How Hierarchical Clustering works ?

Define the essential terms

1. **Dendrogram :** A **dendrogram** is a tree-like diagram that visualizes the results of a hierarchical clustering algorithm. It shows the clusters that were merged (agglomerative) or split (divisive) and the distance at which the operation occurred. By "cutting" the dendrogram at different heights, you can obtain a different number of clusters.

2. **Agglomerative (Bottom-Up) Clustering :** This is the most common type of hierarchical clustering. It starts with each data point as its own cluster and then iteratively **merges the two closest clusters** until only one cluster remains. The "bottom-up" name comes from the way individual points are progressively joined together.

3. **Divisive (Top-Down) Clustering :**This method works in the opposite direction. It starts with all data points in a single cluster and then **recursively splits the most appropriate cluster** into two smaller clusters. This process continues until each data point forms its own individual cluster.

We will explain how **Agglomerative (Bottom-Up) Clustering works**

Define the essential terms

Single linkage, complete linkage, average linkage, and Ward's method are all different ways to measure the **distance between clusters** in **agglomerative hierarchical clustering**. This choice of distance metric, known as a **linkage criterion**, determines how the algorithm decides which clusters to merge at each step.

1. Single Linkage (Closest)
Concept: This method measures the distance between two clusters as the **shortest distance** between any single point in one cluster and any single point in the other cluster.
Behavior: It tends to form long, thin clusters and is sensitive to noise and outliers. It's often referred to as "minimum linkage."

2. Complete Linkage (Furthest)
Concept: This method measures the distance between two clusters as the **greatest distance** between any single point in one cluster and any single point in the other cluster.
Behavior: It tends to form compact, spherical clusters. It is less susceptible to noise but can be sensitive to outliers at the edges of the clusters. It's also known as "maximum linkage."

3. Average Linkage (All Pairs)
Concept: This method measures the distance between two clusters as the **average distance** between all pairs of points, where one point is from each cluster.
Behavior: This approach strikes a balance between single and complete linkage. It's less sensitive to noise than single linkage and less prone to creating spherical clusters than complete linkage.

4. Ward's Method (Variance)
Concept: This method measures the distance between two clusters based on the **increase in the total sum of squares** within the clusters after they are merged. It looks for the pair of clusters that, when combined, causes the smallest increase in variance.
Behavior: This method is effective at forming tight, compact clusters and is generally considered a good all-around choice. It is also more sensitive to outliers than average linkage.

Steps of Agglomerative hierarchical clustering ?

The working mechanism of Hierarchical Clustering can be described as follows:

1. **Initialization :** This is the starting point of the algorithm, where each data point is treated as its own individual cluster. The first step involves finding the two closest points to begin the merging process.

2. **Iterative Merging :** This is the core of the algorithm. It's a repetitive loop where the two closest clusters (which can be individual points or previously formed clusters) are merged into a new, larger cluster. This step continues until the stopping condition is met.

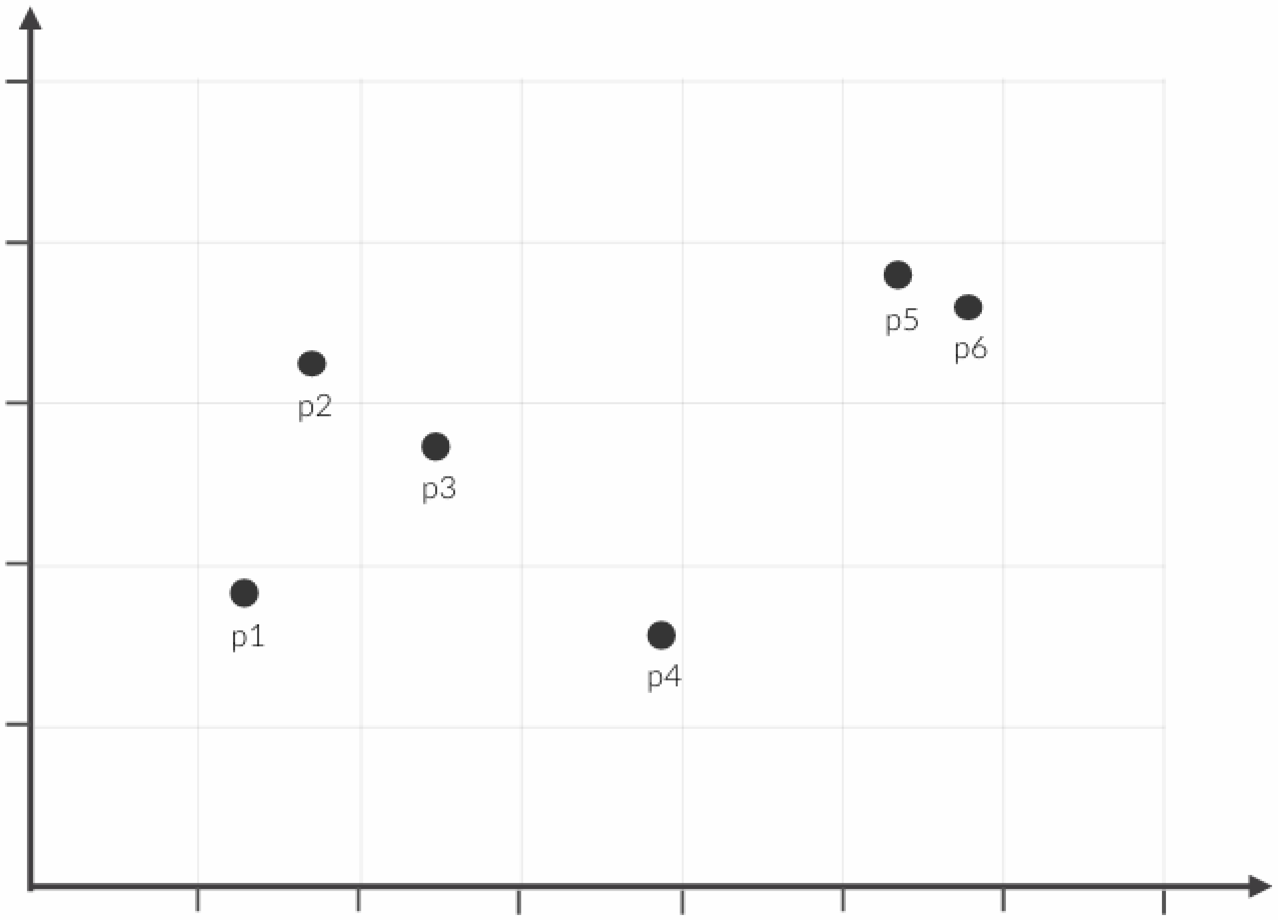
3. **Finalization :** The process concludes when all points have been merged into a single, unified cluster. The result is a completed hierarchy that can be visualized using a **dendrogram**.

Step 1 - Initialization

Step 1 - Initialization

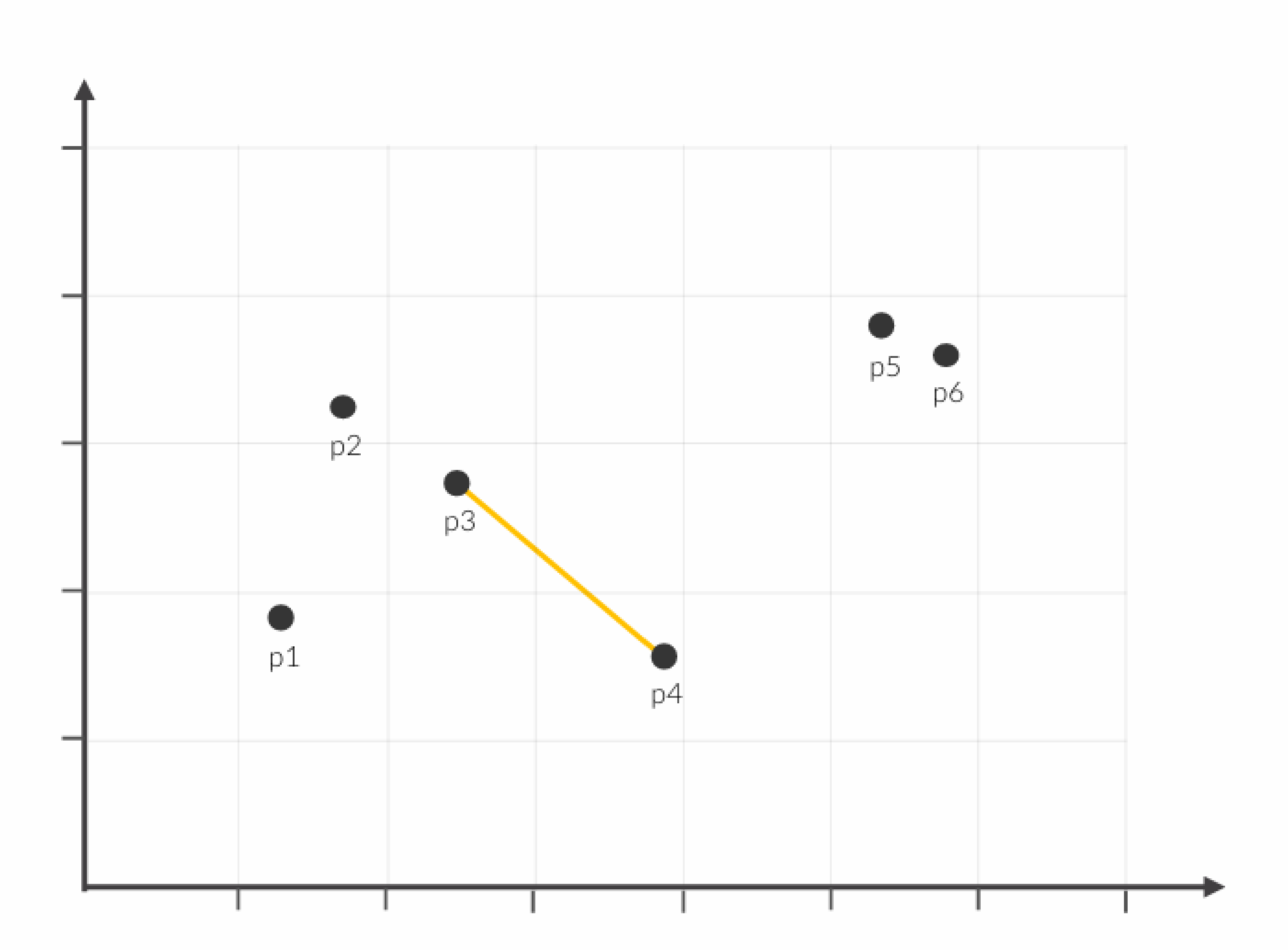
This is the starting point of the algorithm, where each data point is treated as its own individual cluster. The first step involves finding the two closest points to begin the merging process.

