding clusters is not capturing real structure in the data. A model with *k* = 1,000 will have low inertia, but is it any good? What meaning can you take away from 1,000 clusters?

Remember, you want inertia to be low, but if you add more and more clusters with only minimal improvement to inertia, you're only adding complexity without capturing real structure in the data.

**There's not always an obvious elbow.**

In the last example, the elbow was very clear. Often it won't be so obvious. Consider this, for example:




In this case, it seems that the elbow occurs at the three-cluster model, but there's still a considerable decline in inertia from three to four clusters. There's not necessarily a "correct" answer. It might be worth doing some analysis on the cluster assignments of both models to check for yourself which is more meaningful. There are also other tools at your disposal to help you decide.

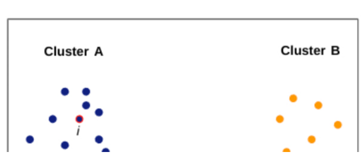
## Silhouette analysis

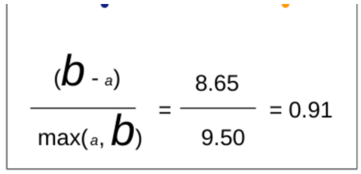
One of these tools is a silhouette analysis. A **silhouette analysis** is the comparison of different models' silhouette scores. To calculate a model's silhouette score, first, a silhouette coefficient is calculated for each instance in the data. An instance's silhouette coefficient is defined by the following formula, where:

- $a$ = the mean distance between the instance and each other instance in the same cluster
- $b$ = the mean distance from the instance to each instance in the nearest other cluster (i.e., excluding the cluster that the instance is assigned to)
- $max(a, b)$ = whichever value is greater, $a$ or $b$

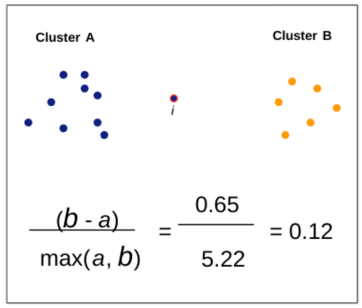
$$\text{Silhouette coefficient} = \frac{(b-a)}{max(a,b)}$$

A silhouette coefficient can range between -1 and +1. A value closer to +1 means that a point is close to other points in its own cluster and well separated from points in other clusters.

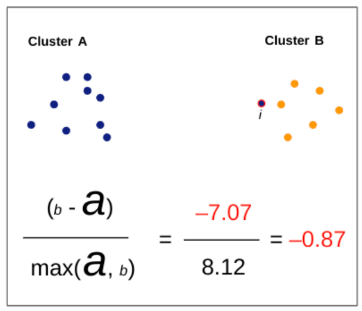





A value closer to zero means that a point is between clusters.

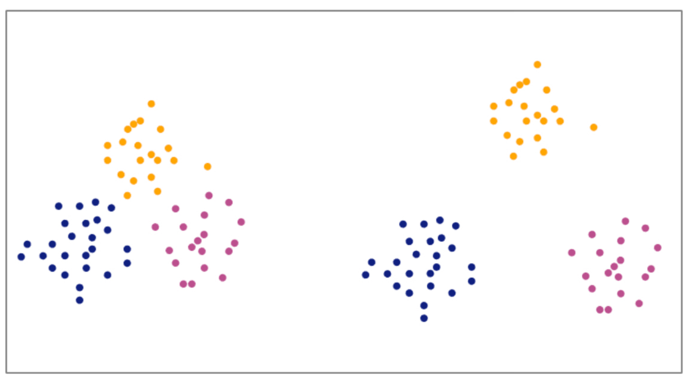



And a value closer to -1 means that a point is probably assigned to the wrong cluster, because it is closer to the points of another cluster than to the points in its own.



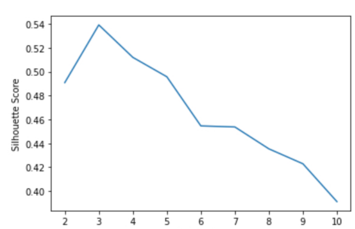

The **silhouette score** is the mean silhouette coefficient over all the observations in a model. The greater the silhouette score, the better defined the model clusters, because the points in a given cluster are closer to each other, and the clusters themselves are more separated from each other.

Note that, unlike inertia, silhouette coefficients contain information about both intracluster distance (captured by the variable *a*) and intercluster distance (captured by the variable *b*).



**The points in the three clusters on the left have lower silhouette coefficients than the points in the three on the right.**

As with inertia values, you can plot silhouette scores for different models to compare them against each other.




In this example, it's evident that a three-cluster model has a higher silhouette score than any other model. This indicates that this model results in individual clusters that are tighter and more separated from other clusters, more so than any other model tried. Based on this diagram, the data is probably best grouped into three clusters.

## Key takeaways

Inertia and silhouette score are useful metrics to help determine how meaningful your model's cluster assignments are. Both are especially helpful when your data has too many dimensions (features) to visualize in 2-D or 3-D space.

Use these metrics together to help inform your decision on which model to select.

**Inertia:**

- Measures intracluster distance
- Equal to the sum of the squared distance between each point and the centroid of the cluster that it's assigned to
- Used in elbow plots
- All else equal, lower values are generally better

**Silhouette score:**

- Measures both intercluster distance and intracluster distance
- Equal to the average of all points' silhouette coefficients
- Can be between -1 and +1 (greater values are better)

## Resources for more information

More detailed information about inertia and silhouette scores can be found here.

- scikit-learn documentation for silhouette_score
- Academic paper: Silhouettes: A graphical aid to the interpretation and validation of cluster analysis