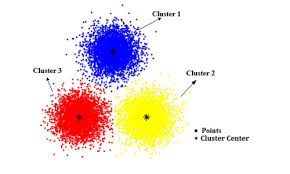
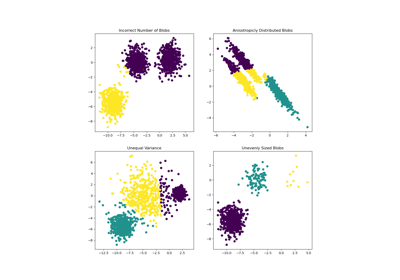
Gaussian Mixture Model

The Gaussian mixture model is a probabilistic density function that assigns labels based on probability.

Before we start, let's talk about k-means algorithms. K means assign value when a record falls into its circular cluster.



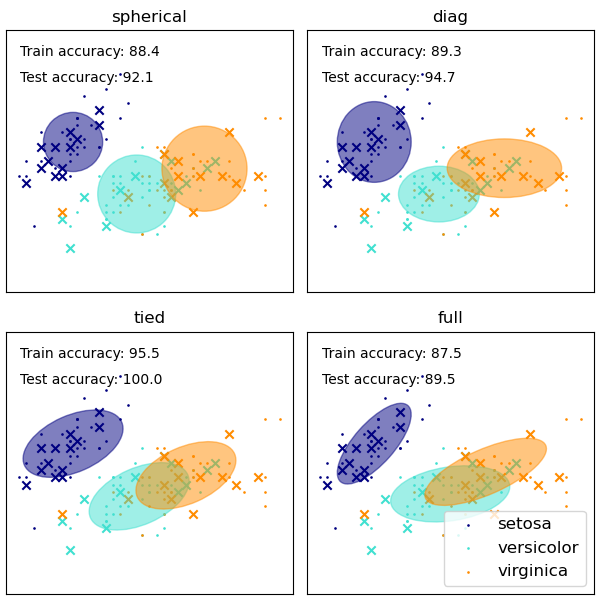
However, most clusters cannot be represented with curricular shape.



K-means doesn't have an intrinsic measure of probability or uncertainty of cluster assignment.

A Gaussian mixture model is a probabilistic model that assumes all the data points are generated from a mixture of a finite number of Gaussian distributions with unknown parameters. One can think of mixture models as generalizing k-means clustering to incorporate information about the covariance structure of the data as well as the centers of the latent Gaussians.

Below models take shapes based on probability distribution.

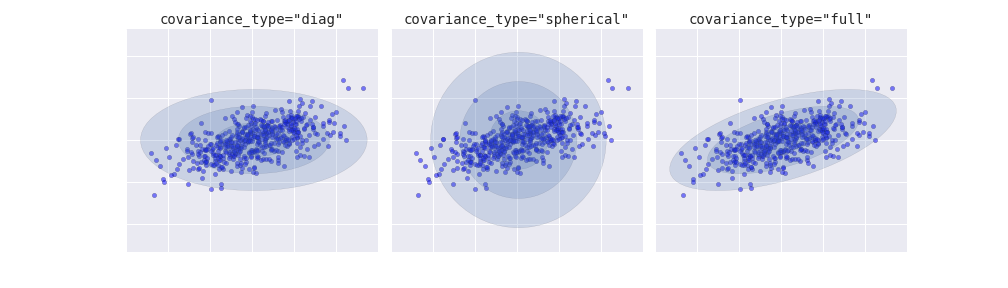


There are three type we can pick in sklearn.

Diag = default → size of the cluster set along the axis

Spherical = considers all dimensions are equal. Similar to k- means but not entirely equivalent.

full= (expensive one, recommended if number of dimensions grows). Take arbitrary elliptical shapes.



Advantages

Flexible can take form of different shapes

Disadvantages

Tuning hyperparameters may require trial and error

Python Sklearn Implementation

Import libraries

Load data

Separate futures and target

(Split data into train and test for some cases to find accuracy, Also we may need to normalize input for some ML algorithm)

Define the model

Fit the Model

Predict