Extra Material for P1W3D3

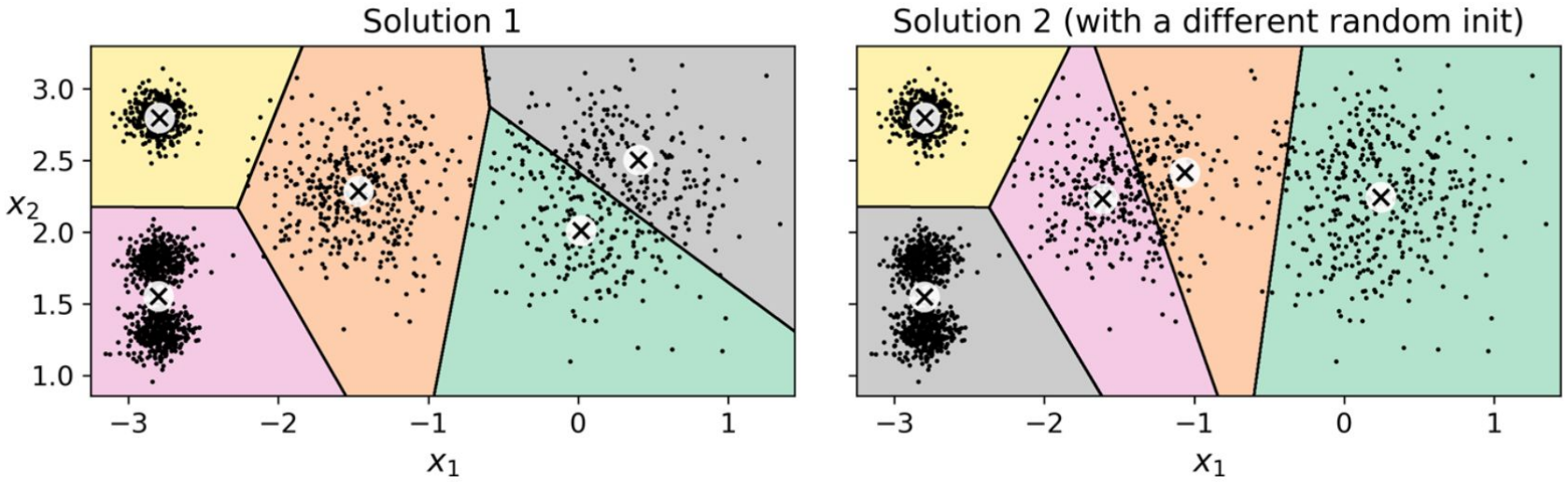
P1 W3 D3 AM
GAUSSIAN MIXTURE MODEL

KMeans - Advantages

- Guaranteed to converge in a finite number of steps (usually quite small iterations).
- One of the fastest clustering algorithm.
- Easy to understand.

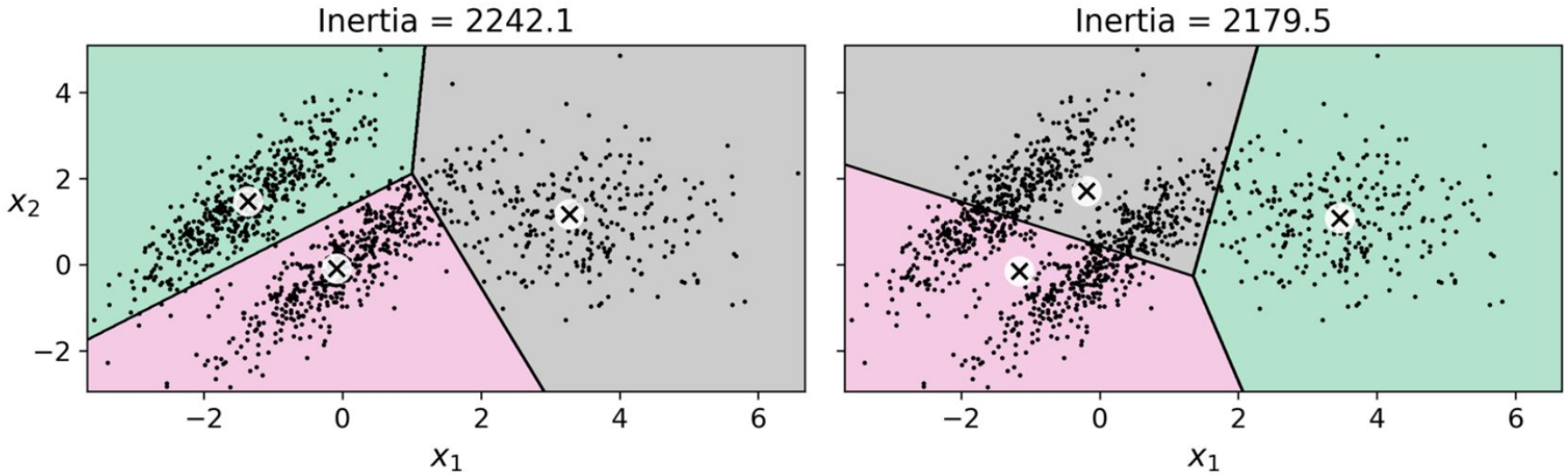
KMeans - Disadvantages (1)

- Depends on the centroid initialization.