Step 1.1 - Find the two closest points, and group them into a cluster



How do you define "closest"?

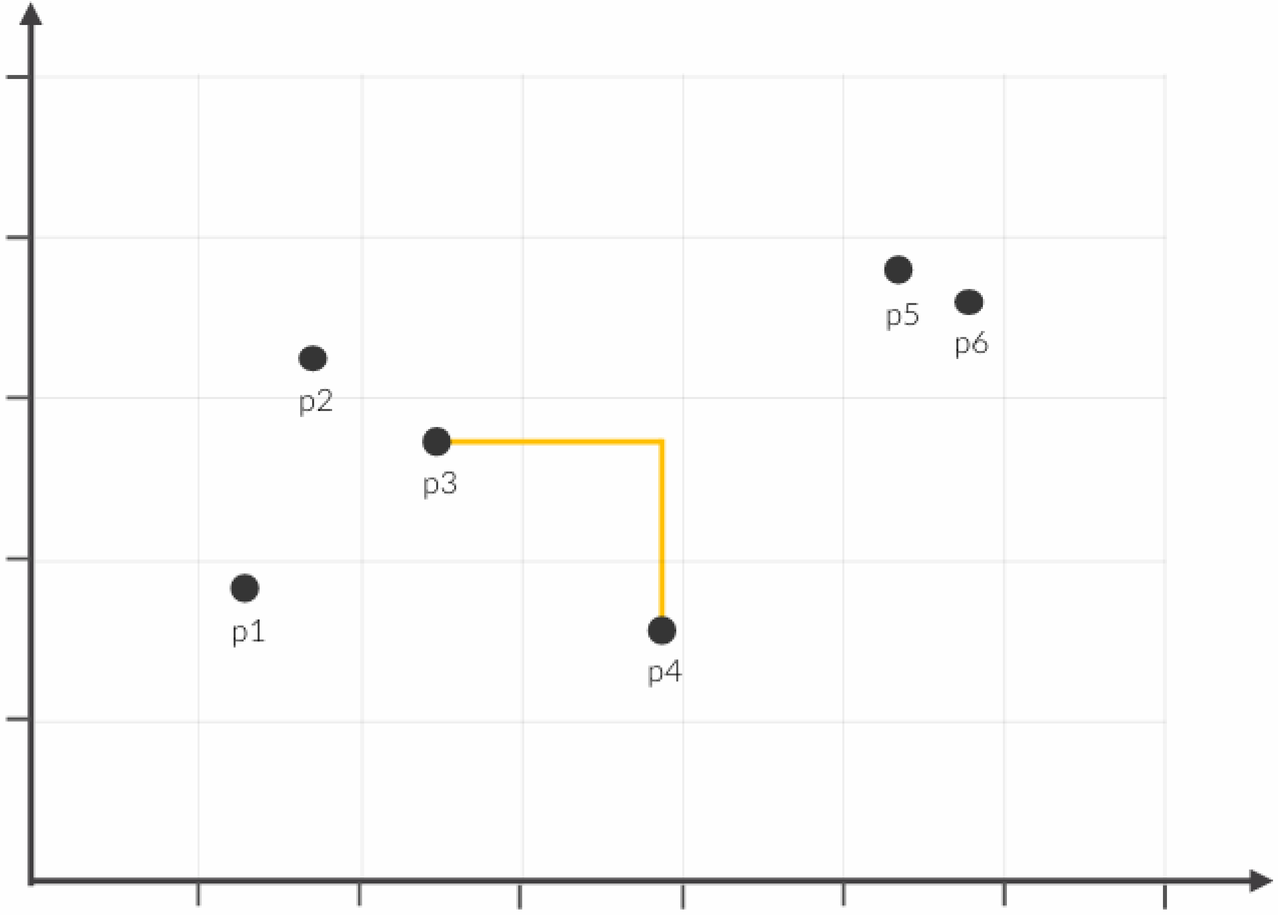
Step 1.2 - Find the two closest points, and group them into a cluster



How do you define "closest"?

- **Most commonly, the Euclidean distance is used**

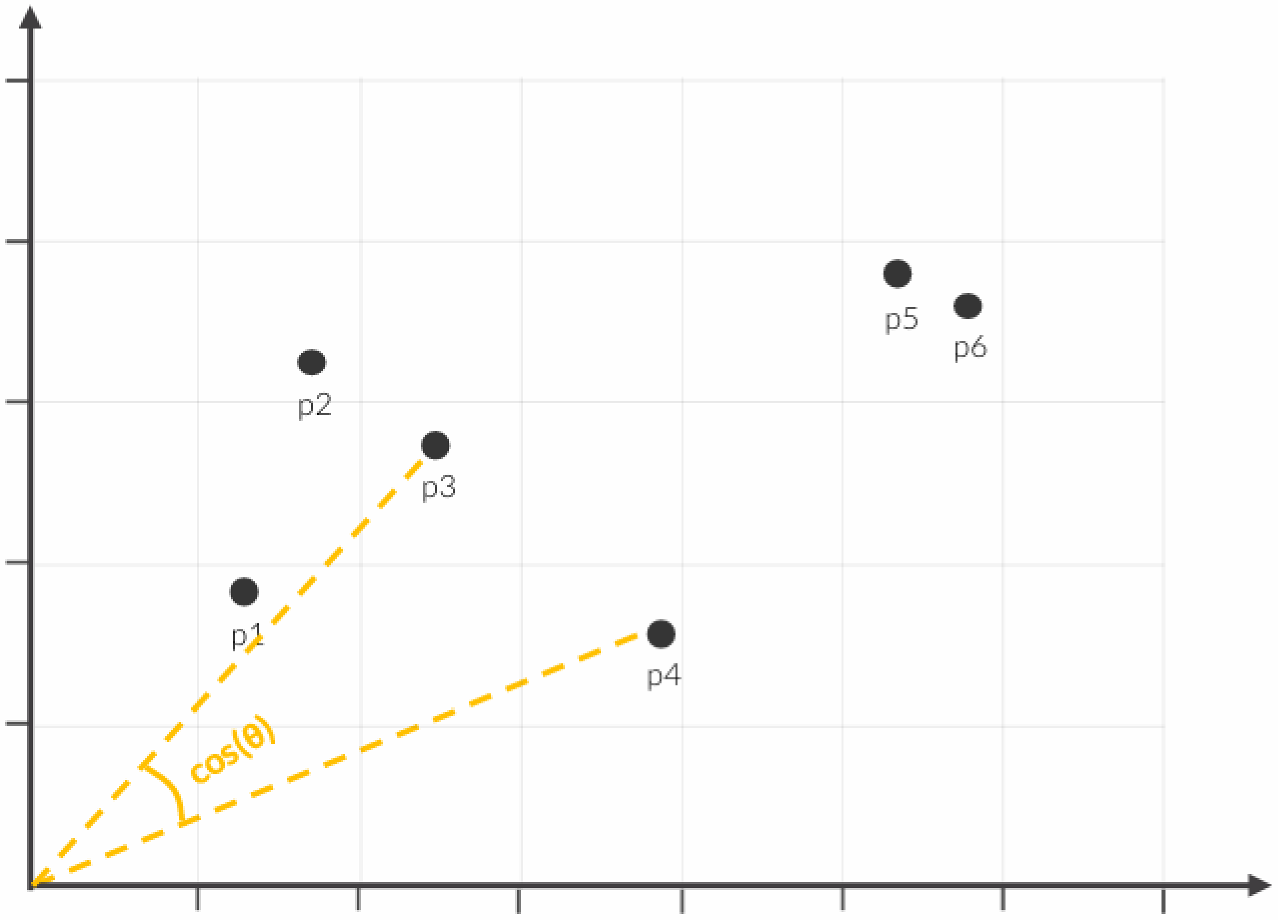
Step 1.3 - Find the two closest points, and group them into a cluster



How do you define "closest"?

