Mixture Models

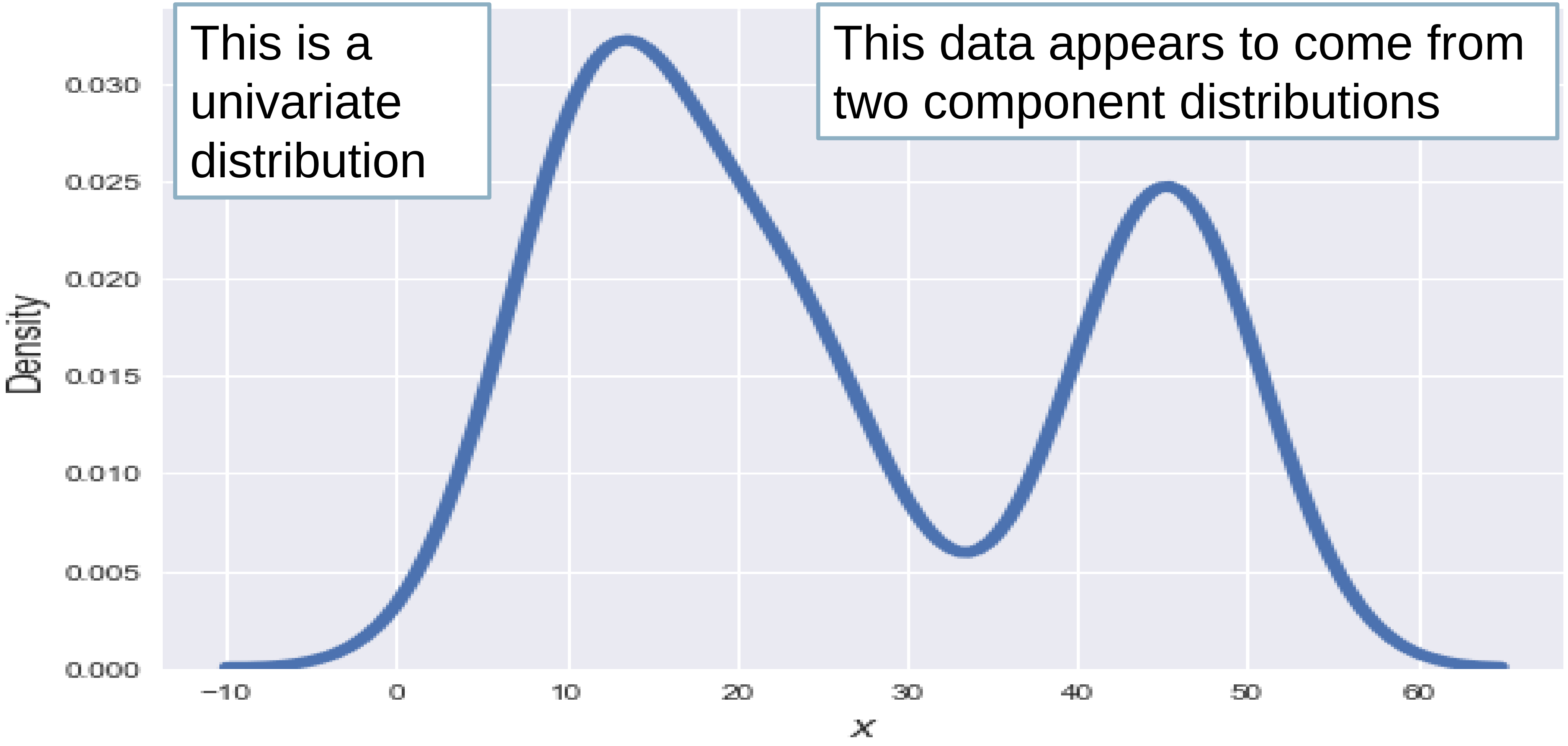
Neba Nfonsang
University of Denver

Mixture Models

- Sometimes, a phenomena or process under study cannot be properly described using a single distribution or model.
- A mixture of models may be more appropriate to describe a phenomena.
- A mixture model consists of a combination of models.
- A mixture model is a mixture of two or more probability distributions.
- The sub distributions represent subpopulations or subgroups in the data.