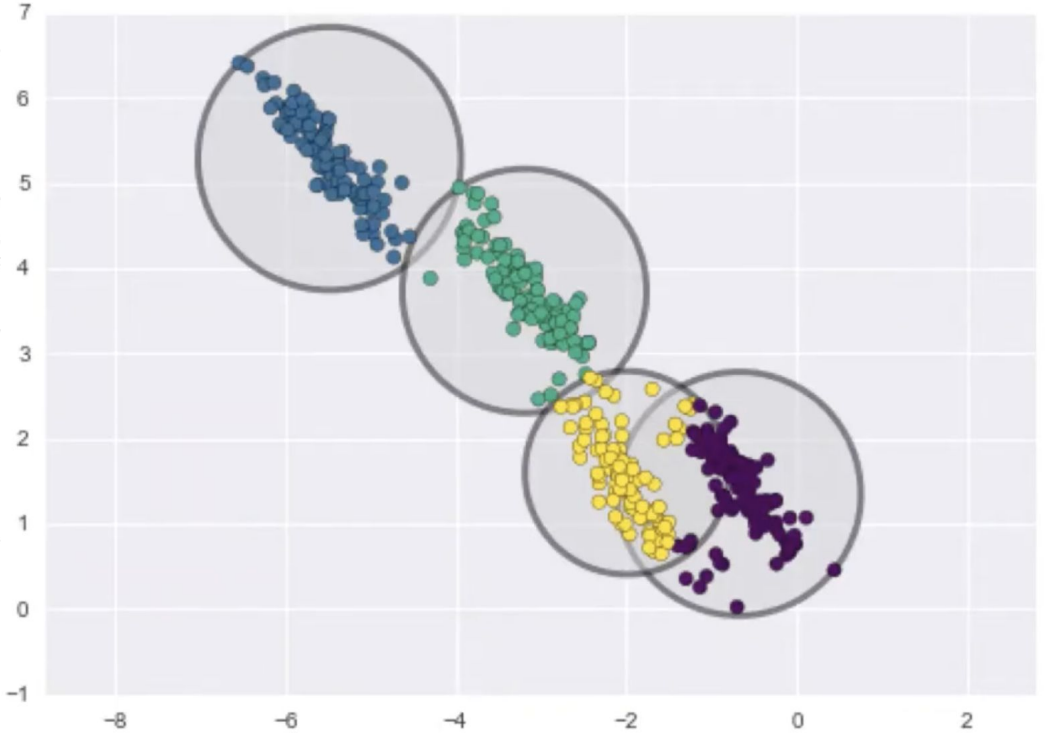
KMeans - Disadvantages (2)

- KMeans does not behave very well when the clusters have varying sizes, different densities, or non spherical shapes.



KMeans - Disadvantages (3)

- KMeans is a hard-clustering technique.



Problem : News Clustering (1)

- Cluster 1 : news related to **Sport**.
- Cluster 2 : news related to **Technology**.

3 Tim Negara Besar Gagal Lolos Piala Dunia 2022

CNN Indonesia

Jumat, 25 Mar 2022 14:00 WIB

Bagikan :

Timnas Italia gagal ke Piala Dunia 2022. (REUTERS/GUGLIELMO MANGIAPANE)

Hal-hal yang Perlu Diketahui dari 5G: Bukan Cuma soal Cepat

Inn | CNN Indonesia

Senin, 13 Dec 2021 04:11 WIB

Bagikan :

Ilustrasi. 5G tak hanya bicara soal internet cepat. Teknologi tersebut membawa pengguna merasakan pengalaman lebih baik di segala lini. (Foto: iStockphoto/alexsl)

Problem : News Clustering (2)

Which cluster for this news ?

FIFA Bakal Uji Teknologi Offside Semi-Otomatis di Piala Arab 2021

jun | CNN Indonesia

Rabu, 01 Dec 2021 01:35 WIB

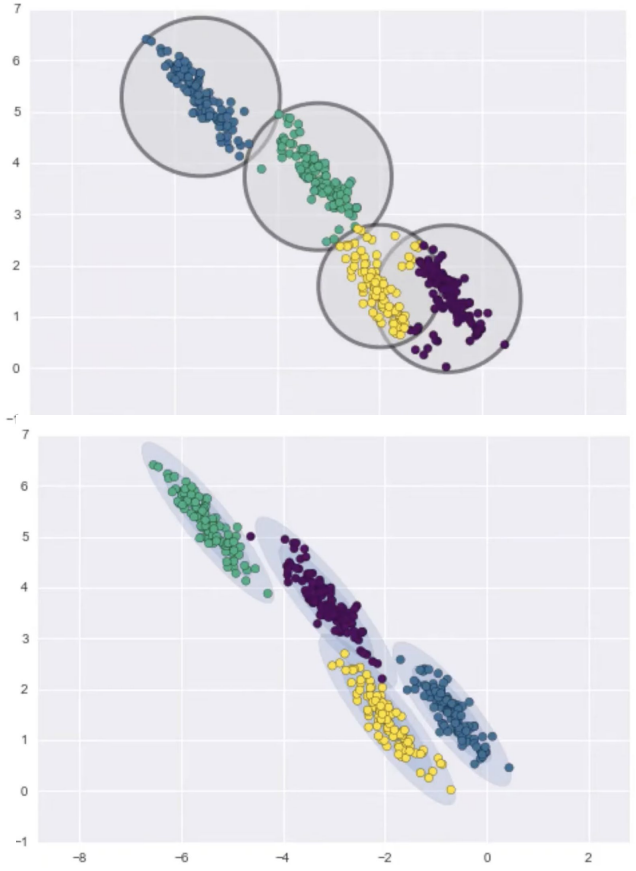
Bagikan :

Ilustrasi offside di pertandingan sepak bola. (AFP PHOTO / JEAN-CHRISTOPHE VERHAEGEN)