- *Most commonly, the Euclidean distance is used*
- ***Alternatively, there's Manhattan distance***

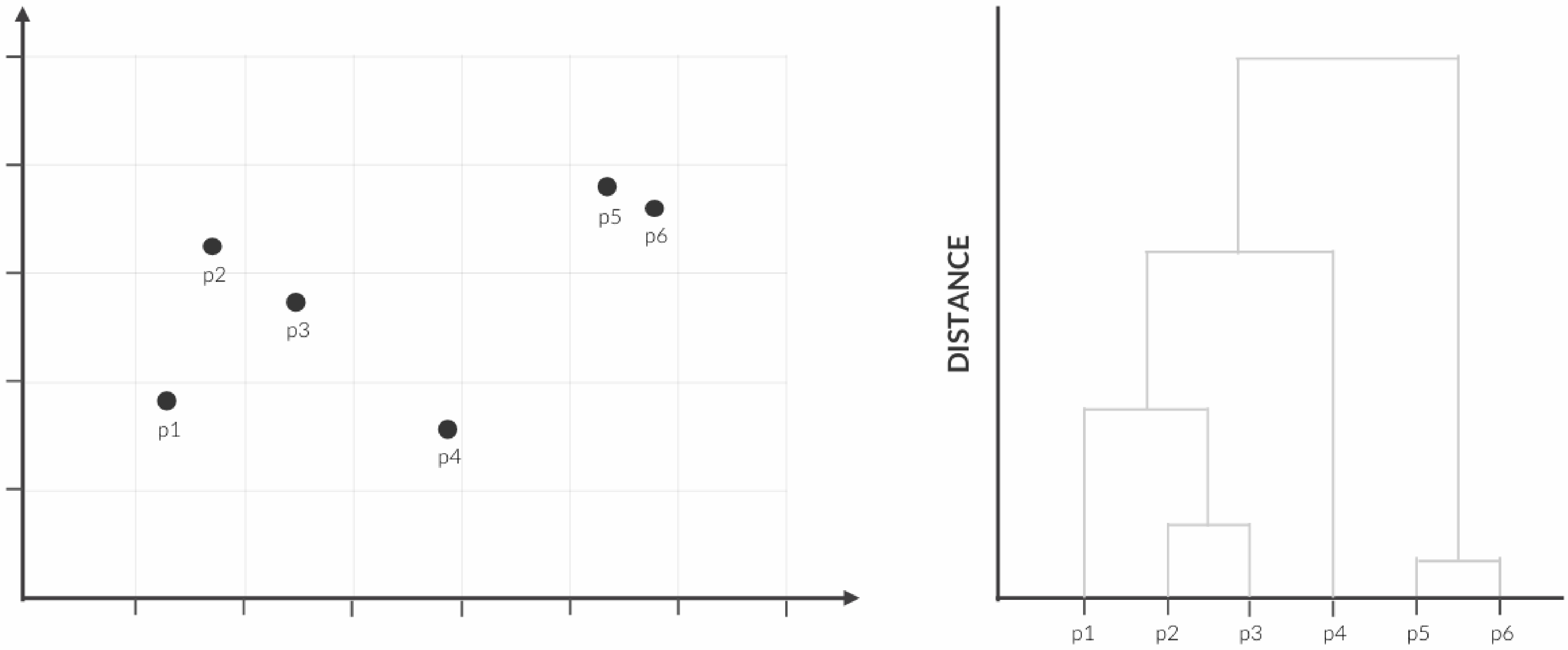
Step 1.4 - Find the two closest points, and group them into a cluster



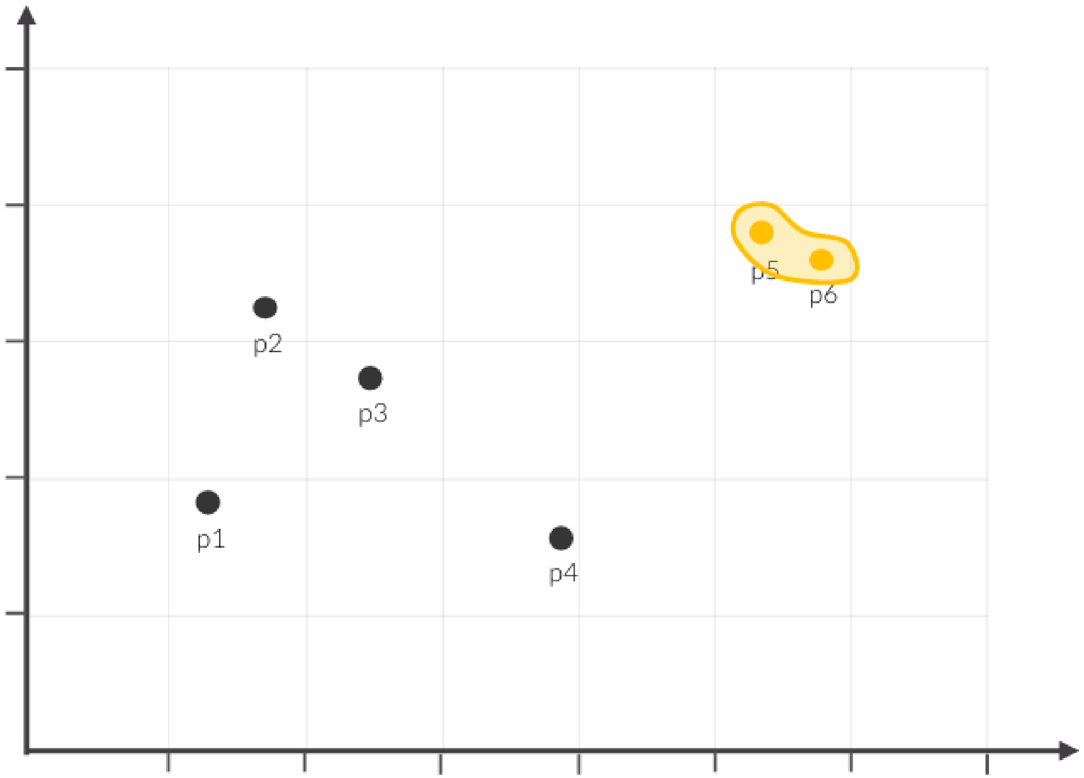
How do you define "closest"?

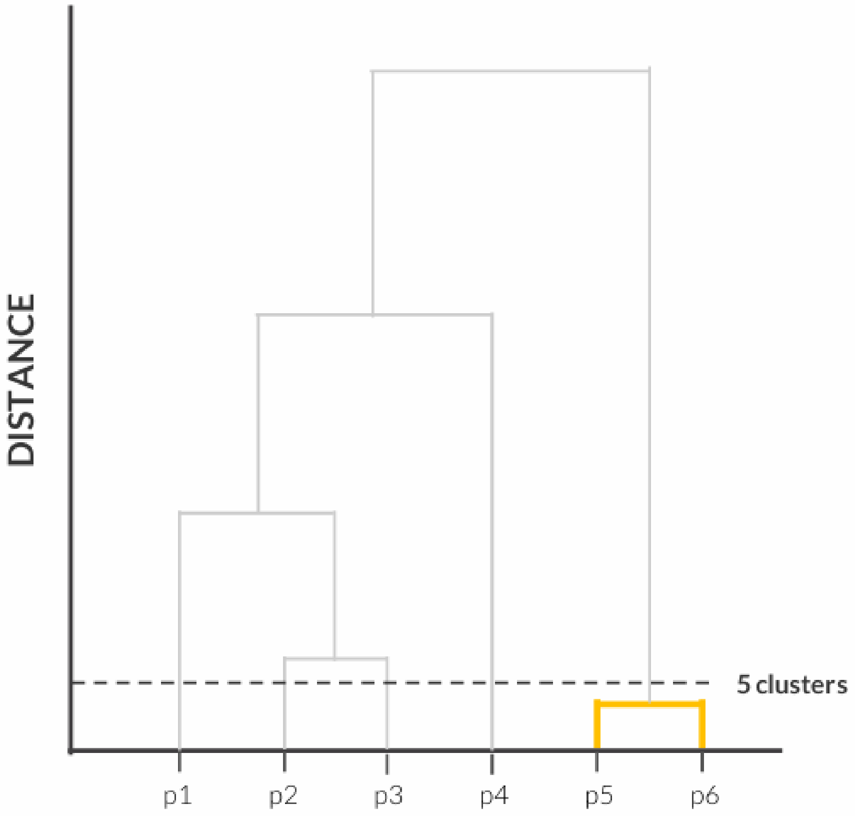
- *Most commonly, the Euclidean distance is used*
- *Alternatively, there's Manhattan distance*
- ***And Cosine distance***