Osvaldo Martin (2016)

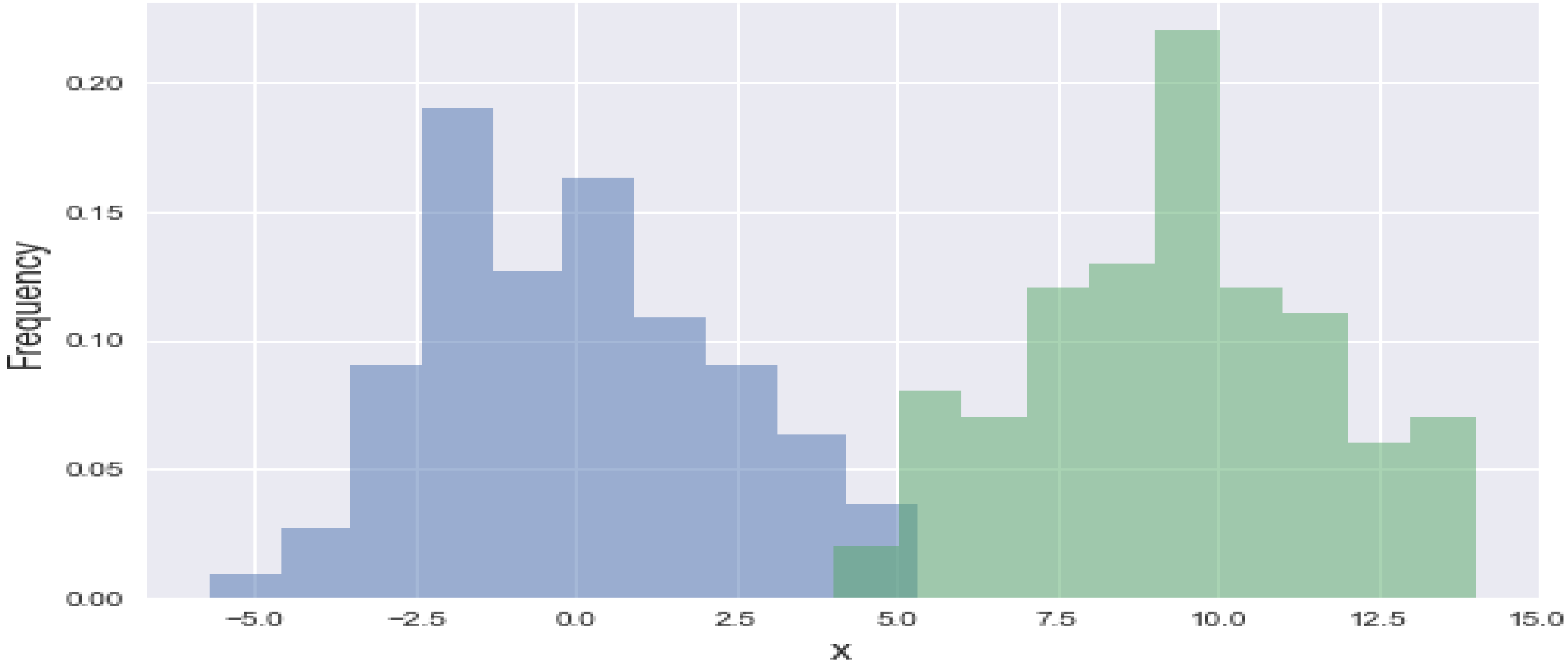
Mixture Model



This is a univariate distribution

This data appears to come from two component distributions

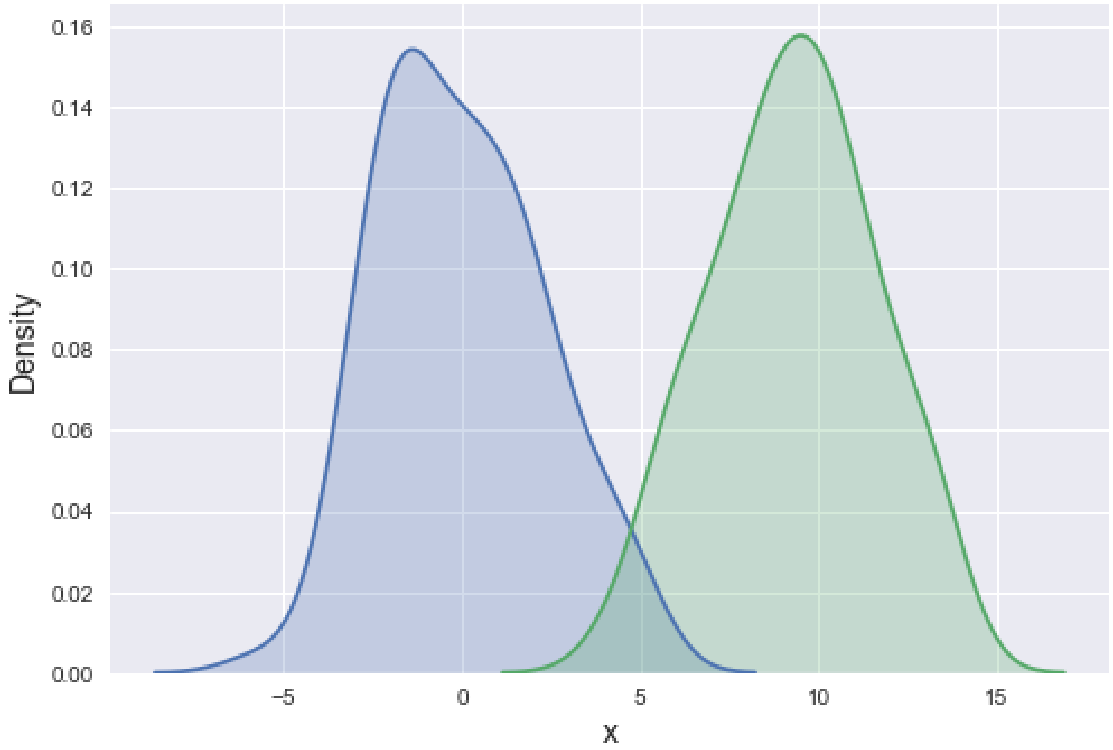
Mixture Models



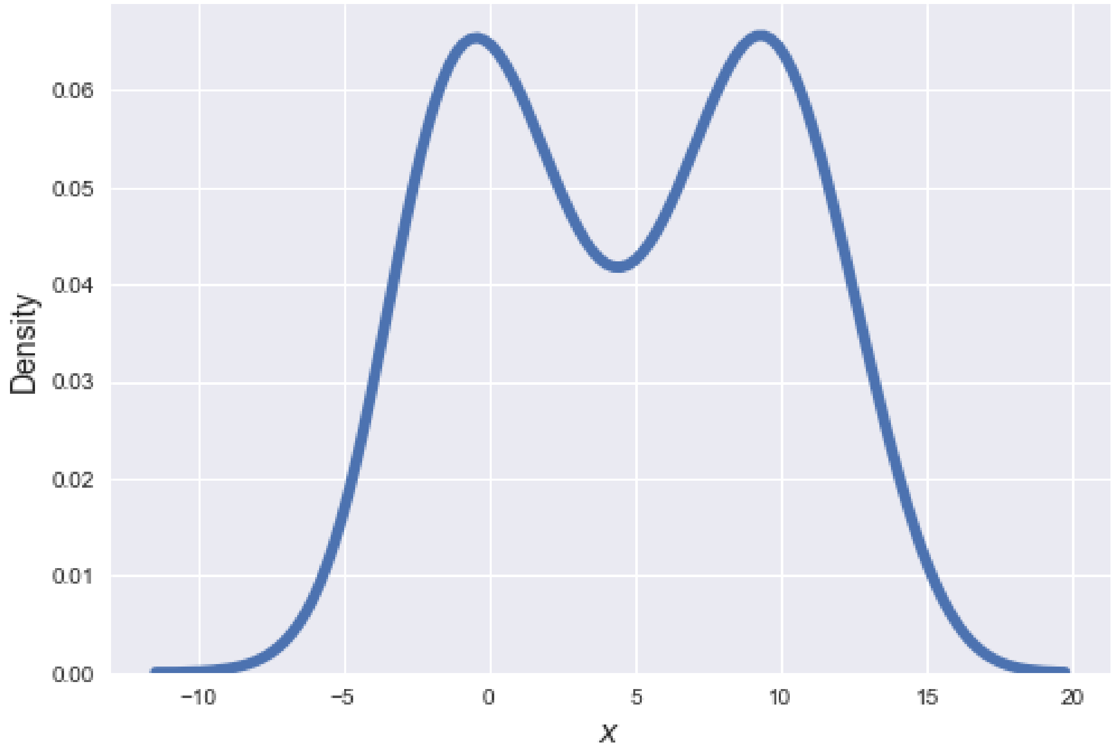
The data has two underlying univariate distributions that make up the univariate mixture distribution

Mixture Models

Underlying univariate distributions