KMeans vs GMM

- KMeans :
  - ➔ Good for spherical-shape
  - ➔ Hard clustering

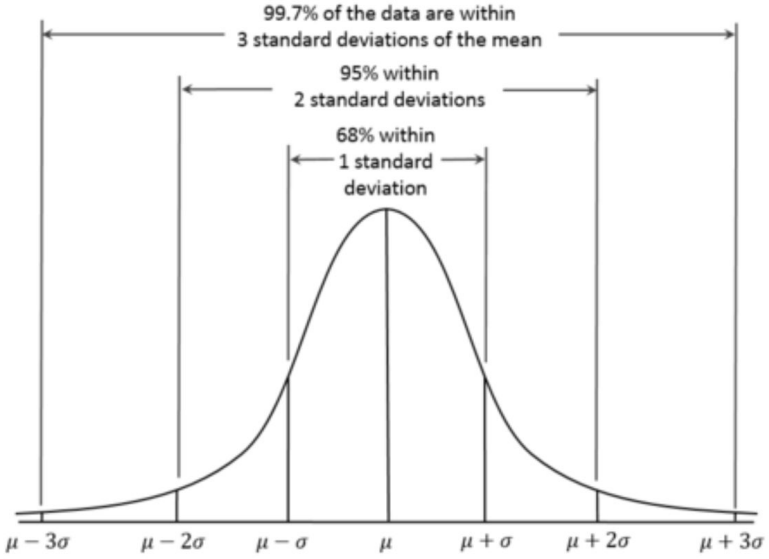
- GMM :
  - ➔ Good for non spherical-shape
  - ➔ Soft clustering

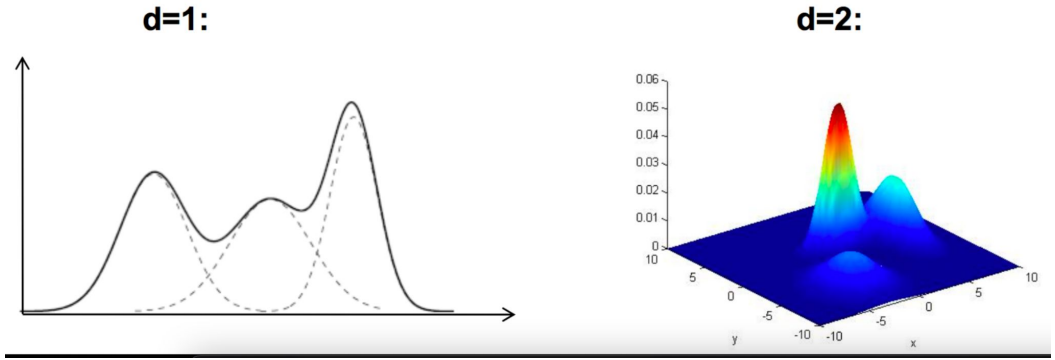


Gaussian Mixture Model (1)

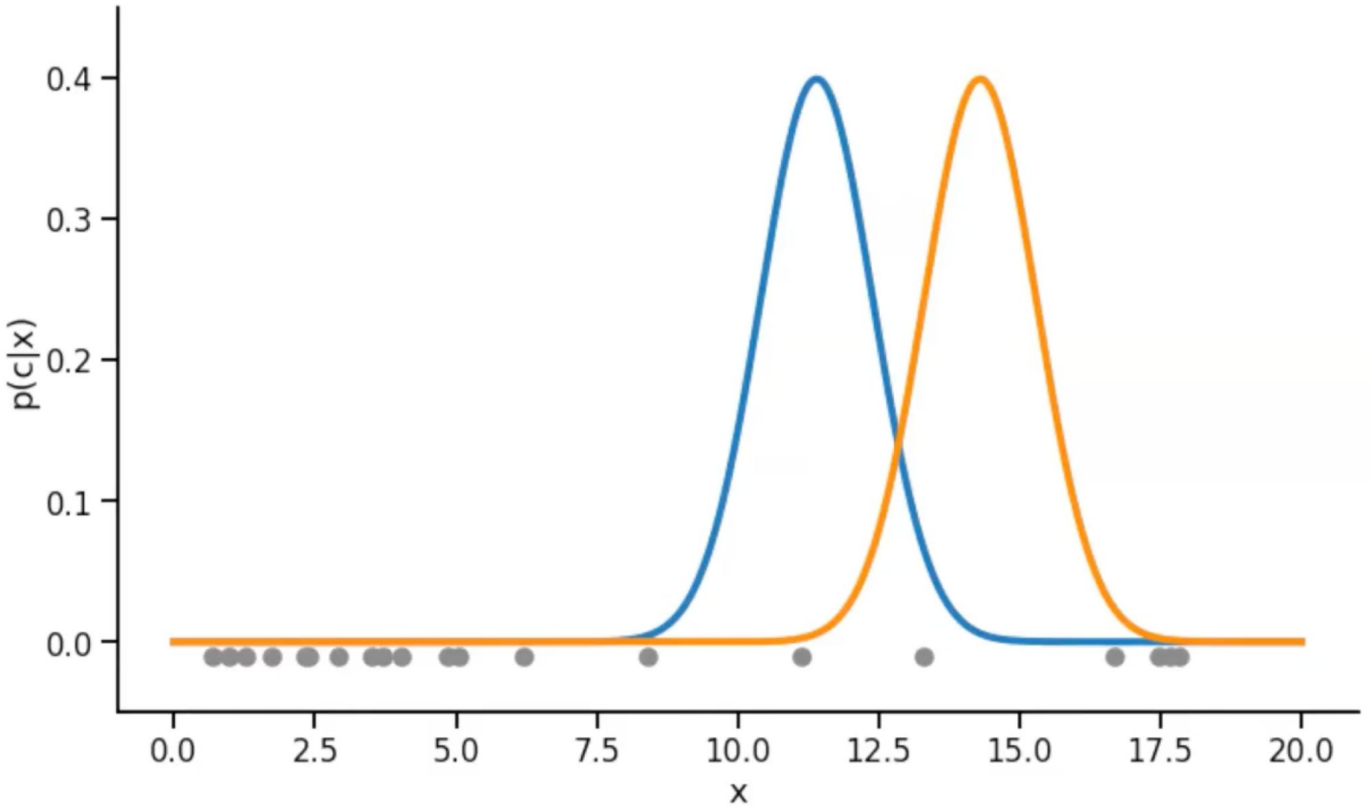
- Using probabilistic approaches.
- Based on Expectation-Maximisation algorithm.
- GMM is a model that assumes data generated from a mixture of K Gaussian (bell shape) components.

