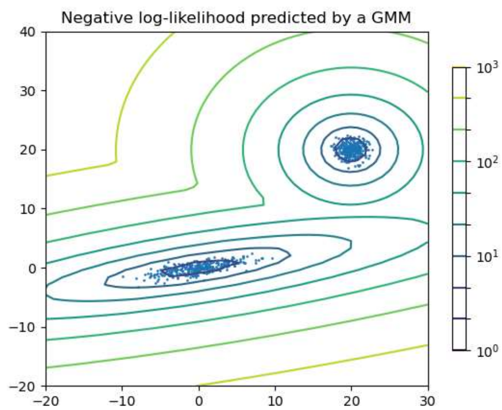
2.1. Gaussian mixture models

`sklearn.mixture` is a package which enables one to learn Gaussian Mixture Models (diagonal, spherical, tied and full covariance matrices supported), sample them, and estimate them from data. Facilities to help determine the appropriate number of components are also provided.



Two-component Gaussian mixture model: *data points, and equi-probability surfaces of the model.* #

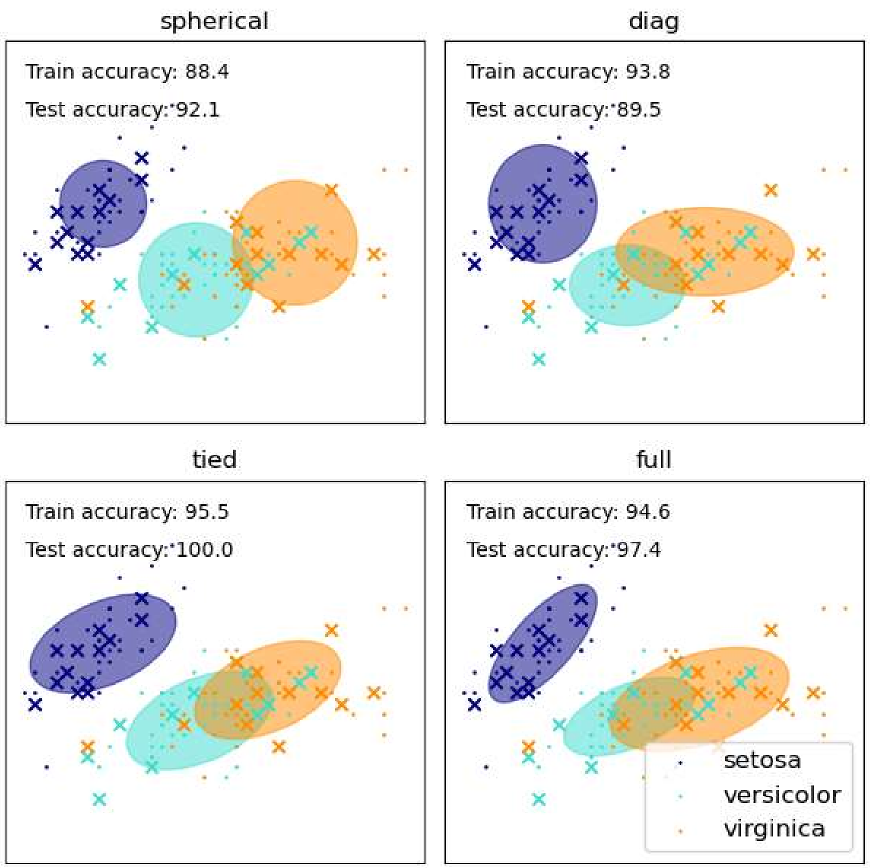
A Gaussian mixture model is a probabilistic model that assumes all the data points are generated from a mixture of a finite number of Gaussian distributions with unknown parameters. One can th of mixture models as generalizing k-means clustering to incorporate information about the covariance structure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

Scikit-learn implements different classes to estimate Gaussian mixture models, that correspond to different estimation strategies, detailed below.