Univariate mixture model



Mixture Models

- Models that assume that the data come from a mixture of distributions (or models) is called mixture model.
- The underlying distributions in the data can:
  - have the same functional form with different parameters,
  - have different functional forms.
- If the underlying models that make up a mixture model are of different forms, the distributions should:
  - have the same dimensions
  - be all discrete or all continuous (for example, the mixture model could consist of a mixture of a normal and t-distributions).

Mixture Models

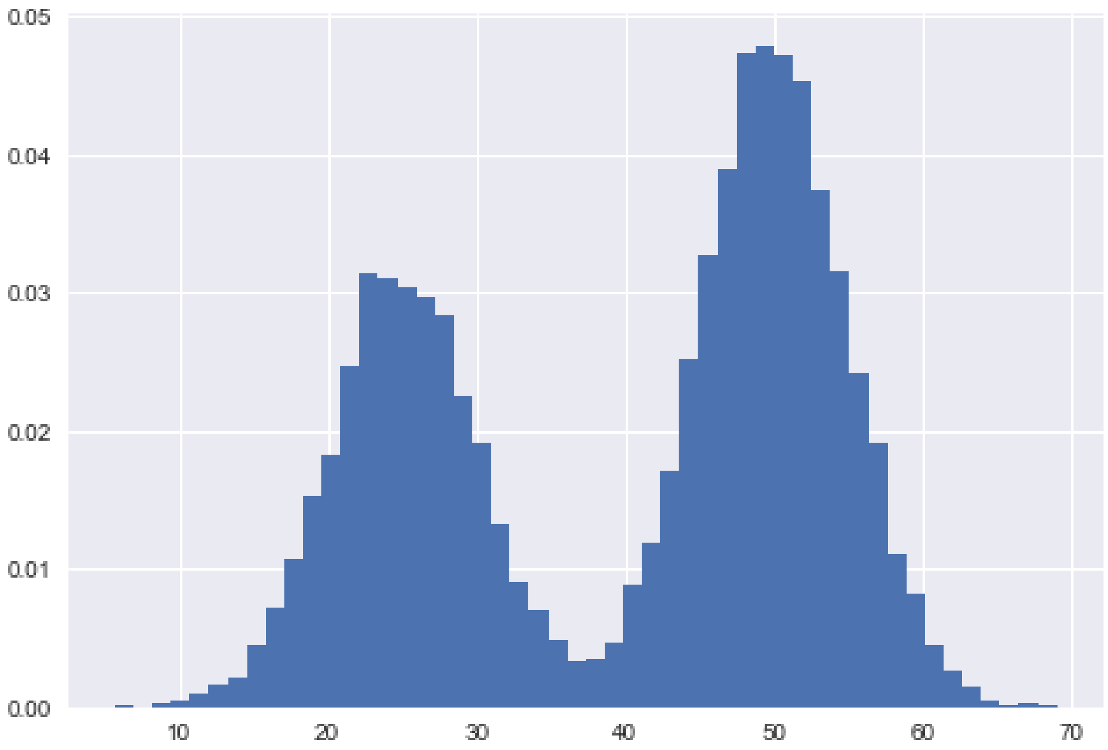
- A mixture model with k component distribution where k << n data points is called a finite mixture model.
- Formally, a distribution f is a mixture of k component distributions f_1, f_2 ... f_k if:

- $f(x) = \sum_{k=1}^{k} \lambda_k f_k(x)$ where
  - $\lambda_k = \text{proportion of data points in group k}$
  - λ_k is also called mixing weights
  - P($\mathrm{z} = k$) = λ_k = probability of a point belonging to group k in the dataset (prevalence)
  - $\sum_{k=1}^{k} \lambda_k = 1$
  - $0 \leq \lambda_k \leq 1$

Mixture Models

Mixture modeling can be used for:

- Density estimation: that is, to estimate an empirical distribution using a mixture of distributions.
- Clustering: That is, to assign data points to an underlying distribution or subgroups in the data in a probabilistic way.



Mixture Models

Fitting Distributions for Subpopulations

- Mixture models can be used to model subpopulations or describe datasets that consist of real subpopulations.
- That means, the parameters of the sub distributions can be estimated.

Fitting Models for Subpopulations

- Mixture models are not limited to estimating parameters of sub distributions but could also be used to estimate parameters of models such as regression model, etc., across unobserved groups.

Mixture Models

Approximating a Mixture Model

- Also, mixture models can be used to mathematically approximate empirical distributions by combining simpler distributions.

- For example, the kernel density estimation technique sums individual Gaussians to approximate the mixture model.
- $f(x) = \sum_{k=1}^{k} \lambda_k f_k(x)$
  - Mixture model = weighted average of the distributions

Mixture Models

Probabilistic Clustering

- Mixture modeling can be viewed as a model-based approach to clustering through the use of statistical distributions.

- Mixture modeling tries to group or assign data points into subgroups or clusters.
- Data points in each cluster are assumed to:
  - be similar
  - come from the same distribution
  - represent a subpopulation

Mixture Models

Subgroups or Clusters

- The subgroups in a mixture model could be age groups, education level, income brackets, ethnic groups, risk level, etc.

- Sometimes, we don't know the group variable (unobserved or latent) but we could find the parameters of the distribution for each subgroups.
  - The unobserved group variable is an indicator latent variable.
  - The indicator variable assigns a group to each data point.

Mixture Models

Mixture Model and Clustering

- A mixture model can be fitted into a dataset using the EM algorithm.
- That is, the parameters of the distributions that make up the mixture model are estimated using the EM algorithm.
- Once the parameters of the distributions in the mixture model are estimated:
  - The mixture model is then used to predict the latent variable value of each data point. Each data points is assigned to a specific subgroup.