Gaussian Mixture Model (2)

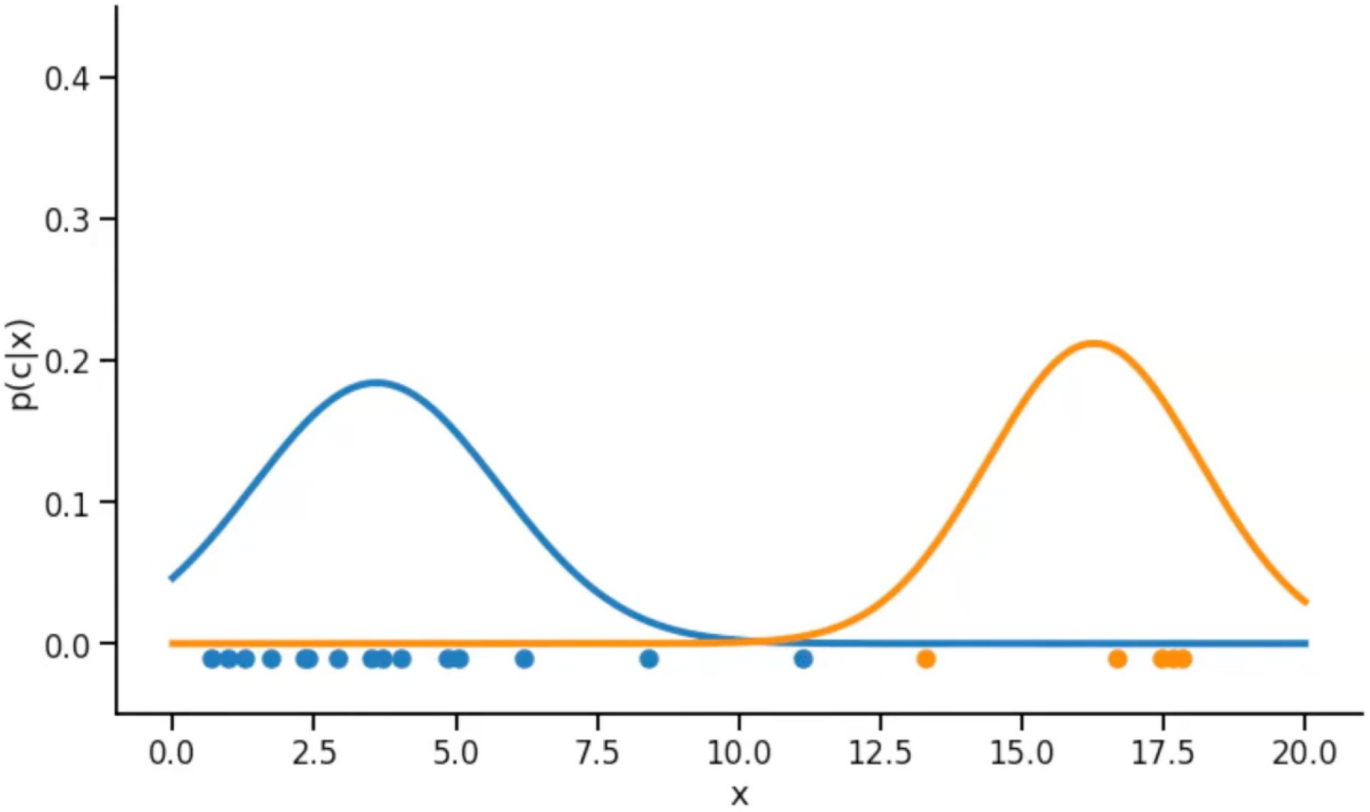




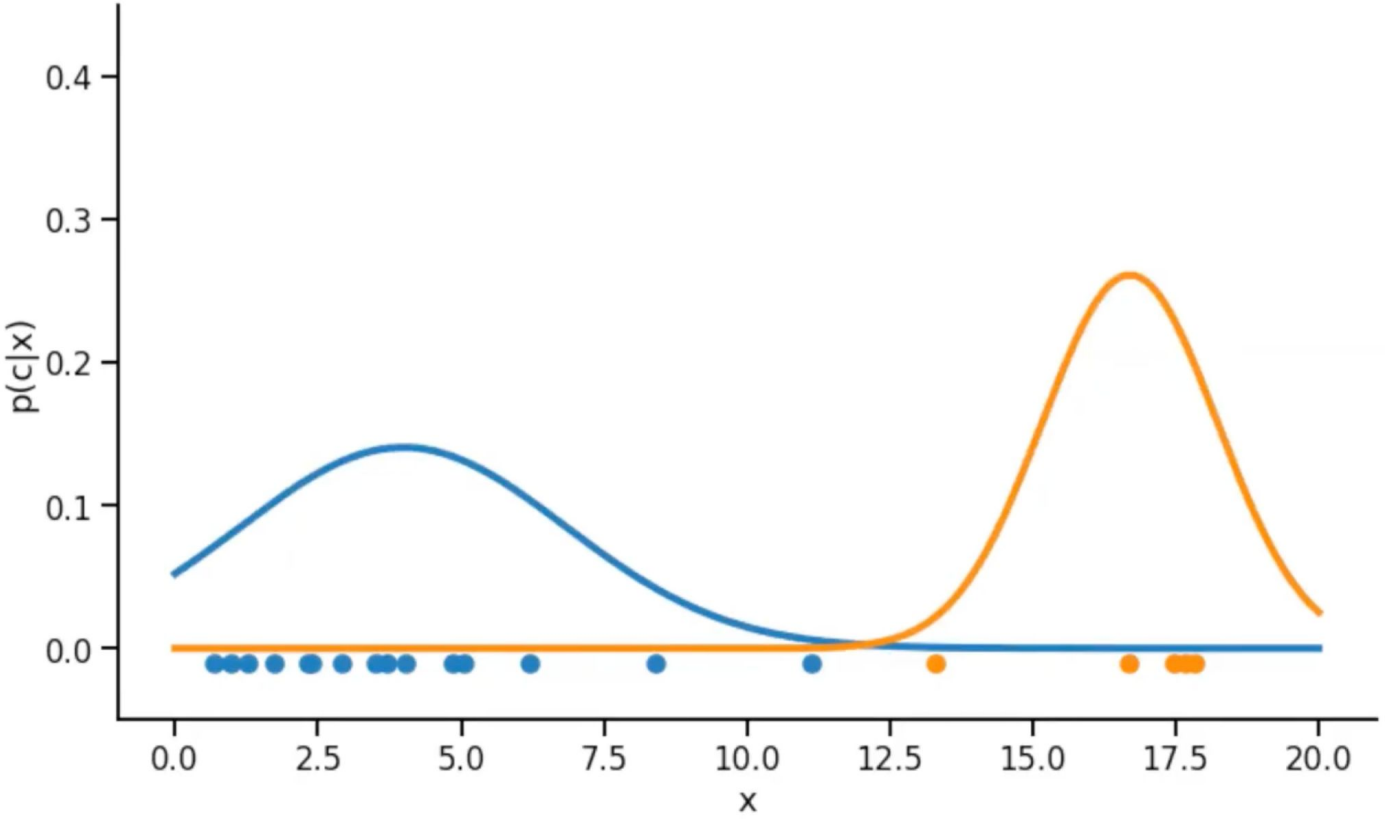