2.1.1. Gaussian Mixture

The **`GaussianMixture`** object implements the [expectation-maximization](#) (EM) algorithm for fitting mixture-of-Gaussian models. It can also draw confidence ellipsoids for multivariate models, and compute the Bayesian Information Criterion to assess the number of clusters in the data. A **`GaussianMixture.fit`** method is provided that learns a Gaussian Mixture Model from train data. Given test data, it can assign to each sample the Gaussian it most probably belongs to using the **`GaussianMixture.predict`** method.

The `GaussianMixture` comes with different options to constrain the covariance of the difference classes estimated: spherical, diagonal, tied or full covariance.



Examples

- See GMM covariances for an example of using the Gaussian mixture as clustering on the iris dataset.
- See Density Estimation for a Gaussian mixture for an example on plotting the density estimation.

Pros and cons of class GaussianMixture

Selecting the number of components in a classical Gaussian Mixture model

Estimation algorithm expectation-maximization

Choice of the Initialization method

2.1.2. Variational Bayesian Gaussian Mixture

The `BayesianGaussianMixture` object implements a variant of the Gaussian mixture model with variational inference algorithms. The API is similar to the one defined by `GaussianMixture` .

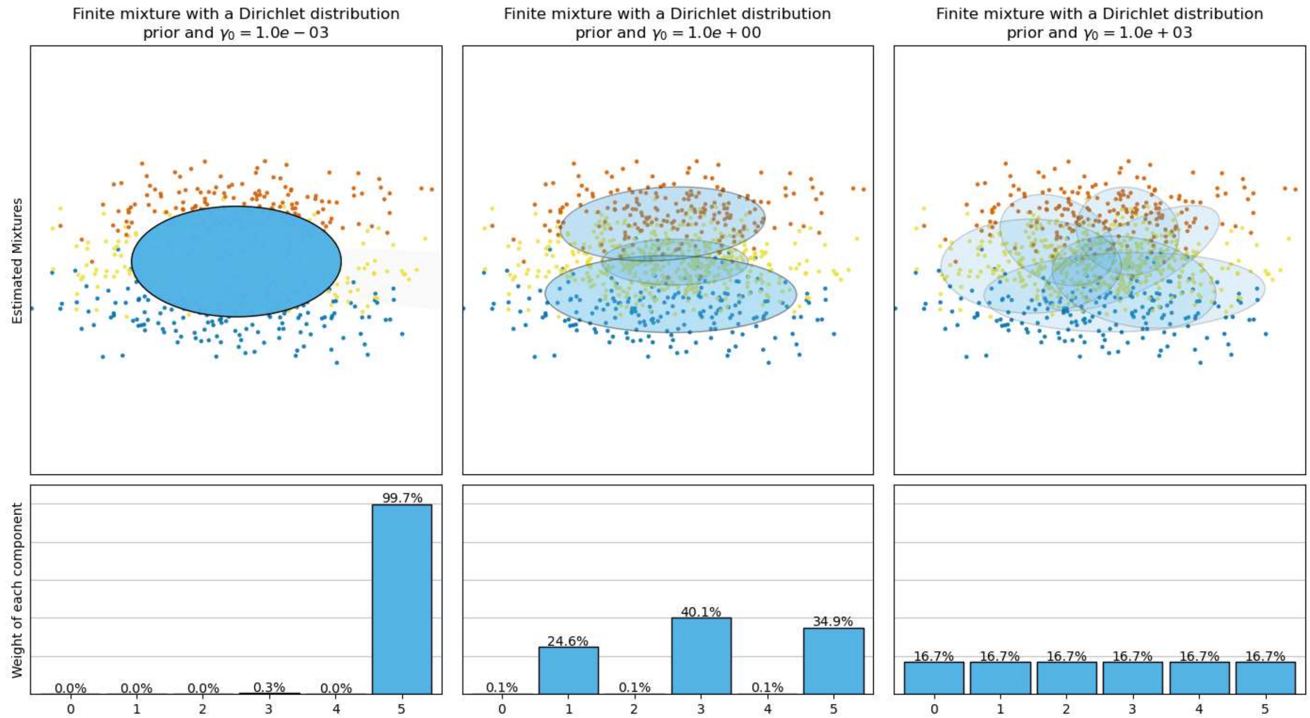
Estimation algorithm: variational inference