Step 1.5 - Find the two closest points, and group them into a cluster



Step 1.6 - Find the two closest points, and group them into a cluster



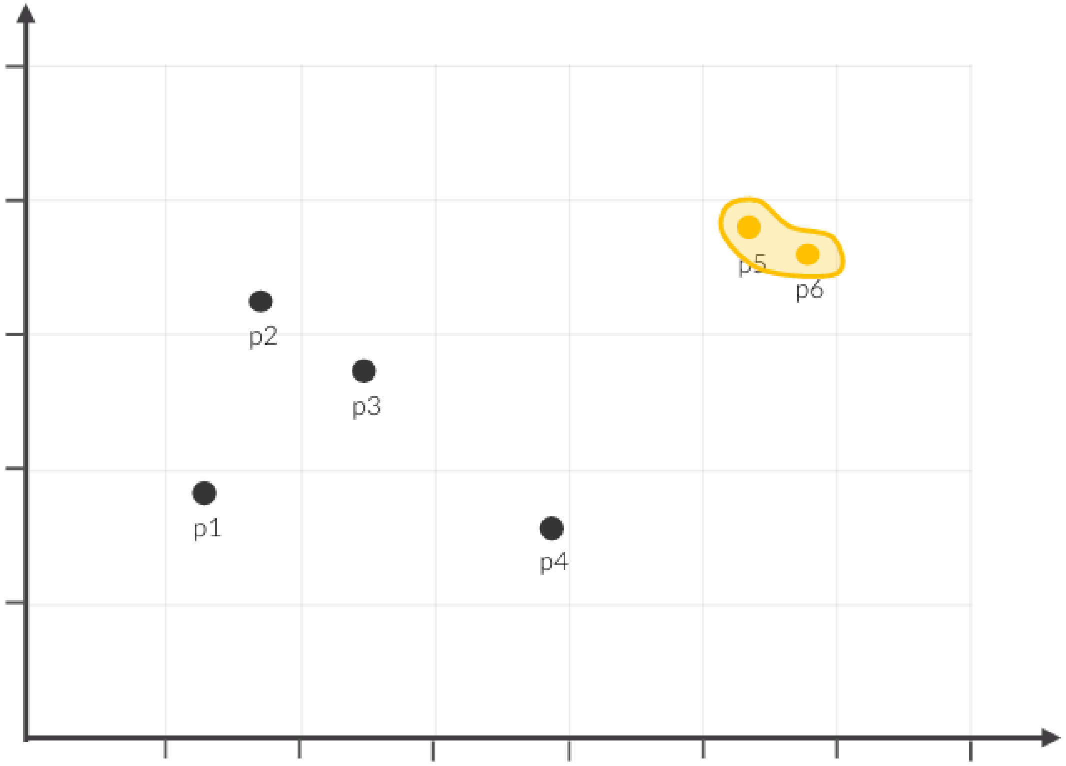


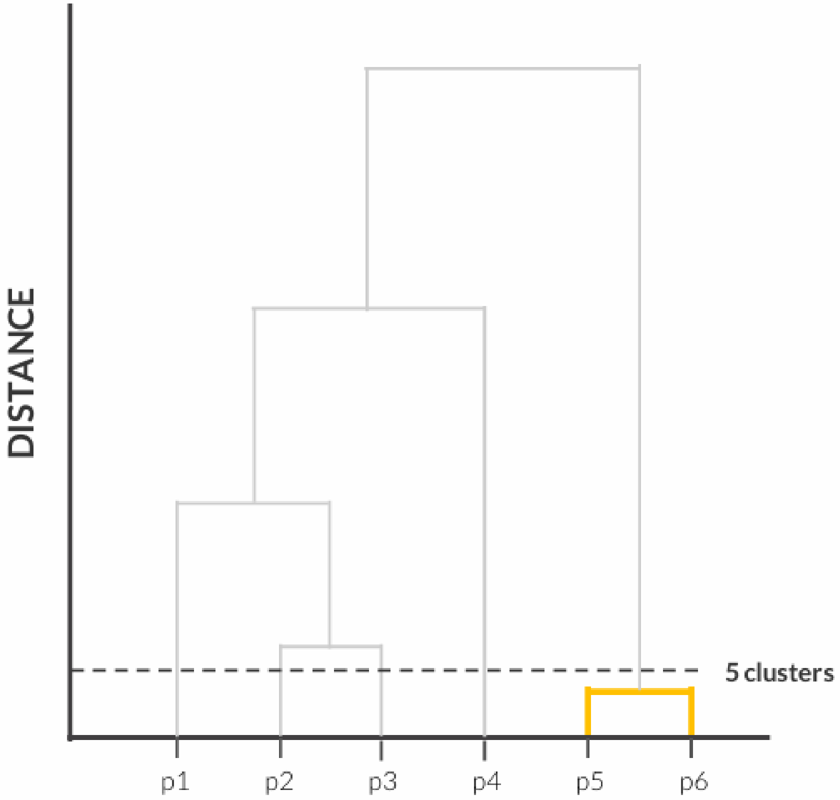
Step 2 - Iterative Merging

Step 2 – Iterative Merging

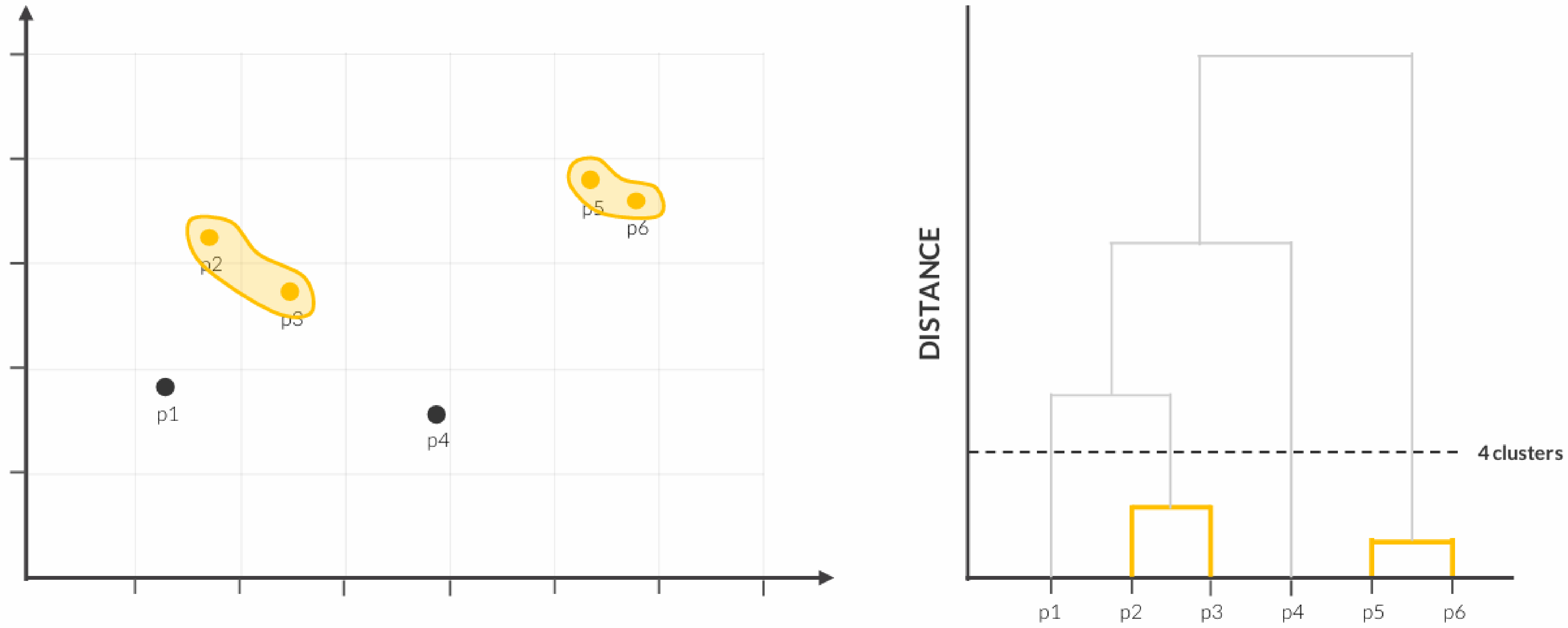
This is the core of the algorithm. It's a repetitive loop where the two closest clusters (which can be individual points or previously formed clusters) are merged into a new, larger cluster. This step continues until the stopping condition is met.

