

# **Analizzare e Interpretare i dati usando il Clustering Gerarchico**

**Prova Finale – Corso di Laurea in Informatica**

**Candidato: Sara Caviglia**

**Relatore: Prof. Nicoletta Noceti**

# Obiettivi del lavoro

- Studiare l'algoritmo di Clustering Gerarchico, che rappresenta una soluzione semplice ma molto descrittiva al problema del clustering
- Esplorare l'utilizzo dell'algoritmo, per comprenderlo meglio, su un insieme di dati che rappresentano le caratteristiche tecniche di brani musicali



**Università  
di Genova**

**DIBRIS** DIPARTIMENTO  
DI INFORMATICA, BIOINGEGNERIA,  
ROBOTICA E INGEGNERIA DEI SISTEMI

# **Il Clustering Gerarchico**

# Metodo del Clustering Gerarchico

## Cosa vuol dire «clustering»

Il clustering è la partizione degli elementi di un insieme in modo tale che gli elementi nello stesso gruppo (ovvero nello stesso cluster) siano più simili rispetto agli elementi presenti negli altri gruppi (cluster).

Fonte: [\*Cluster analysis or Clustering - Wikipedia\*](#)

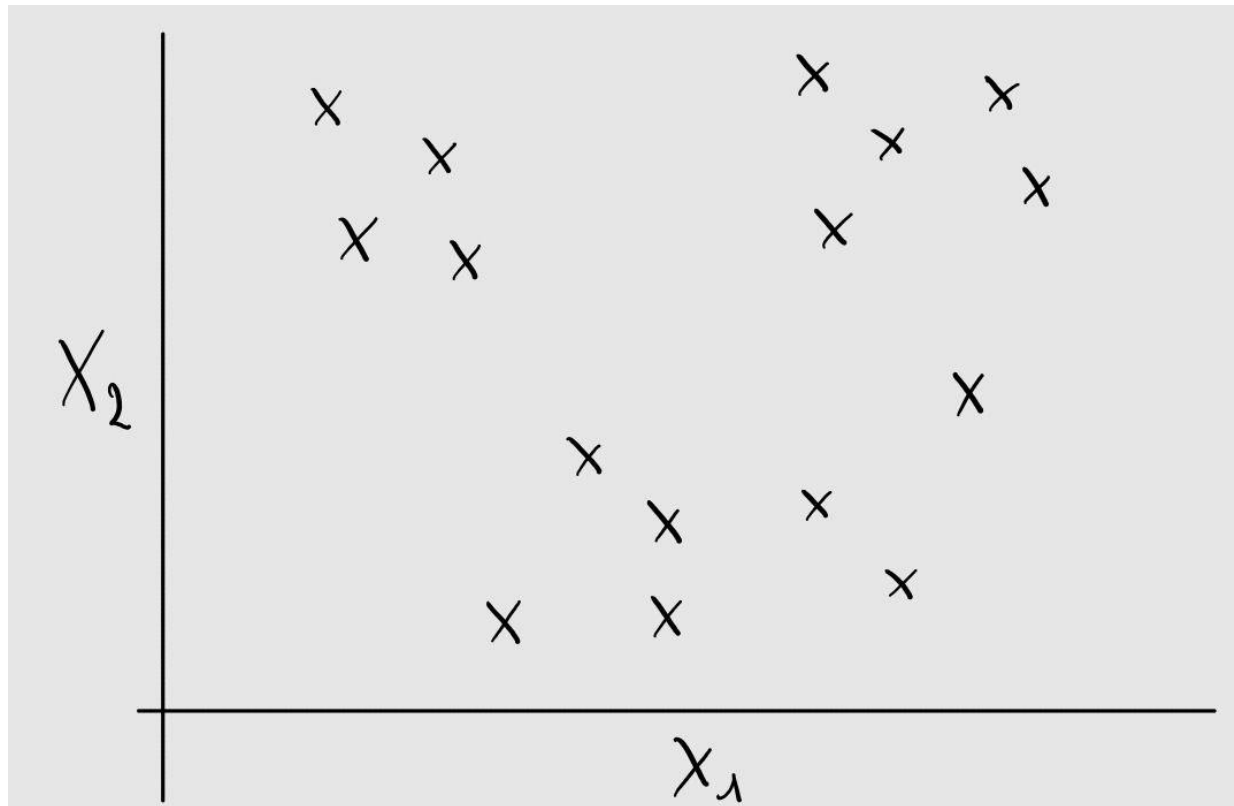
# Metodo del Clustering Gerarchico

## Diverse tipologie di Clustering Gerarchico

- In base alla formula usata per calcolare le distanze tra insiemi di punti
- In base all'approccio utilizzato
  - Agglomerativo o bottom-up
  - Divisivo o top-down