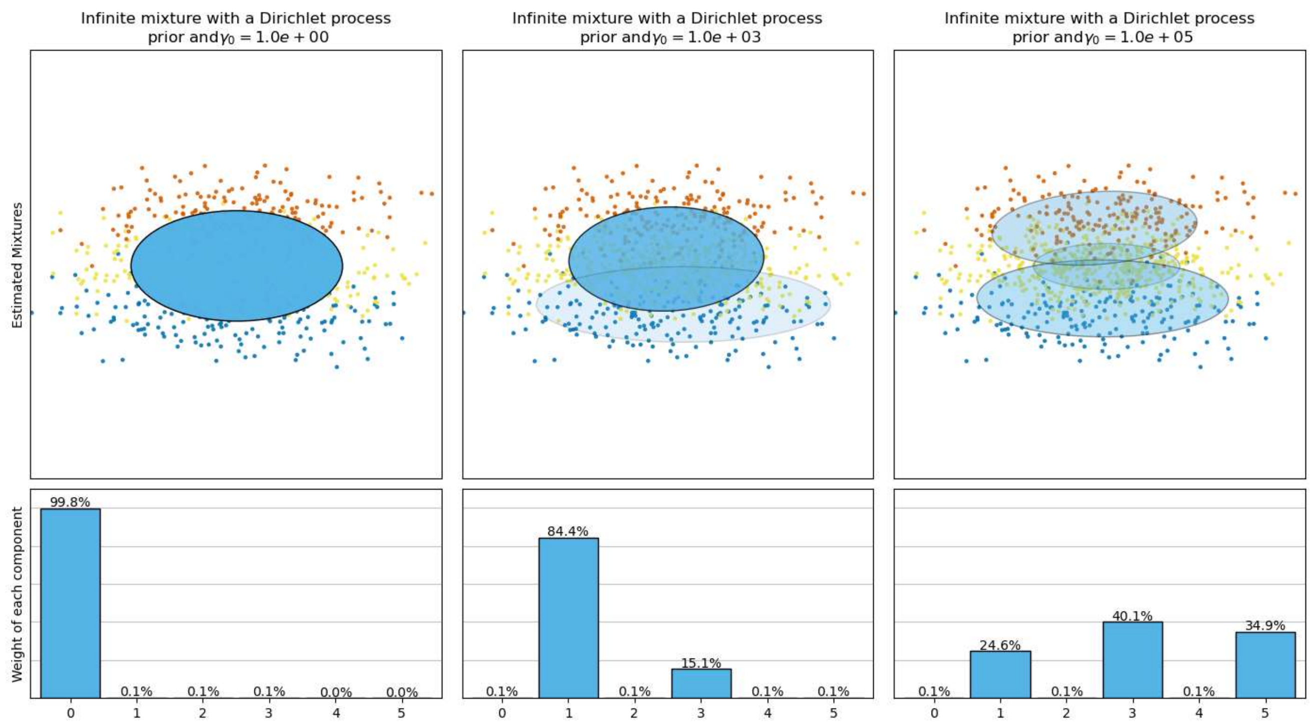
Variational inference is an extension of expectation-maximization that maximizes a lower bound o model evidence (including priors) instead of data likelihood. The principle behind variational methods is the same as expectation-maximization (that is both are iterative algorithms that altern between finding the probabilities for each point to be generated by each mixture and fitting the mixture to these assigned points), but variational methods add regularization by integrating information from prior distributions. This avoids the singularities often found in expectation-maximization solutions but introduces some subtle biases to the model. Inference is often notably slower, but not usually as much so as to render usage unpractical.

Due to its Bayesian nature, the variational algorithm needs more hyperparameters than expectatic maximization, the most important of these being the concentration parameter `weight_concentration_prior` . Specifying a low value for the concentration prior will make the model put most of the weight on a few components and set the remaining components' weights very close to zero. High values of the concentration prior will allow a larger number of component to be active in the mixture.

