Introduction

January 14, 2025

[ ]:
```
'''
   Clustering -->

   Clustering is similar to classification, but the basis is different.
   In Clustering you don't know what you are looking for, and you are trying␣
↪to identify
   some segments or clusters in your data. When you use clustering algorithms␣
↪on your dataset,
   unexpected things can suddenly pop up like structures, clusters and␣
↪groupings you would
   have never thought of otherwise.
'''
```

[ ]:
```
'''
   Key Characteristics of Clustering -->

   Unsupervised Learning: No labeled data; the goal is to find structure in␣
↪the data.
   Similarity: Data points in the same cluster are more similar to each other␣
↪than to those in other clusters.
   Use Cases: Market segmentation, image compression, social network analysis,␣
↪anomaly detection, etc.
'''
```

[ ]:
```
'''
   Popular Clustering Algorithms -->

   1. K-Means Clustering

   Concept: Partitions data into k clusters by minimizing intra-cluster␣
↪variance.

   Steps:
   Randomly initialize k centroids.
   Assign each data point to the nearest centroid.
   Update centroids by computing the mean of assigned points.
   Repeat until centroids stabilize.
```

```
   Pros: Simple, fast for large datasets.
   Cons: Requires k to be specified, sensitive to outliers.

   2. Hierarchical Clustering

   Concept: Builds a tree-like hierarchy of clusters (dendrogram).

   Types:
   Agglomerative: Starts with individual points and merges them iteratively.
   Divisive: Starts with all points in one cluster and splits them iteratively.

   Pros: Visualizes cluster relationships, no need to specify the number of␣
↪clusters.
   Cons: Computationally expensive for large datasets.
'''
```

[ ]:
```
'''
   Applications of Clustering -->

   Customer Segmentation: Group customers based on behavior for targeted␣
↪marketing.
   Anomaly Detection: Identify unusual patterns or outliers in data.
   Document Clustering: Group similar documents for information retrieval.
   Image Segmentation: Divide an image into meaningful parts.
   Genomics: Analyze gene expression data.
'''
```