Mixture Model Example

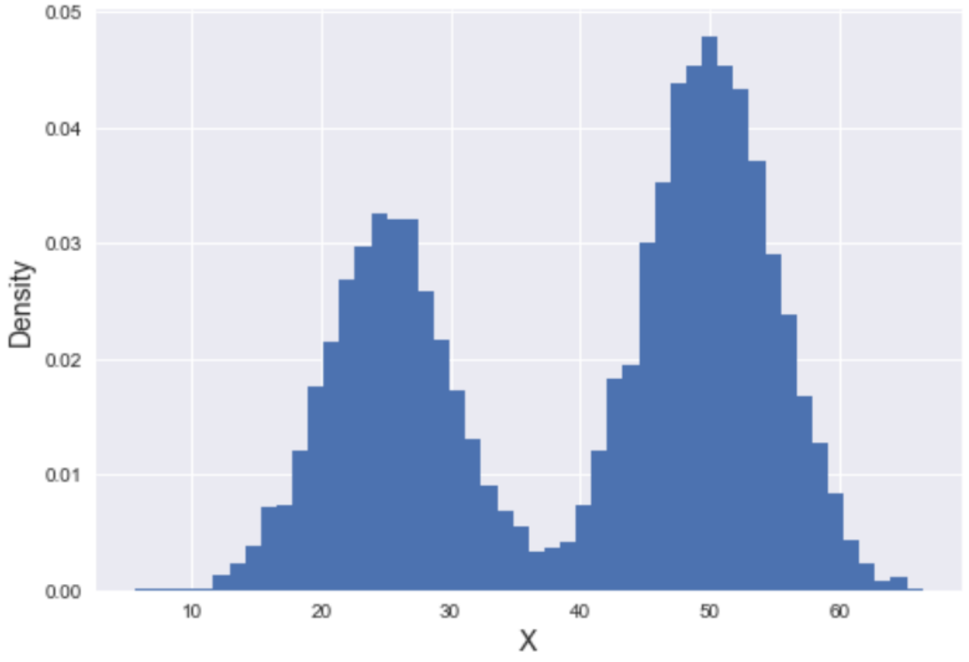
```
# Generate data from two mixture distributions
np.random.seed(1234)
X1 = np.random.normal(loc=25, scale=5, size=4000)
X2 = np.random.normal(loc=50, scale=5, size=6000)
X = np.hstack((X1, X2)).round(2)
# view the first 50 data points
X[0:50]
```

The two distributions are Gaussian distributions.

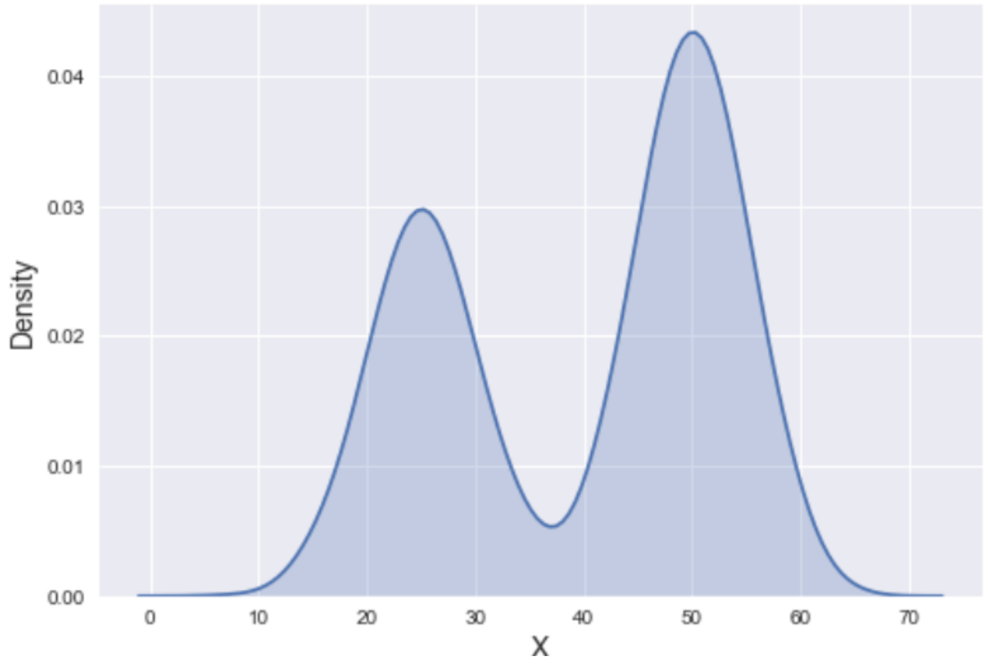
```
array([27.36, 19.05, 32.16, 23.44, 21.4 , 29.44, 29.3 , 21.82, 25.08,
       13.79, 30.75, 29.96, 29.77, 14.89, 23.33, 25.01, 27.03, 26.45,
       31.61, 17.27, 23.99, 21.72, 25.97, 27.77, 31.59, 22.65, 28.38,
       15.91, 24.08, 30.29, 23.01, 26.69, 30.24, 30.23, 29.32, 24.39,
       25.62, 23.39, 29.21, 36.95, 25.38, 22.17, 25.18, 14.63, 26.24,
       20.51, 24.32, 25.09, 28.78, 26.08])
```