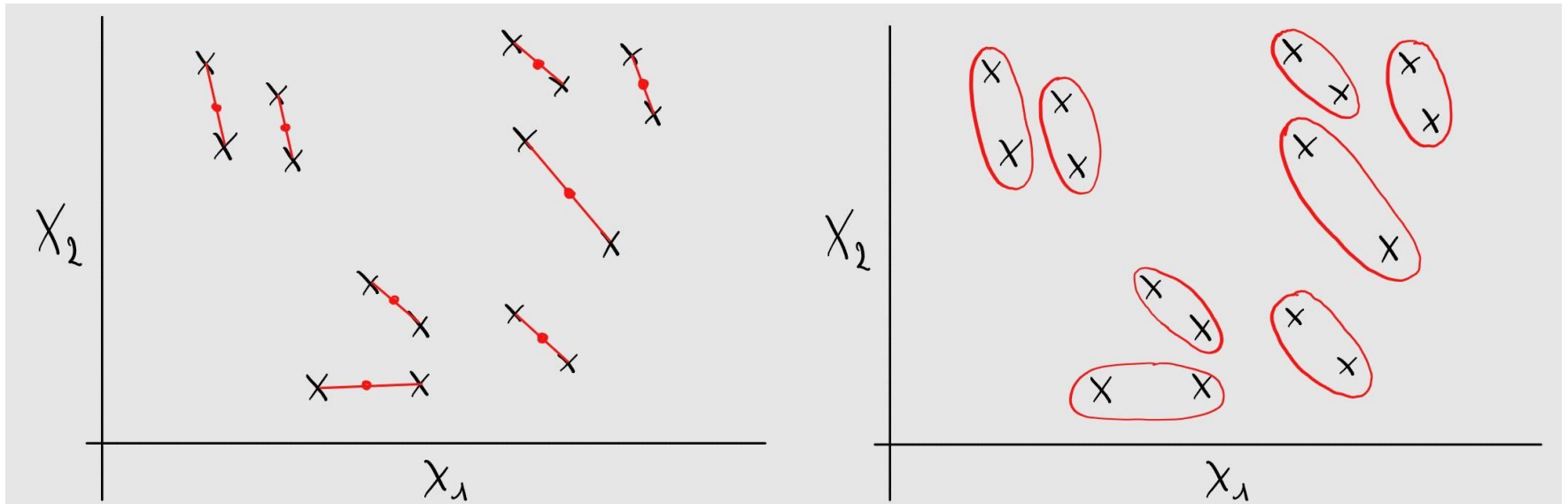
# Metodo del Clustering Gerarchico

## Esempio



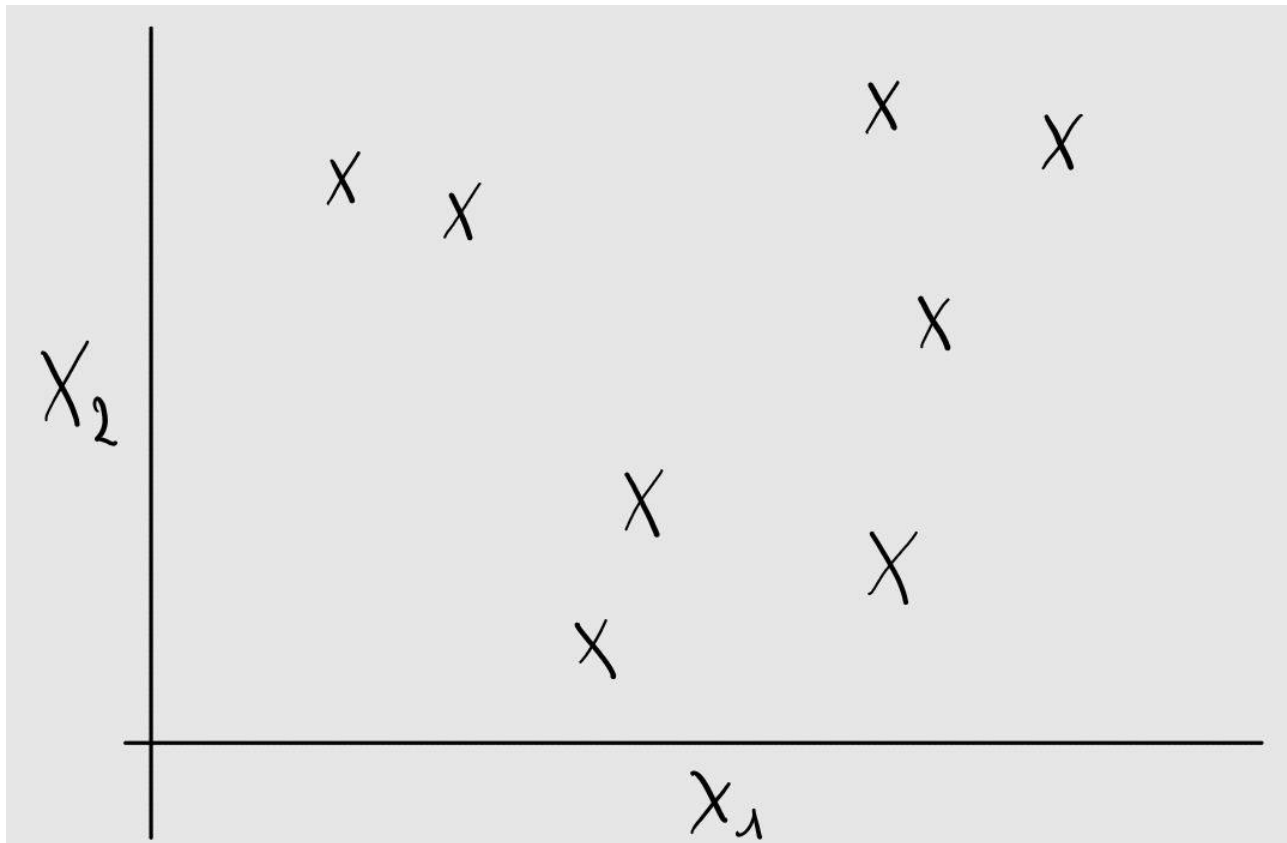
# Metodo del Clustering Gerarchico

## Esempio



# Metodo del Clustering Gerarchico

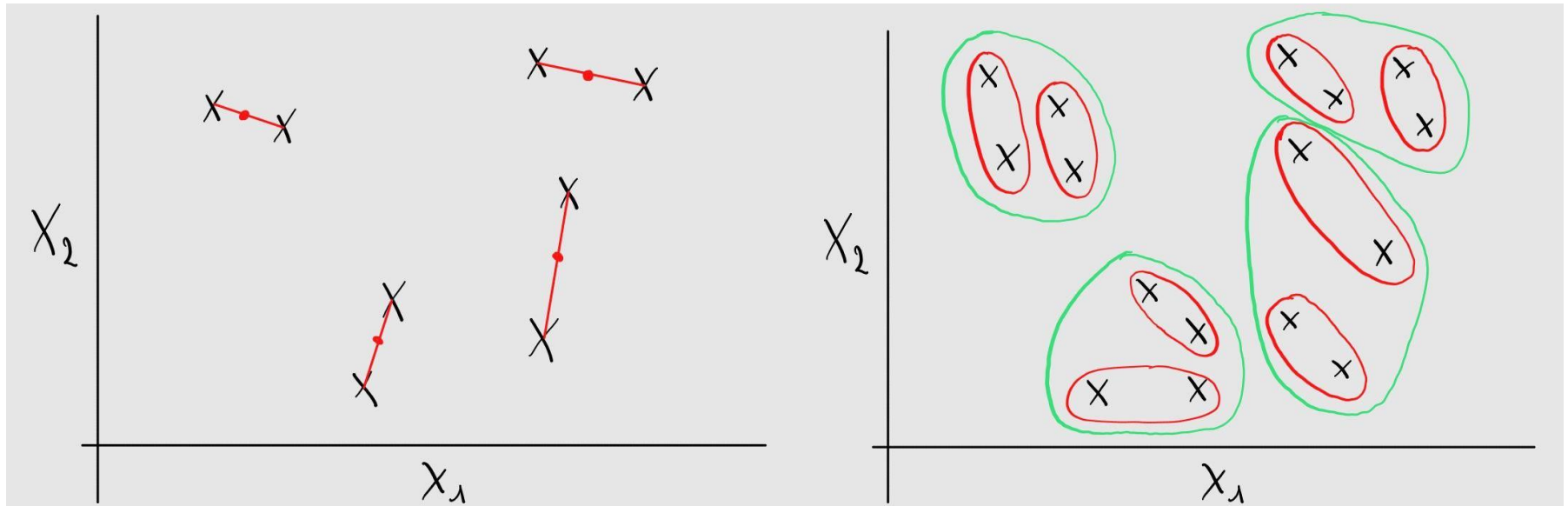