Gaussian Mixture Model - Illustration (Iteration 1)



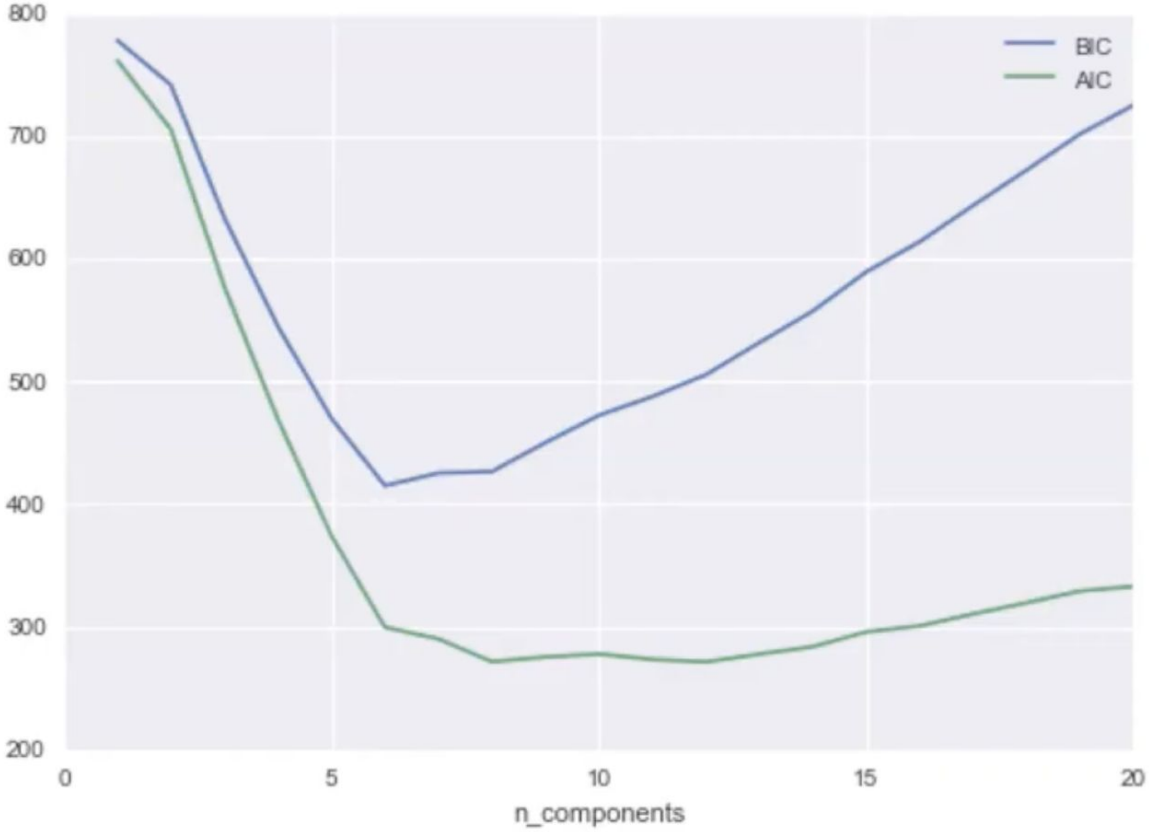
Gaussian Mixture Model - Illustration (Iteration 2)