Mixture Model Example

```
plt.hist(X, bins=50, density=True)
plt.xlabel("X", fontsize=14)
plt.ylabel("Density", fontsize=14);
```




```
sns.kdeplot(X, shade=True)
plt.xlabel("X", fontsize=14)
plt.ylabel("Density", fontsize=14);
```



Mixture Model Example

```
from sklearn.mixture import GaussianMixture

# fit the Gaussian mixture model
# for a univariate distribution, reshape to a column vector
X = X.reshape(len(X), 1)
model = GaussianMixture(n_components=2, max_iter=1000, random_state=0)
model.fit(X)
```

```
GaussianMixture(covariance_type='full', init_params='kmeans', max_iter=1000,
        means_init=None, n_components=2, n_init=1, precisions_init=None,
        random_state=0, reg_covar=1e-06, tol=0.001, verbose=0,
        verbose_interval=10, warm_start=False, weights_init=None)
```

Mixture Model Example

```
# estimated means
model.means_
```

```
array([[50.09062186],
       [25.14853071]])
```

```
# actual means
print(np.mean(X1))
print(np.mean(X2))
```

```
25.10460209547166
50.06465243674555
```

```
# estimated covariances
## variances of univariate distributions
model.covariances_
```

```
array([[[24.76177353]],

       [[24.84105672]]])
```

```
# actual covariance
print(np.var(X1))
print(np.var(X2))
```

```
24.233187715331105
25.109863320433266
```

The means, covariances (variances for univariate distributions) and weights were estimated and compared with the actual parameters from the simulated data.

Mixture Model Example

The mixing weight of the each distribution represents the proportion of data points assigned to the distribution.

```
# estimated weights
model.weights_
```

```
array([0.59867139, 0.40132861])
```

```
# actual weights
len(X1)/len(X) ,len(X2)/len(X)
```

```
(0.4, 0.6)
```

Mixture Model Example

```python
# predit the latent values of each data point
predictions = model.predict(X)
print(predictions[0:150])
print(predictions[-150:])
```