## Esempio





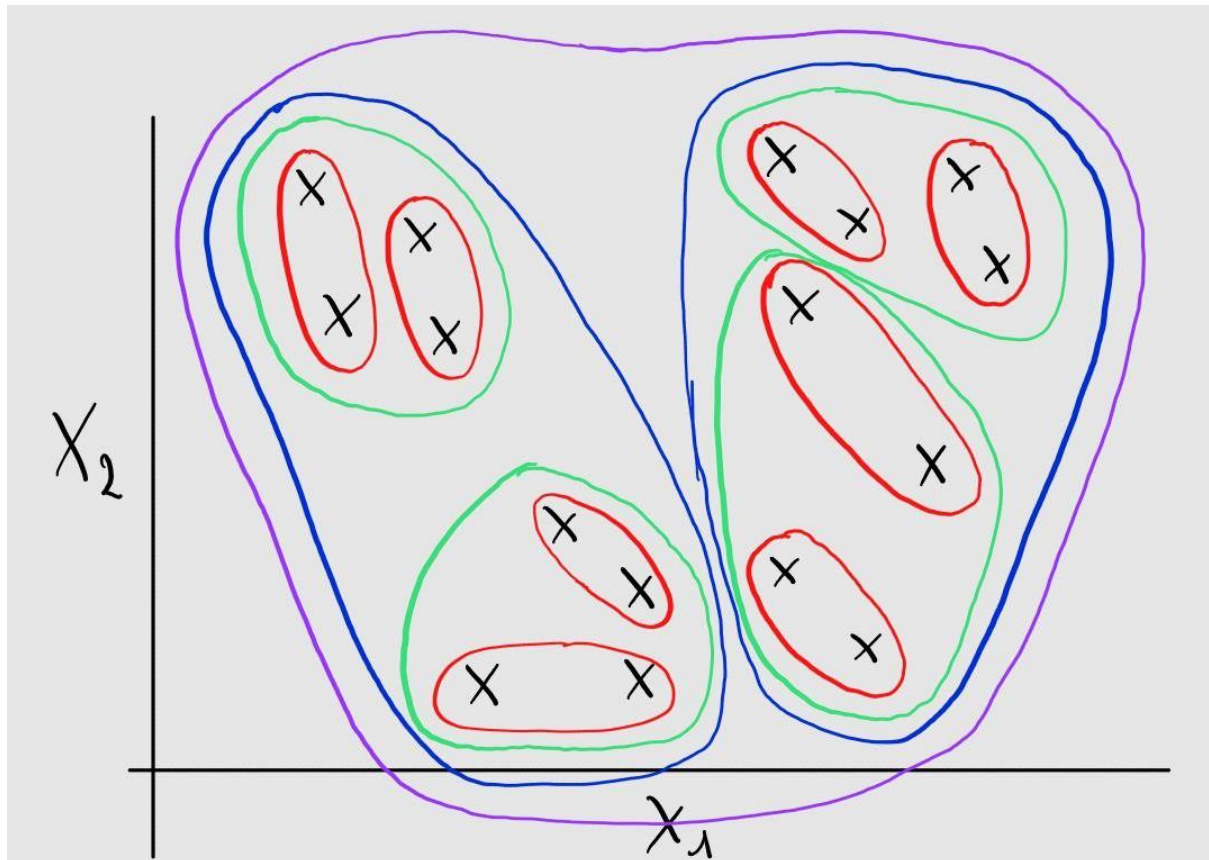
# Metodo del Clustering Gerarchico

## Esempio

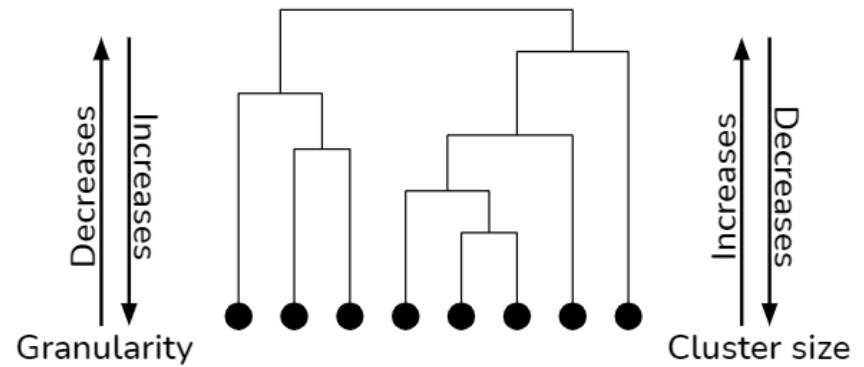
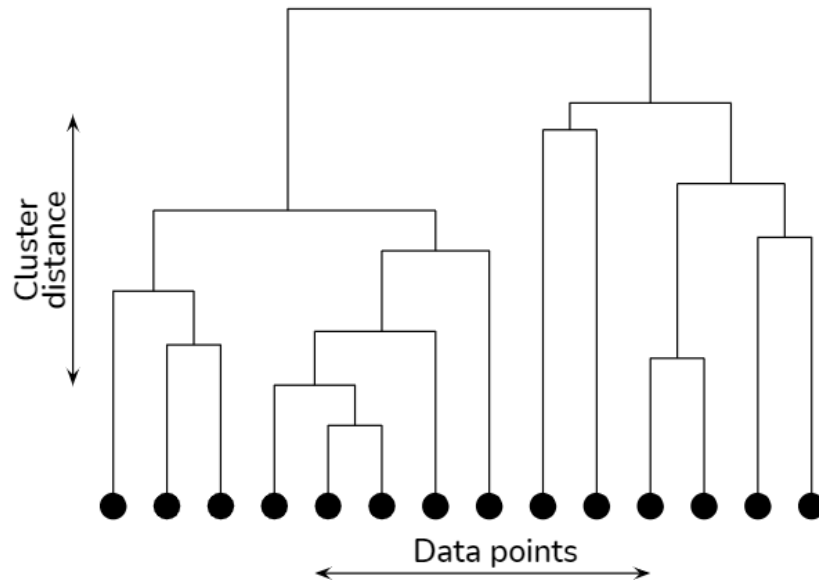


# Metodo del Clustering Gerarchico

## Esempio



# Dendrogramma



Fonte: [Hierarchical Clustering explained](#)



**Università  
di Genova**

**DIBRIS** DIPARTIMENTO  
DI INFORMATICA, BIOINGEGNERIA,  
ROBOTICA E INGEGNERIA DEI SISTEMI

# **Applicazione del Clustering Gerarchico**

# Dataset di Spotify: Generi musicali

- 'acousticness'
- 'danceability'
- 'energy'
- 'instrumentalness'



```
genre
Other      123884
Acoustic/Folk  19784
Rock       19068
Jazz       2794
Hip-Hop/Rap 1011
Pop        787
Electronic/Dance 39
Name: count, dtype: int64
```