Step 2.1 - Find the next two closest points/clusters, and group them together





Step 2.2 - Find the next two closest points/clusters, and group them together

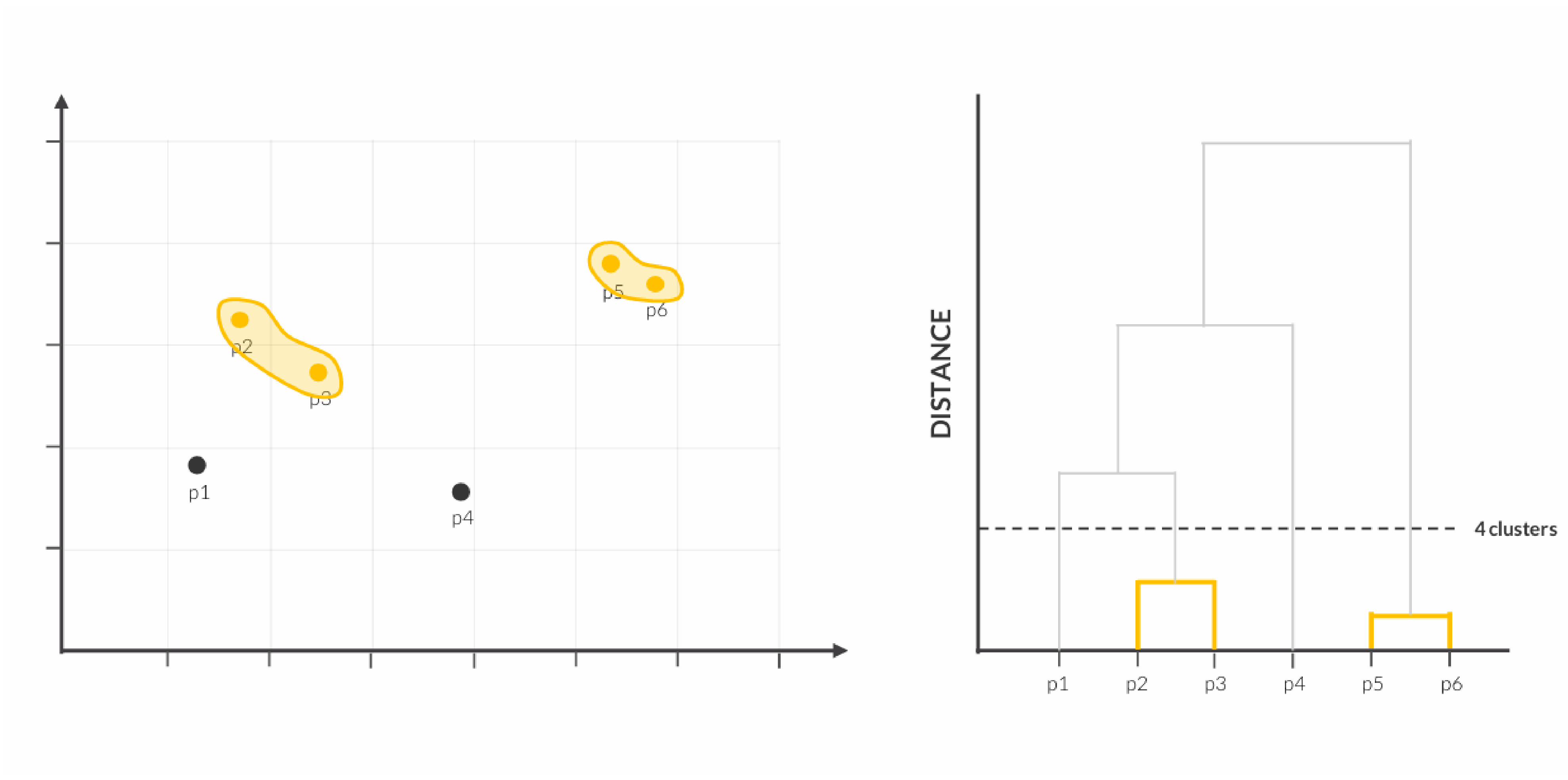


Step 3 - Finalization

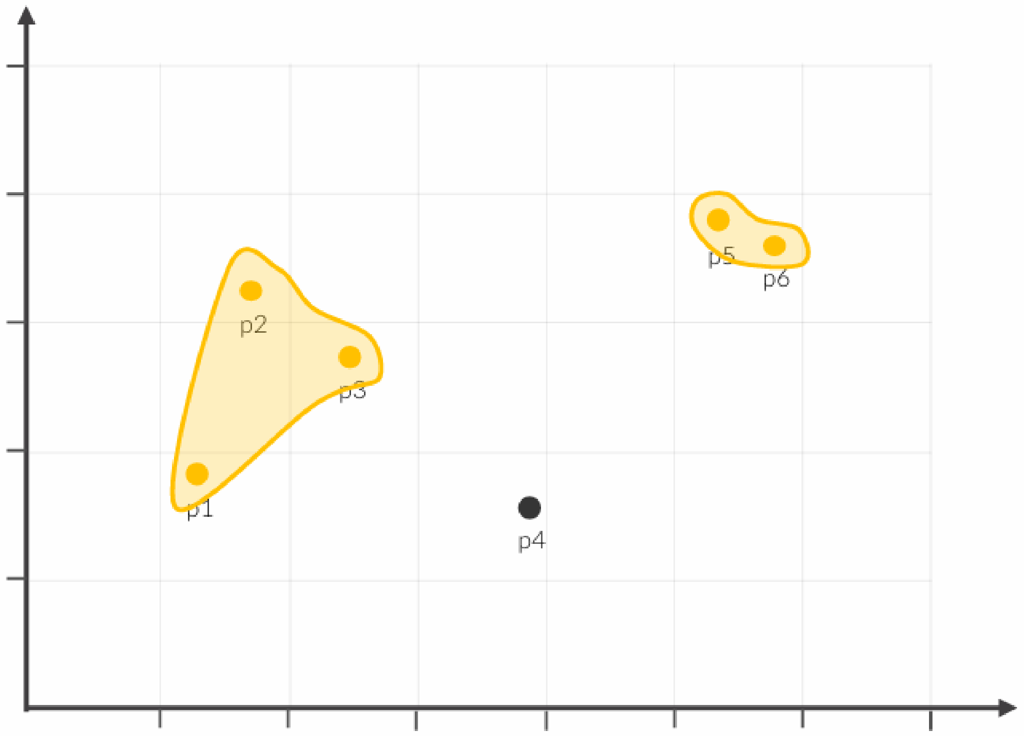
Step 3 – Finalization

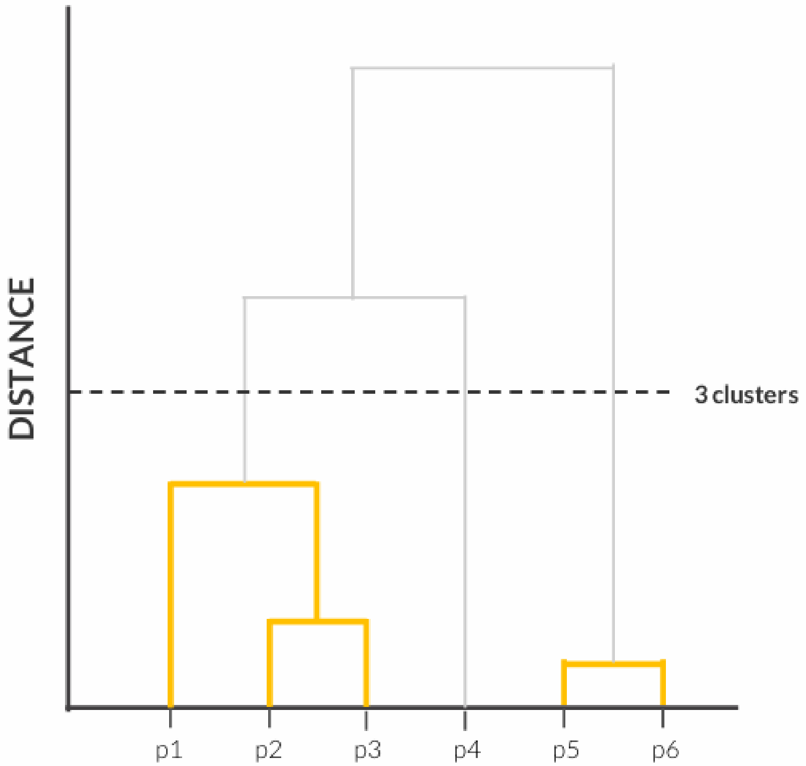
The process concludes when all points have been merged into a single, unified cluster. The result is a completed hierarchy that can be visualized using a **dendrogram**.

Step 3.1 - Repeat the process until all points are part of the same cluster

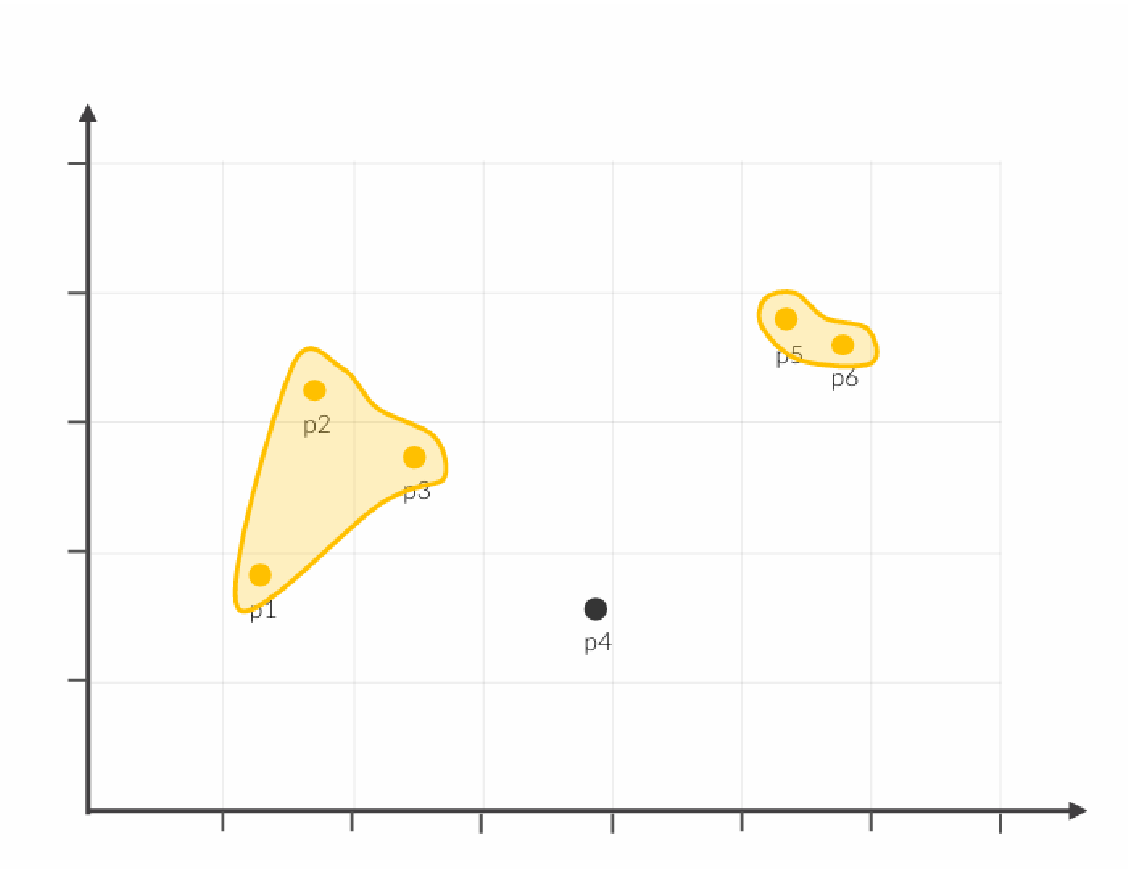


Step 3.2 - Repeat the process until all points are part of the same cluster



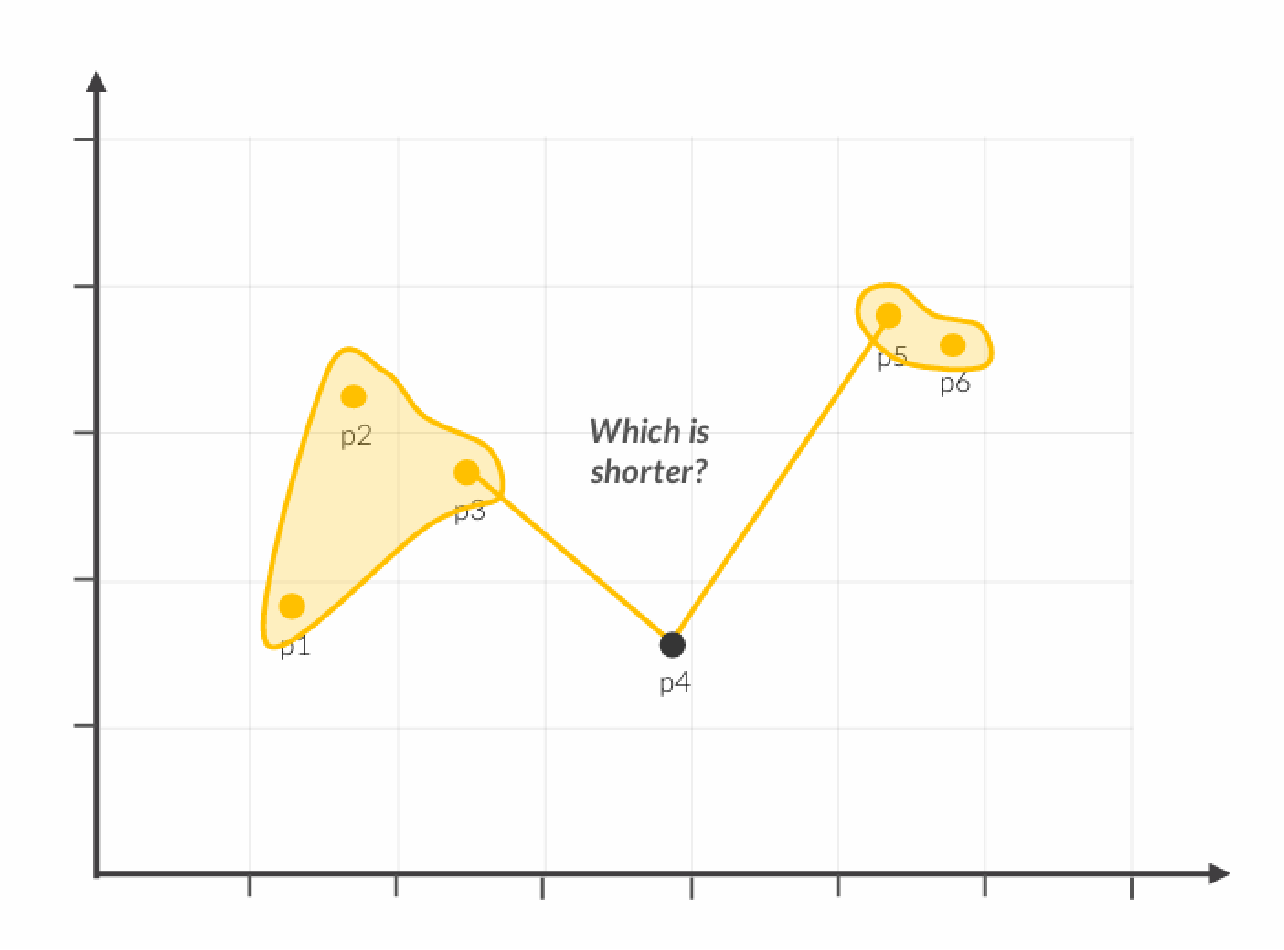


Step 3.3 - Repeat the process until all points are part of the same cluster



How do you define distance between clusters?