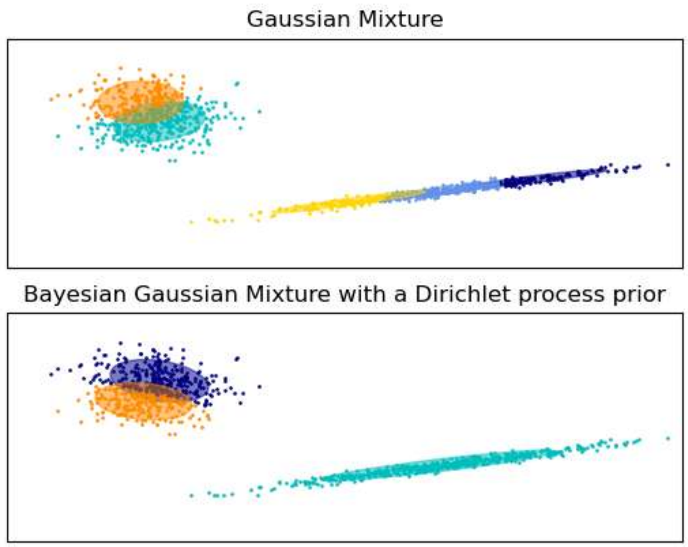
The parameters implementation of the `BayesianGaussianMixture` class proposes two types of pric for the weights distribution: a finite mixture model with Dirichlet distribution and an infinite mixtu model with the Dirichlet Process. In practice Dirichlet Process inference algorithm is approximated and uses a truncated distribution with a fixed maximum number of components (called the Stick-breaking representation). The number of components actually used almost always depends on the data.

The next figure compares the results obtained for the different type of the weight concentration prior (parameter `weight_concentration_prior_type`) for different values of `weight_concentration_prior` . Here, we can see the value of the `weight_concentration_prior` parameter has a strong impact on the effective number of active components obtained. We can al notice that large values for the concentration weight prior lead to more uniform weights when the type of prior is 'dirichlet_distribution' while this is not necessarily the case for the 'dirichlet_process type (used by default).