Link al dataset: [160k Spotify songs from 1921 to 2020 - Kaggle](#)

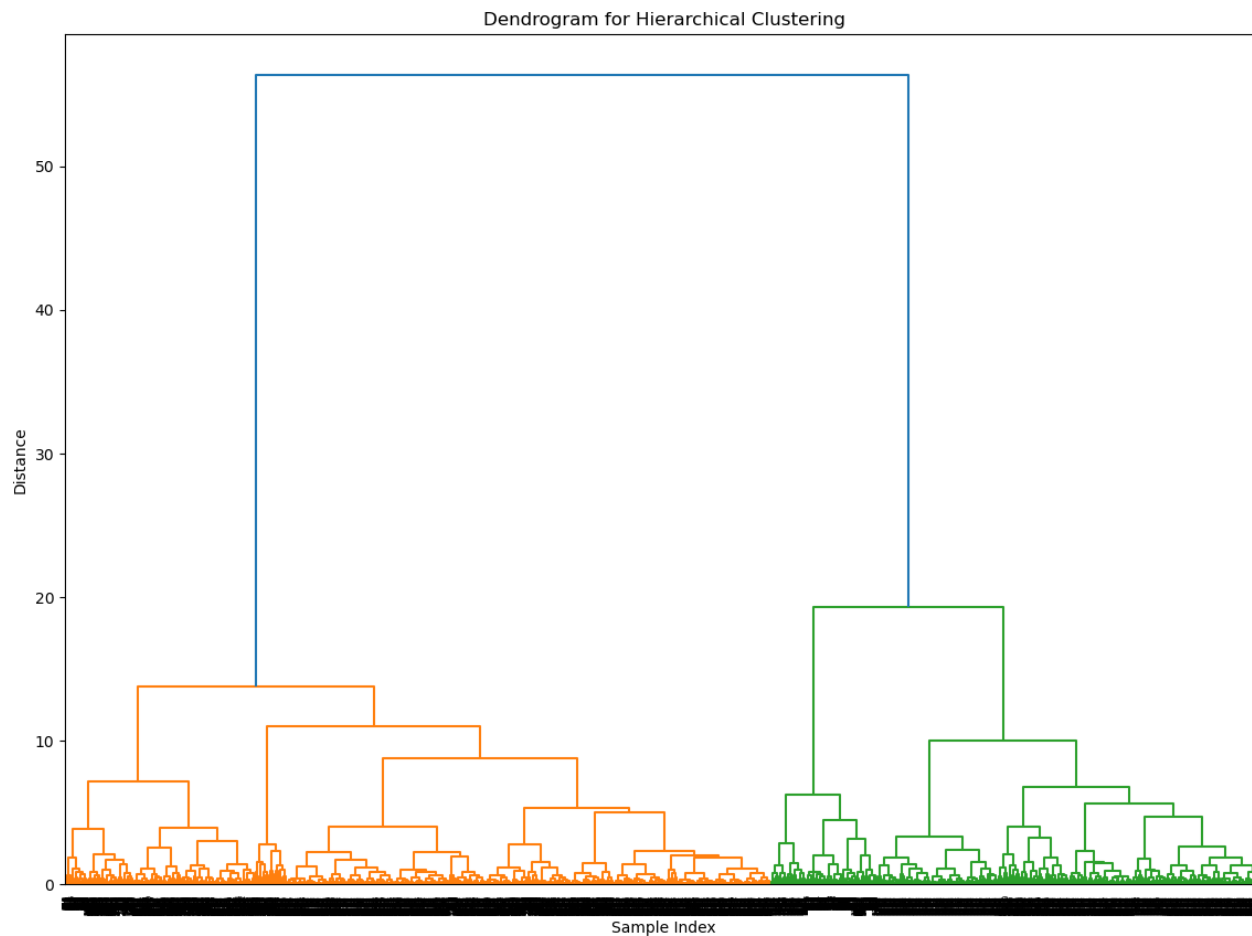
# Utilizzo del Clustering Gerarchico

## Obiettivo e motivazioni

- Come si aggregano i dati?
- Il clustering rispecchia l'annotazione?