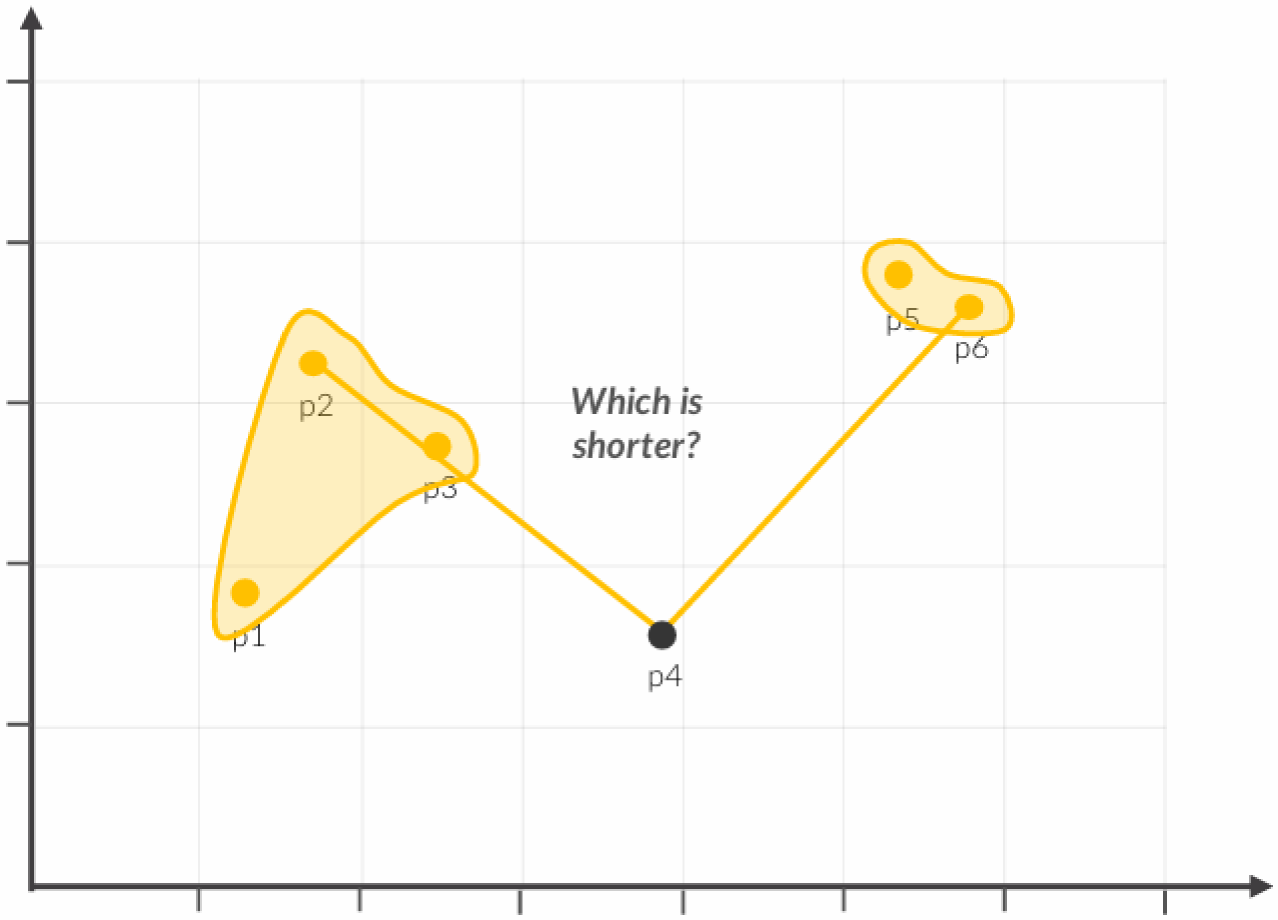
Step 3.4 - Repeat the process until all points are part of the same cluster



How do you define distance between clusters?

- **Single linkage (closest)**

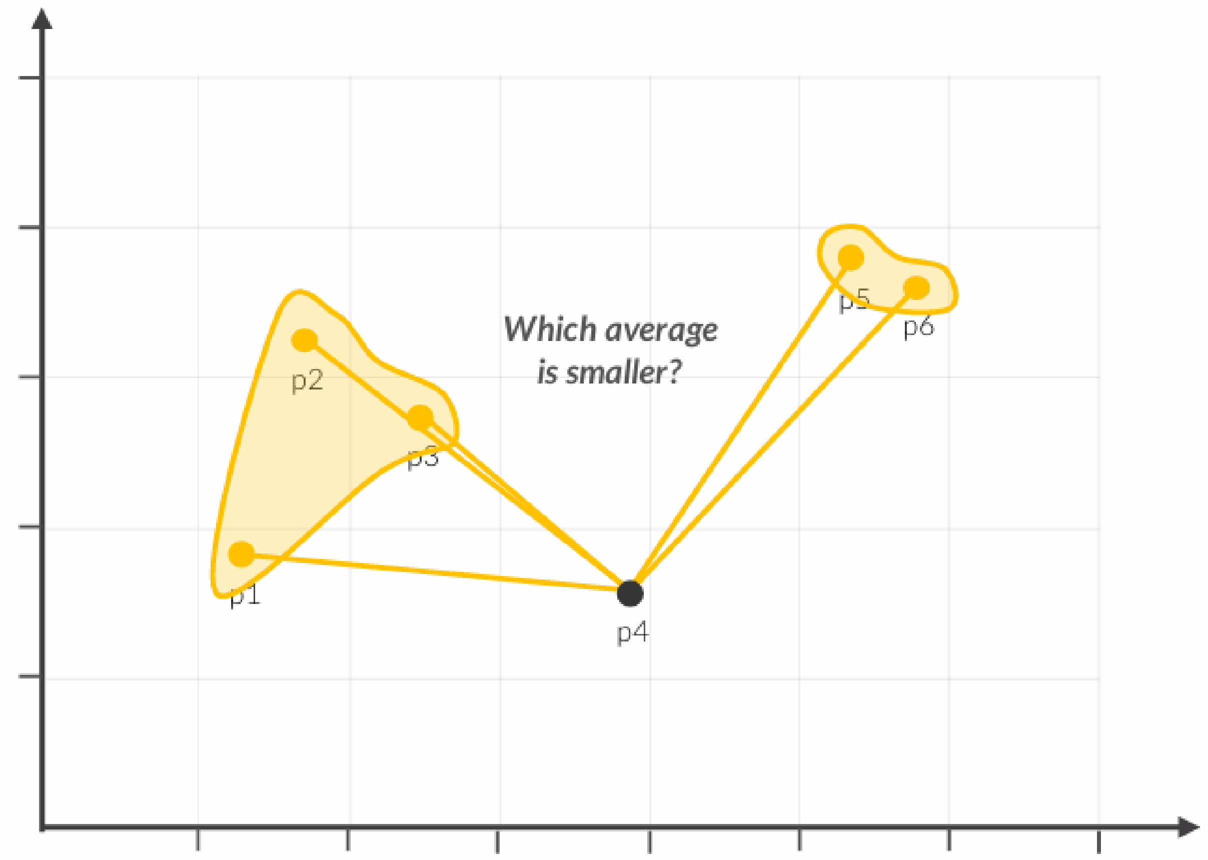
Step 3.5 - Repeat the process until all points are part of the same cluster



How do you define distance between clusters?

