***Gaussian Mixture Models (GMMs) Clustering***

**Laporan Tugas**

**Kelas Penambangan Data**

**Taufik Fathurahman**

**NIM: 1301160790**

**Refka Muhammad Furqan**

**NIM: 1301162748**

**Haddad Alwi Yafie**

**NIM: 1301162756**



**Program Studi Sarjana Informatika**

**Fakultas Informatika**

**Universitas Telkom**

**Bandung 2019**

# Latar Belakang *Gaussian Mixture Models*

*K-means clustering* model merupakan model yang sangat sederhana dan relatif mudah untuk dipahami, namun dengan kesederhanaan dan kemudahannya itu memiliki kekurangan ketika diterapkan terhadap situasi dunia nyata karena sifat dari *k-means* yang non-probabilistik. Dari kekurangan yang muncul di *k-means* menjadi latar belakang munculnya metode *gaussian mixture models (GMMs).*

Untuk membuktikan kelemahan dari *k-means*, kami membuat simulasi terhadap *sample data* sebanyak 400 sampel. Akan ada dua jenis data yang akan diujikan terhadap *k-means*, yaitu data yang sederhana, dan data yang lebih kompleks untuk dikelompokkan.

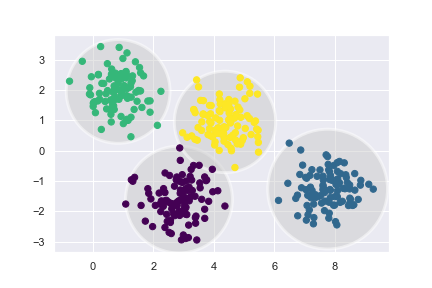


Figure k-means untuk data sederhana

Dari Figure 1, didapatkan hasil yang baik dari *k-means* ketika melakukan *clustering* terhadap data yang sederhana. Lalu bagaimana bila data yang harus dikelompokkan lebih kompleks, seperti persebaran data yang tidak berbentuk lingkaran seperti diatas? Mari kita perhatikan Figure 2 dibawah ini.

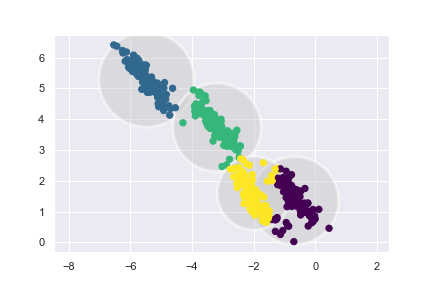


Figure Kmeans untuk data yang kompleks

Dari Figure 2 didapatkan hasil bahwa *k-means* ini lebih tidak fleksibel terhadap sebaran data yang kecenderungannya tidak berbentuk lingkaran. Dari kelemahan ini lah yang menjadi motivasi untuk munculnya *gaussian mixture models (GMMs)*.

# Distribusi Gaussian

Formula untuk distribusi gaussian disebut juga dengan *probability density function (PDF).* *PDF* merupakan fungsi dari variable acak kontinu, yang integral dalam suatu interval memberikan probabilitas bahwa nilai variable berada dalam interval yang sama. Distribusi gaus dan formula *PDF* adalah digambarkan pada Figure 3 dan 4 berikut ini.

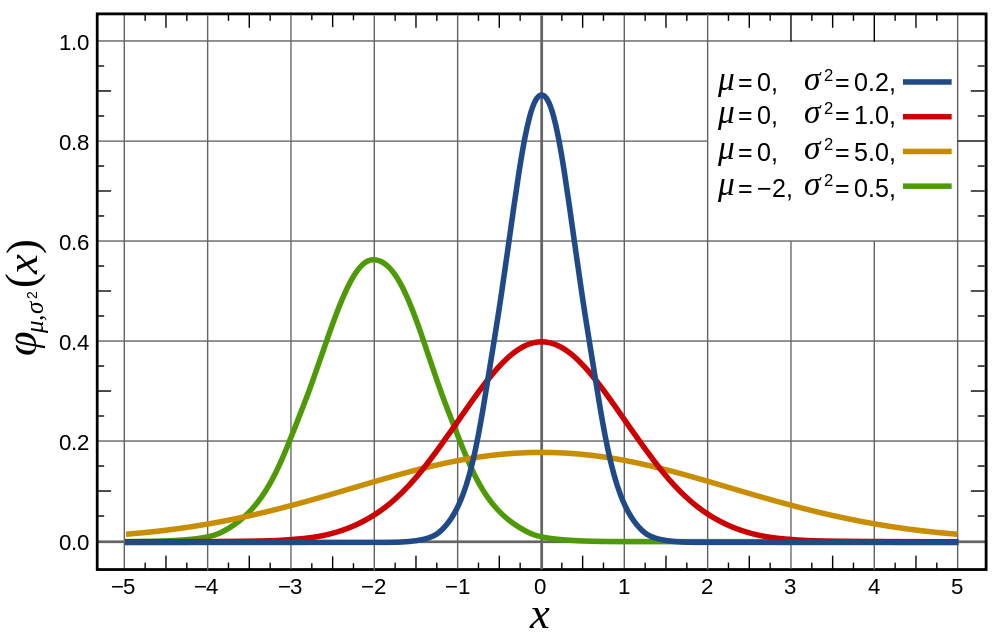


Figure Distribusi gaussian

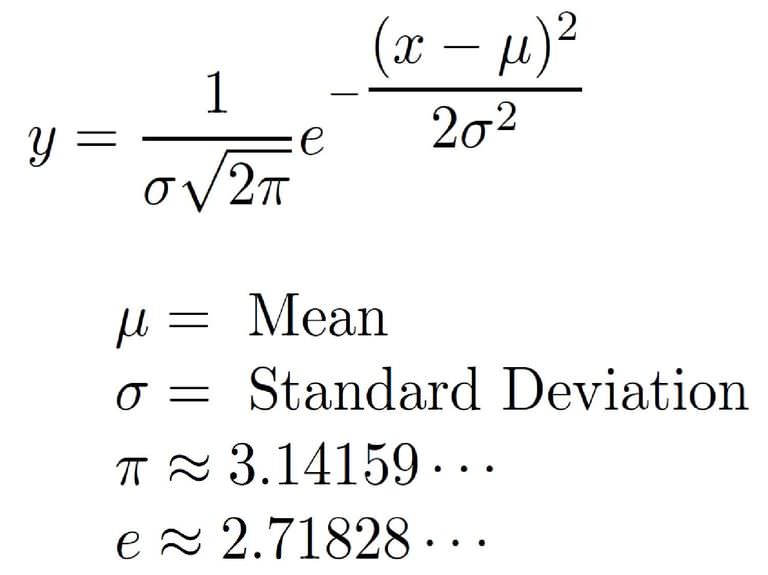


Figure Probabilistik density function