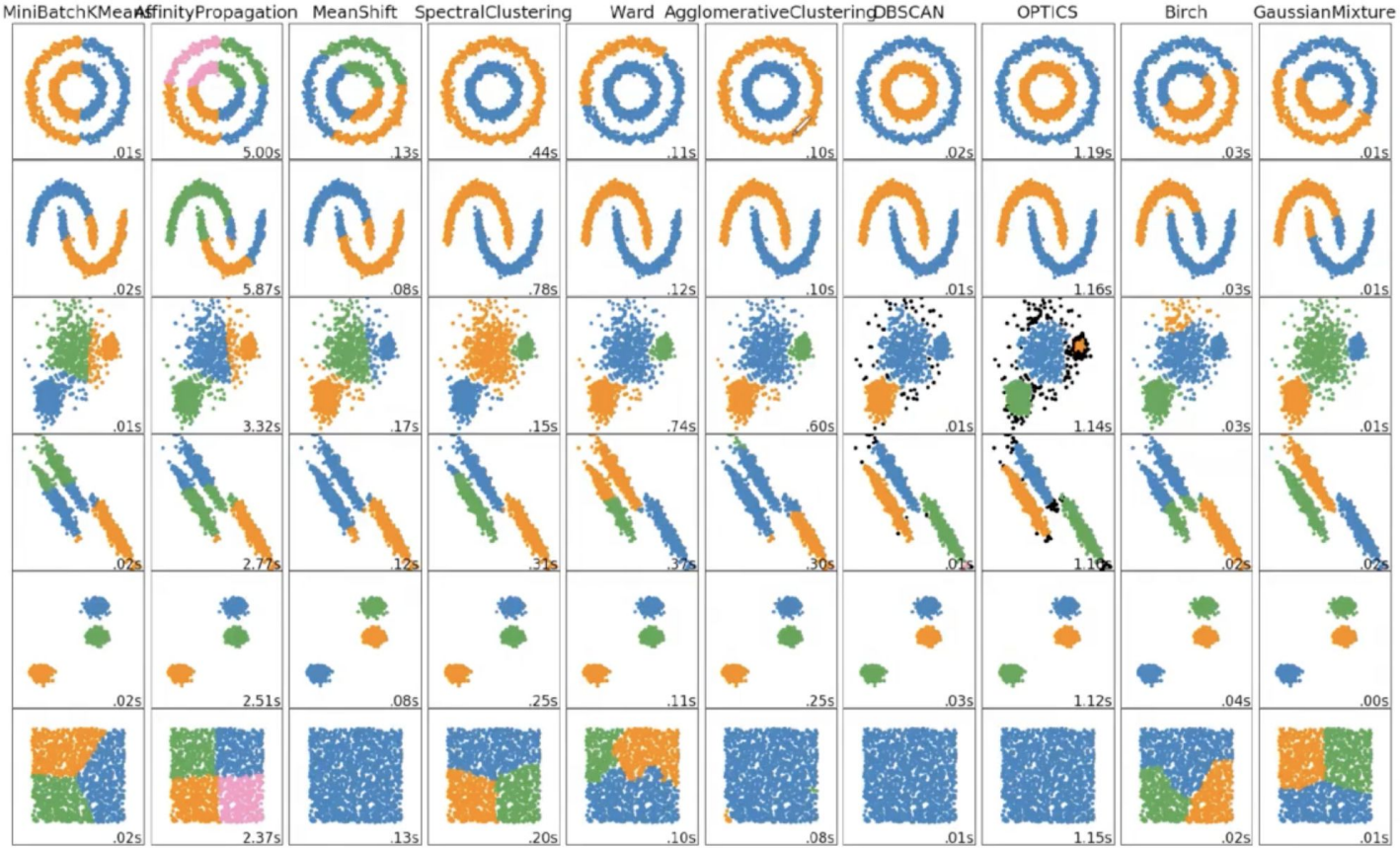
Gaussian Mixture Model - Illustration (Iteration 3)