- Single linkage (closest)
- **Complete linkage(furthest)**

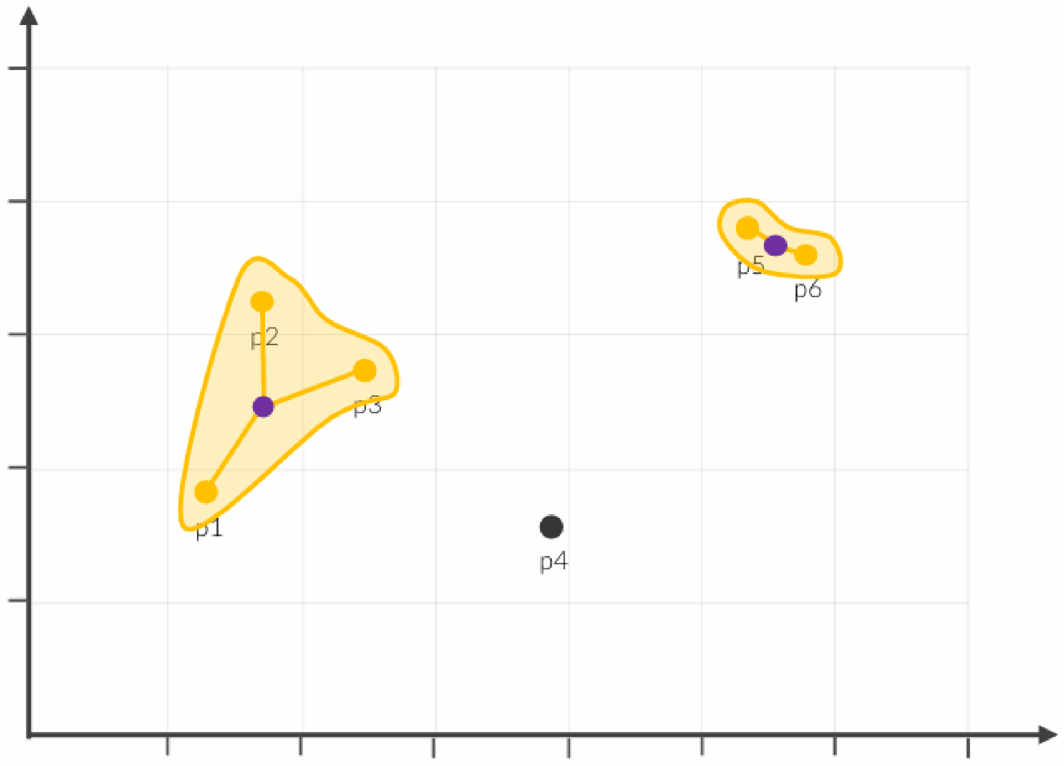
Step 3.6 - Repeat the process until all points are part of the same cluster



How do you define distance between clusters?

- Single linkage (closest)
- Complete linkage(furthest)
- ***Average linkage(all pairs)***

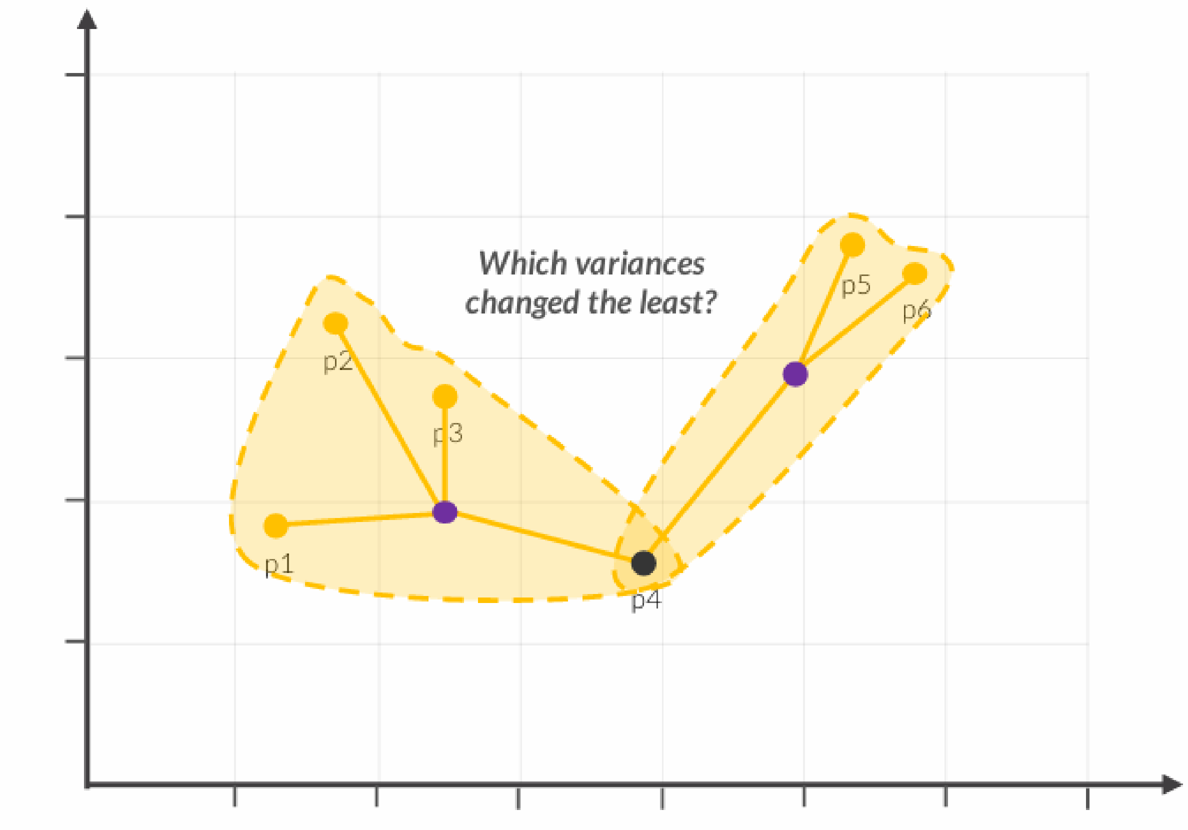
Step 3.7 - Repeat the process until all points are part of the same cluster



How do you define distance between clusters?

- *Single linkage (closest)*
- *Complete linkage(furthest)*
- *Average linkage(all pairs)*
- **Ward's method(variance)**

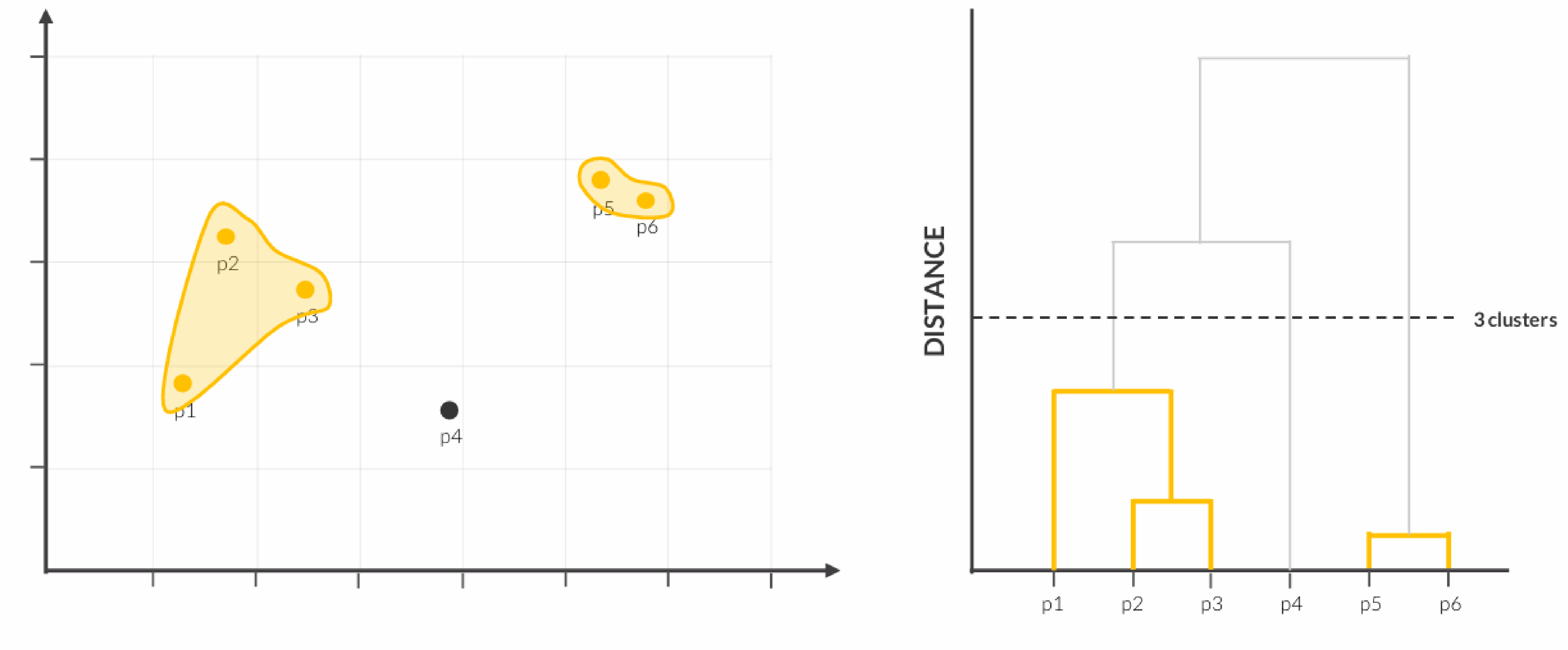
Step 3.8 - Repeat the process until all points are part of the same cluster



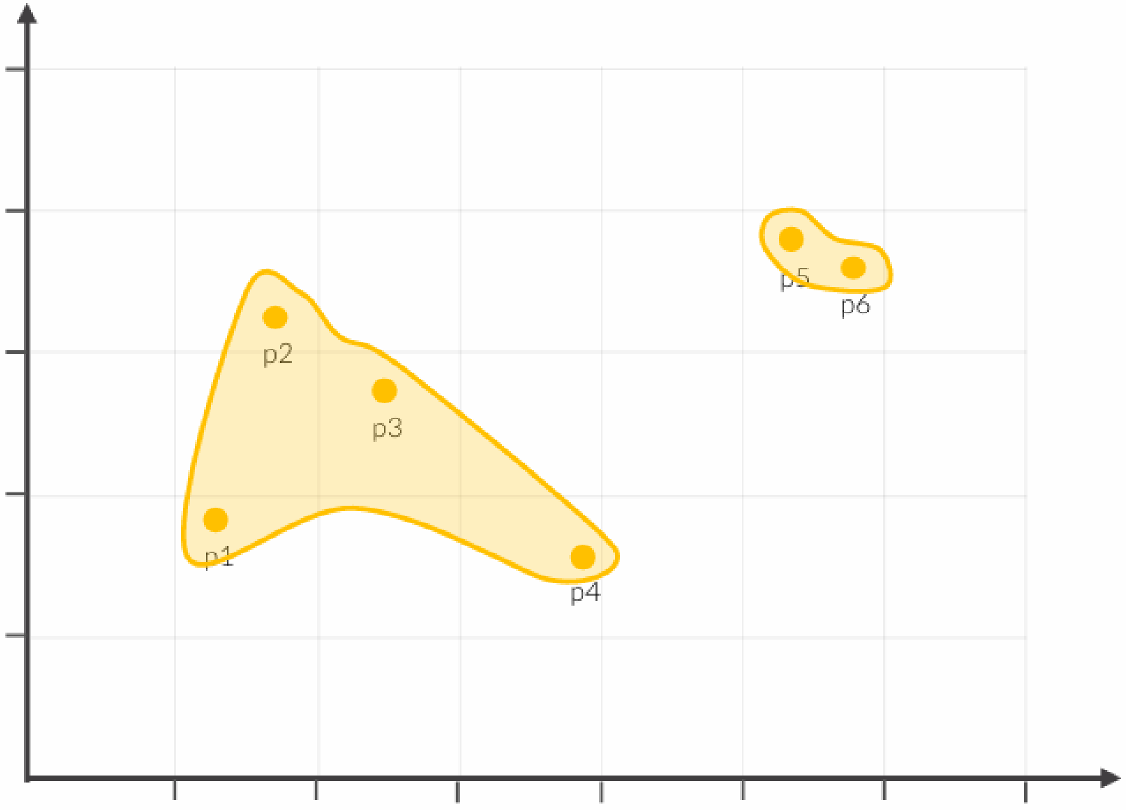
How do you define distance between clusters?