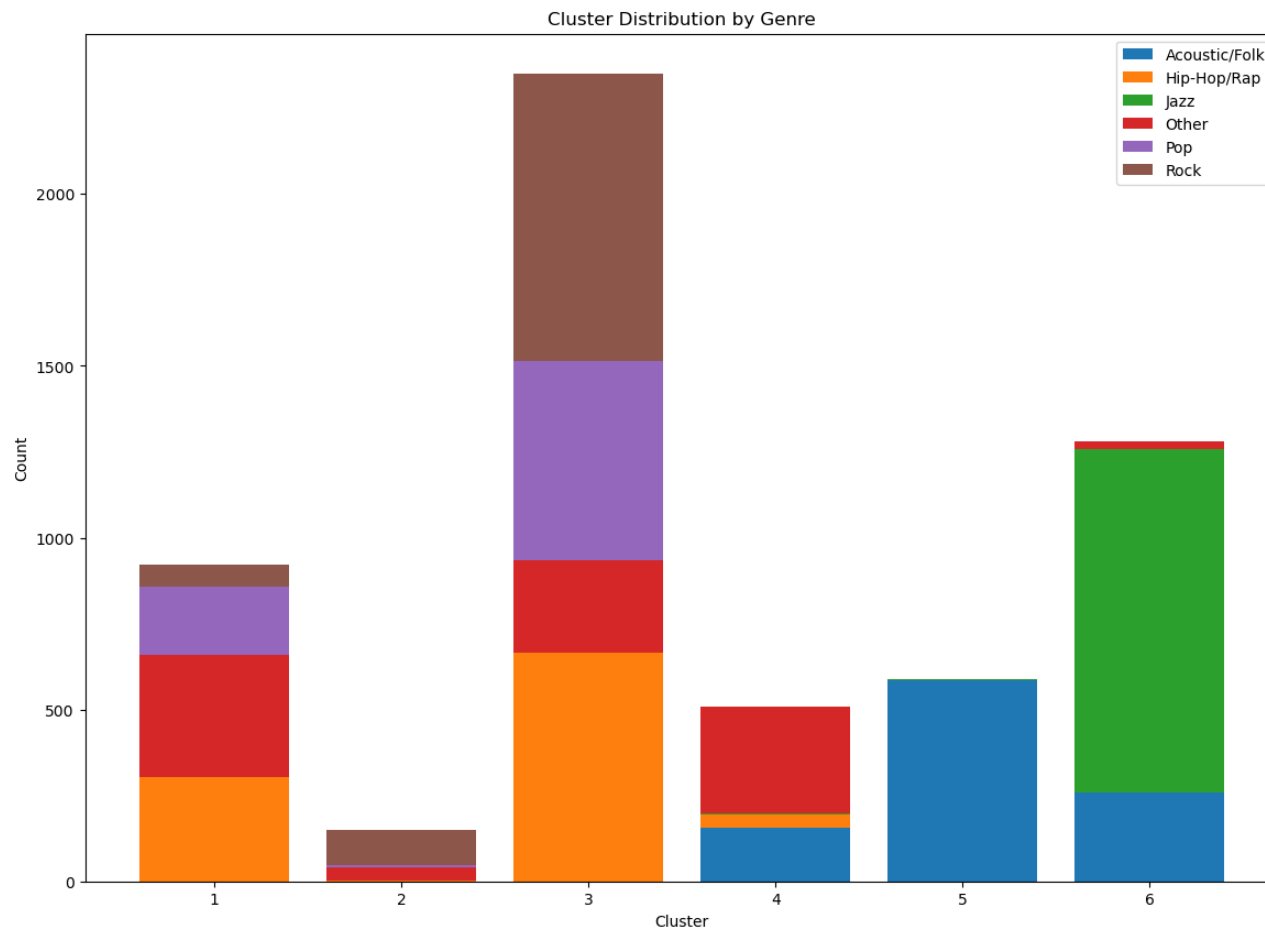
# Utilizzo del Clustering Gerarchico

## Dendrogramma



# Utilizzo del Clustering Gerarchico

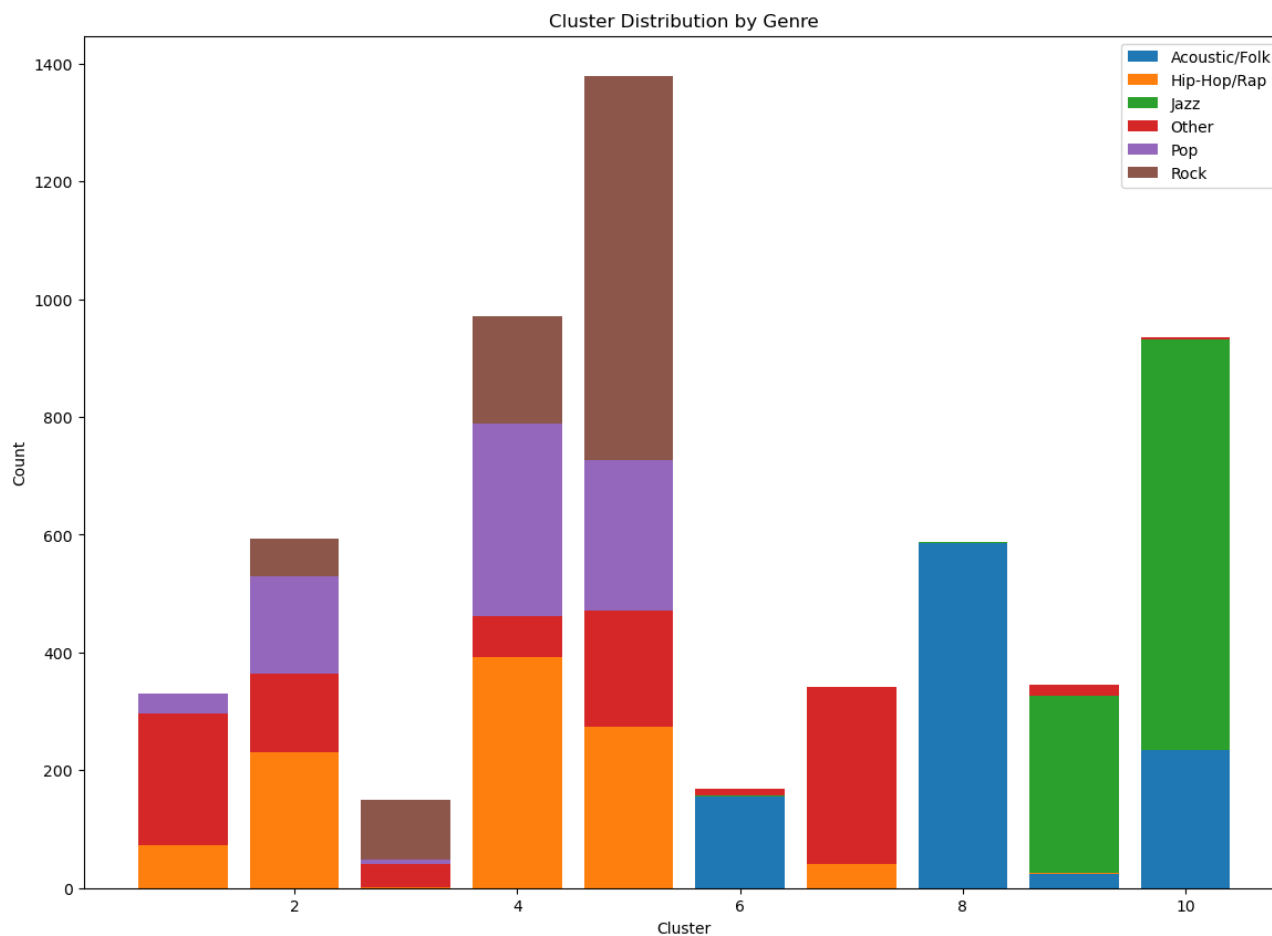
## Generi suddivisi nei cluster





# Utilizzo del Clustering Gerarchico

## Generi suddivisi nei cluster



# Osservazioni

- Generi simili «mescolati» negli stessi cluster
- Suddivisione dei generi negli stessi cluster, indipendentemente dal loro numero

**UniGe**  

---

**DIBRIS**