Mentor Help
Ask a mentor on our Q&A platform

Clustering Recap

We just covered a bunch of information! Here is a quick recap!

I. Clustering

You learned about clustering, a popular method for unsupervised machine lea...
three ways to identify clusters in your dataset.

1. **Visual Inspection** of your data.
2. **Pre-conceived** ideas of the number of clusters.
3. **The elbow method**, which compares the average distance of each point different numbers of centers.

II. K-Means

You saw the k-means algorithm for clustering data, which has 3 steps:

1. Randomly place k-centroids amongst your data.

Then repeat the following two steps until convergence (the centroids don't cha...

2. Look at the distance from each centroid to each point. Assign each point
3. Move the centroid to the center of the points assigned to it.

III. Concerns with K-Means

Finally, we discussed some concerns with the k-means algorithm. These concer...

1. **Concern:** The random placement of the centroids may lead to non-optim...

Solution: Run the algorithm multiple times and choose the centroids that crea...
distance of the points to the centroids.

2. **Concern:** Depending on the scale of the features, you may end up with di...
your points.

Solution: Scale the features using Standardizing, which will create features wit...
deviation 1 before running the k-means algorithm.