BIC & AIC



Algorithm Comparison

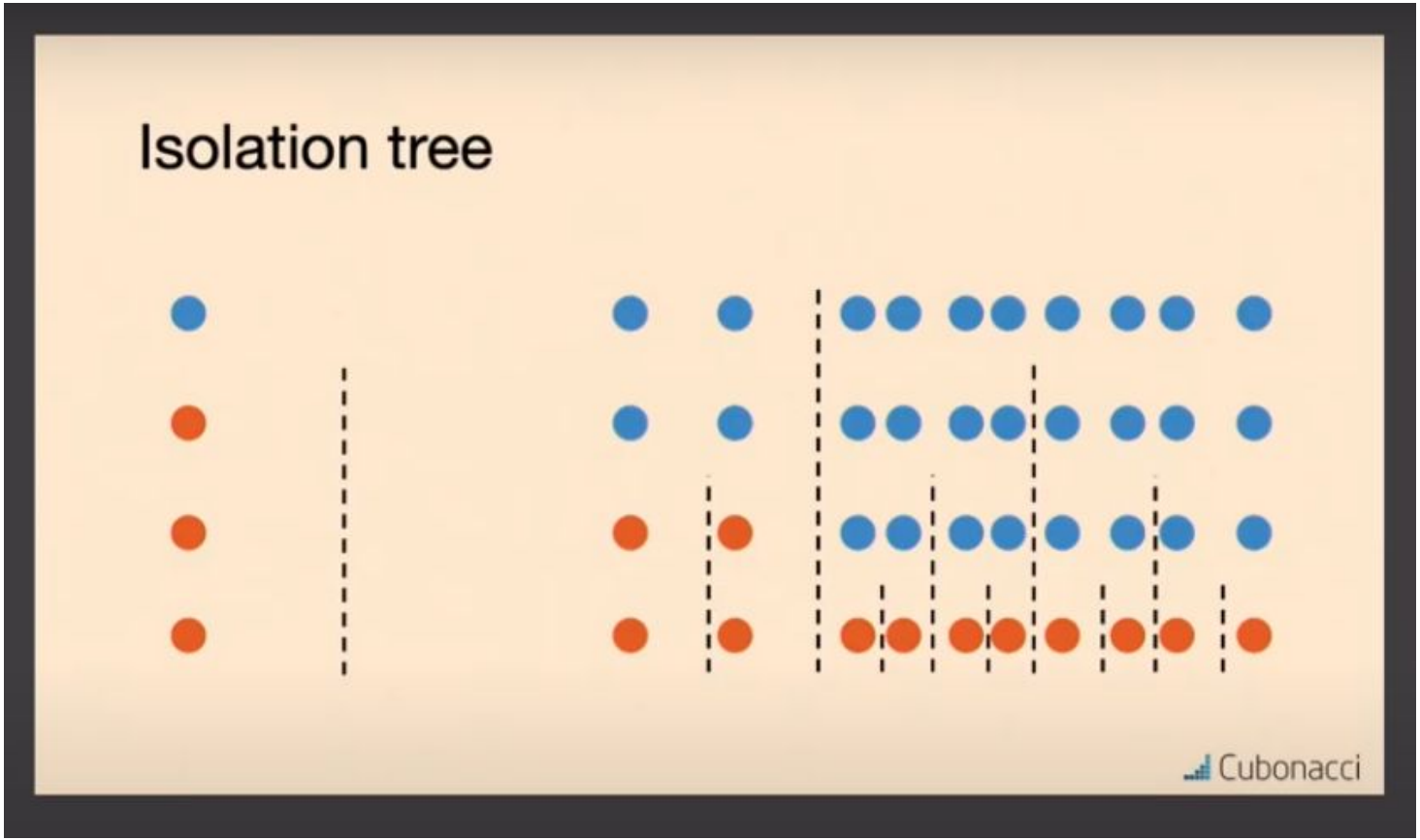


P1 W3 D3 PM

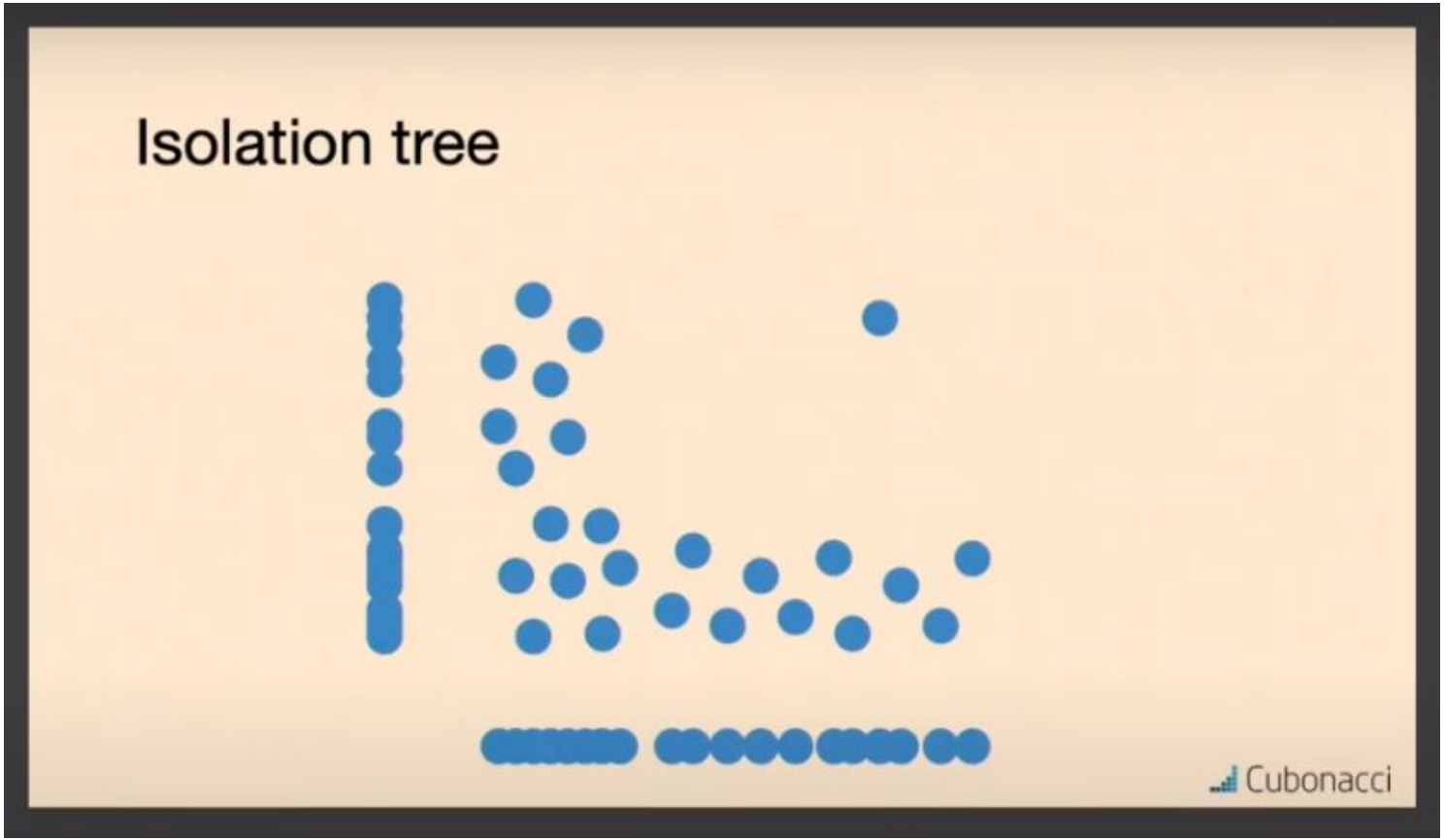
Illustration (Outlier)