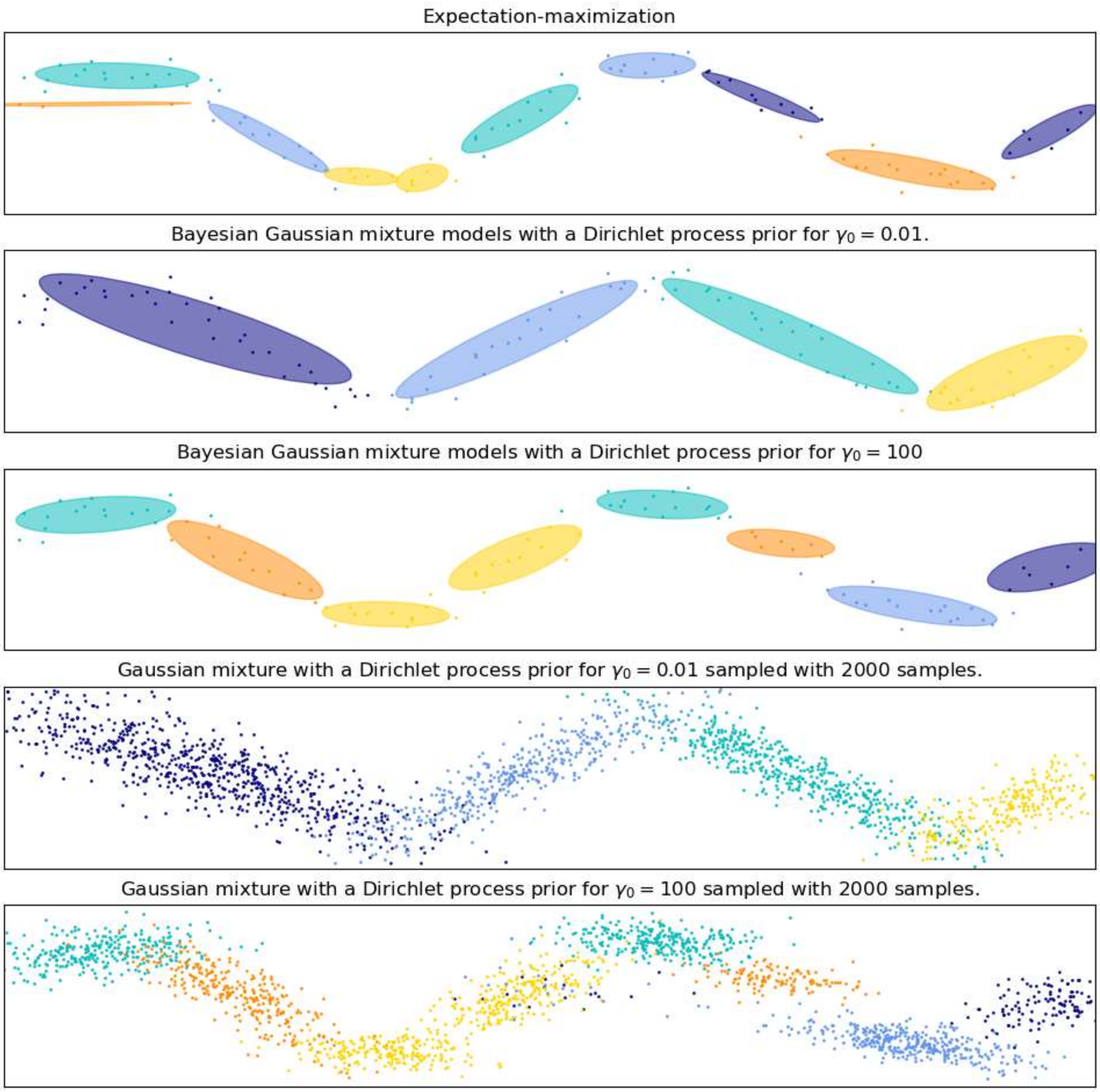




The examples below compare Gaussian mixture models with a fixed number of components, to th variational Gaussian mixture models with a Dirichlet process prior. Here, a classical Gaussian mixtu is fitted with 5 components on a dataset composed of 2 clusters. We can see that the variational Gaussian mixture with a Dirichlet process prior is able to limit itself to only 2 components whereas the Gaussian mixture fits the data with a fixed number of components that has to be set a priori b the user. In this case the user has selected `n_components=5` which does not match the true generative distribution of this toy dataset. Note that with very little observations, the variational Gaussian mixture models with a Dirichlet process prior can take a conservative stand, and fit only one component.



On the following figure we are fitting a dataset not well-depicted by a Gaussian mixture. Adjusting the `weight_concentration_prior`, parameter of the `BayesianGaussianMixture` controls the numb of components used to fit this data. We also present on the last two plots a random sampling generated from the two resulting mixtures.



Examples

- See Gaussian Mixture Model Ellipsoids for an example on plotting the confidence ellipsoids fo both `GaussianMixture` and `BayesianGaussianMixture` .
- Gaussian Mixture Model Sine Curve shows using `GaussianMixture` and `BayesianGaussianMixture` to fit a sine wave.
- See Concentration Prior Type Analysis of Variation Bayesian Gaussian Mixture for an example plotting the confidence ellipsoids for the `BayesianGaussianMixture` with different `weight_concentration_prior_type` for different values of the parameter `weight_concentration_prior` .

Pros and cons of variational inference with BayesianGaussianMixture

2.1.2.1. The Dirichlet Process

Here we describe variational inference algorithms on Dirichlet process mixture. The Dirichlet proce is a prior probability distribution on *clusterings with an infinite, unbounded, number of partitions*. Variational techniques let us incorporate this prior structure on Gaussian mixture models at almost no penalty in inference time, comparing with a finite Gaussian mixture model.

An important question is how can the Dirichlet process use an infinite, unbounded number of clusters and still be consistent. While a full explanation doesn't fit this manual, one can think of its stick breaking process analogy to help understanding it. The stick breaking process is a generative story for the Dirichlet process. We start with a unit-length stick and in each step we break off a