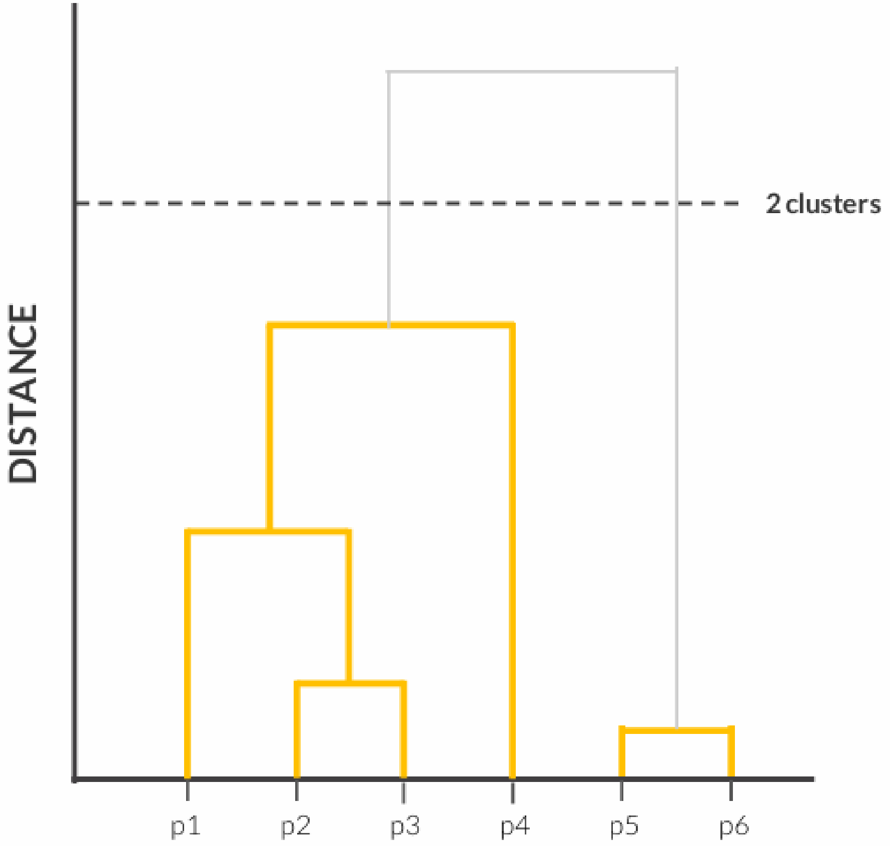
- *Single linkage (closest)*
- Complete linkage(furthest)
- *Average linkage(all pairs)*
- **Ward's method(variance)**

Step 3.9 - Repeat the process until all points are part of the same cluster

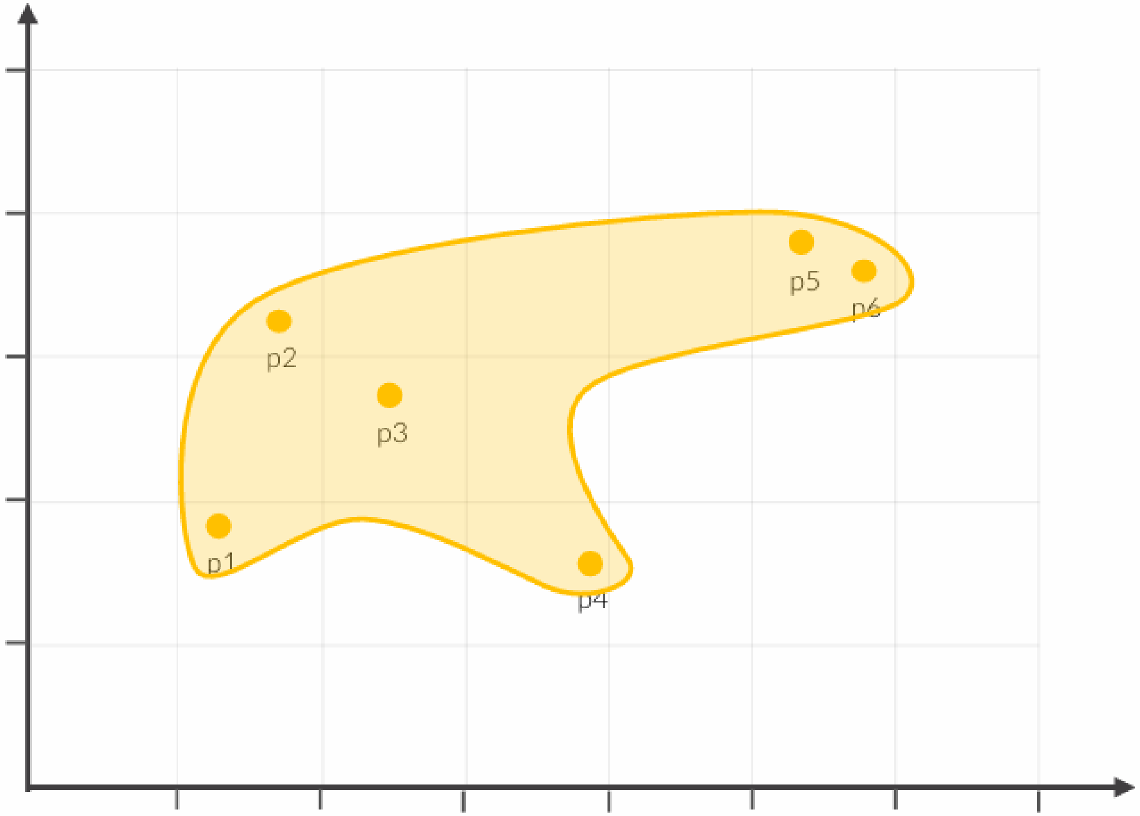


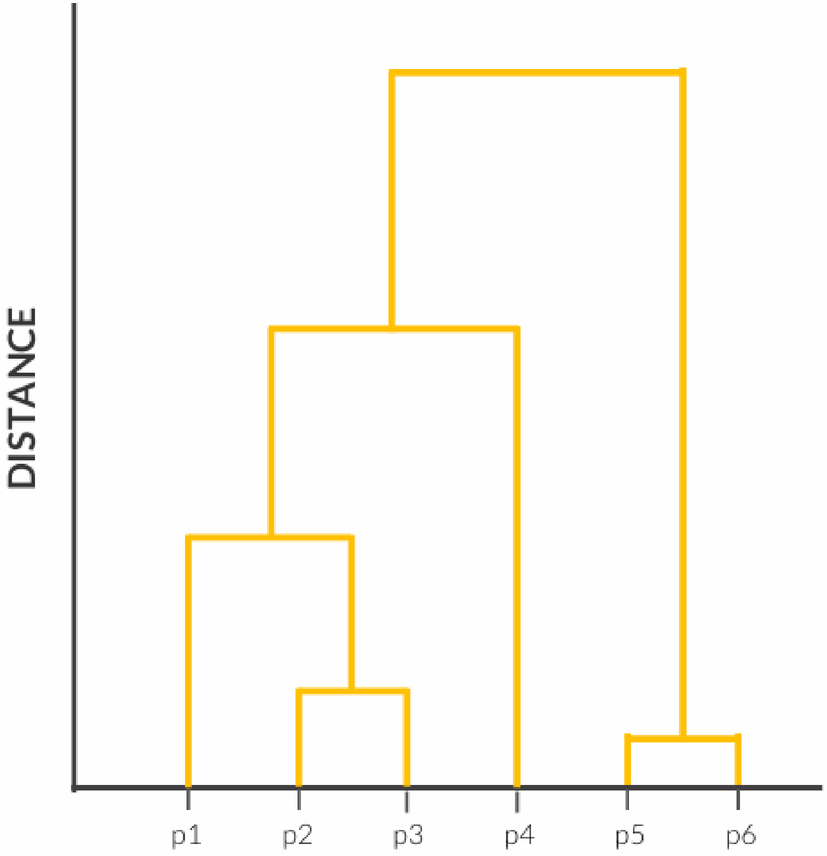
Step 3.10 - Repeat the process until all points are part of the same cluster





Step 3.11 - Repeat the process until all points are part of the same cluster





Agglomerative Hierarchical Clustering summarization

Agglomerative hierarchical clustering is a bottom-up approach that treats each data point as an individual cluster and then iteratively merges them.

- **Initialization:** The process begins by considering every single data point in the dataset as its own cluster.
- **Iterative Merging:** The algorithm finds the two closest clusters and merges them into a single new cluster. The "closest" is determined by a chosen linkage criterion (e.g., single, complete, average, or Ward's).
- **Finalization:** This merging process is repeated until all data points have been grouped into one single, large cluster. The final output is a hierarchy of clusters that can be visualized as a dendrogram.

LET’S GO