Illustration (Novelty)

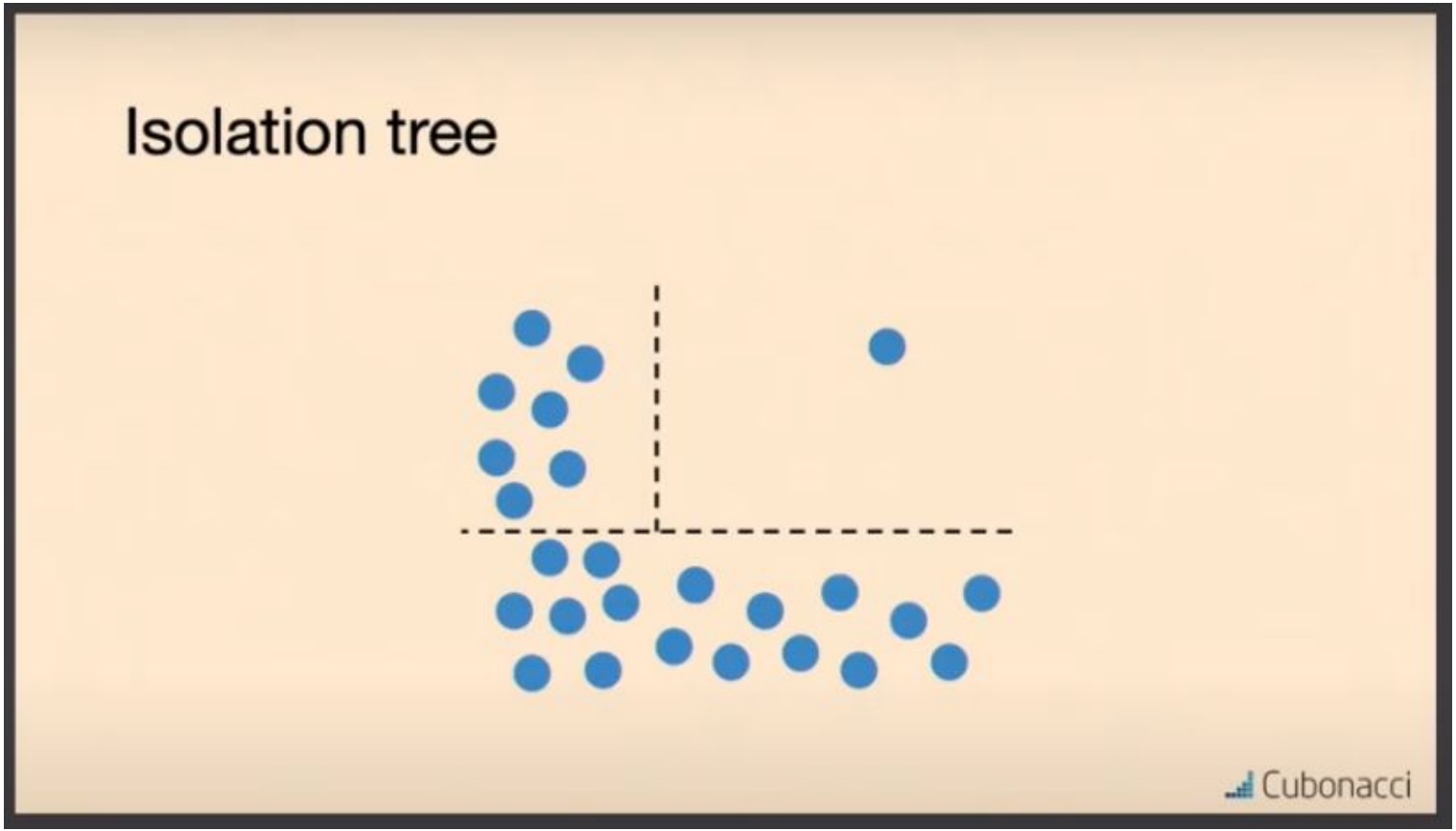
Isolation Forest (1)



Isolation Forest (2)



Isolation Forest (3)



Isolation Forest (4)