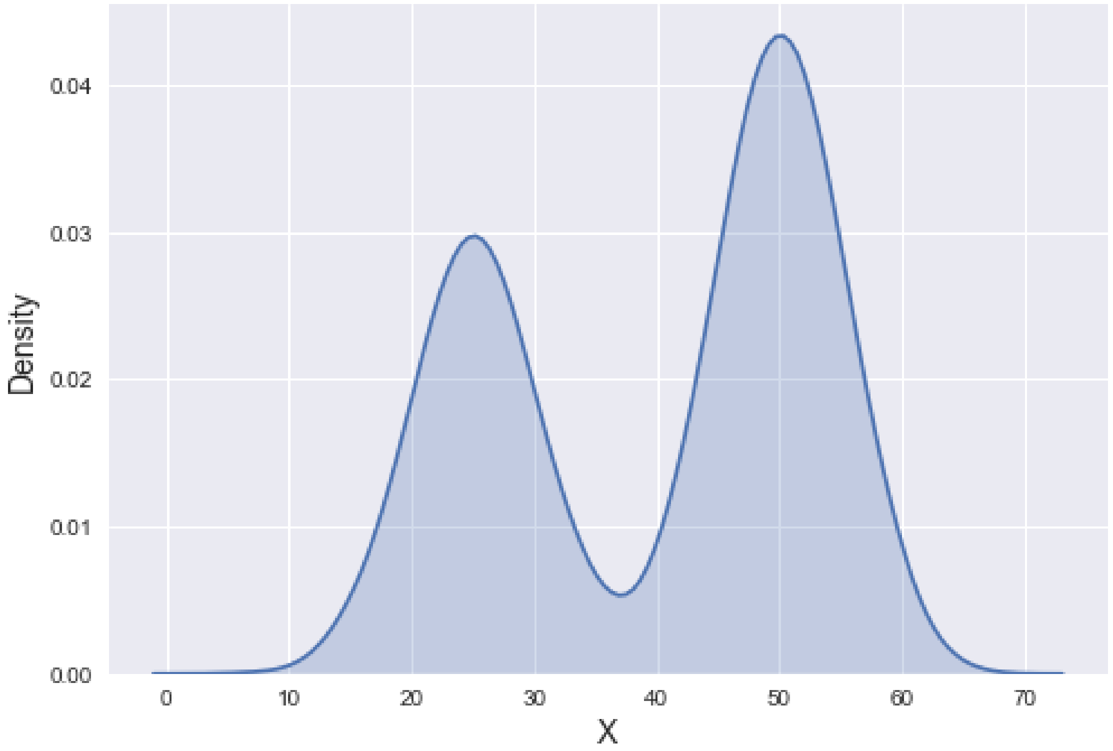
```
[0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
 0 0]
[1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
 1 0 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
 1 1]
```

Gaussian Mixture Models

- A Gaussian mixture model is a combination of underlying Gaussian distributions.
- That is, a Gaussian Mixture model consists of a mixture of Gaussian or normal distributions.
- The Gaussians distributions in a mixture model can be univariate or multivariate distributions
  - a univariate normal distribution models data for a single variable.
  - a multivariate normal distribution models data points in a high dimensional space (with many variables).

Univariate Gaussian Mixture Models

- A univariate Gaussian mixture model consist of a mixture of univariate normal distributions.
- The mixture model can be written as a weighted sum of the underlying normal densities:
  - $f(x) = \sum_{i=1}^{k} \lambda_i f_i(x)$
  - $for\ k = 2{:}\ f(x) = \lambda_1 f_1(x) + \lambda_1\ f_1(x)$
  - f_i is distribution of the ith class, group or component



Univariate Gaussian Mixture Models

- A random variable, X, with mean μ, and variance δ^2 is said to follows a Gaussian distribution if it's probability density function is:
  - $f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\delta}\right)^2} = N(u, \sigma),\ -\infty < x < \infty$
- A Univariate Gaussian Mixture model f is given by:
  - $f(x) = \lambda_1 f_1(x) + \lambda_1\, f_1(x);\ \ for\ k = 2$ component distributions
  - $f(x) = \lambda_1 N(\mu_1, \sigma_1) + \lambda_2 N(\mu_2, \sigma_2)$
  - $f(x) = \lambda_1 \frac{1}{\sqrt{2\pi\sigma_1^2}} e^{-\frac{1}{2}\left(\frac{x-\mu_1}{\sigma_1}\right)^2} + \lambda_2 \frac{1}{\sqrt{2\pi\sigma_2^2}} e^{-\frac{1}{2}\left(\frac{x-\mu_2}{\sigma_2}\right)^2}$

Univariate Gaussian Mixture Models

Generally, a univariate Gaussian mixture model f is written as a combination of k components f_1, f_2 ... f_k :

$$f(x) = \sum_{i=1}^{k} \lambda_i \frac{1}{\sqrt{2\pi\sigma_i^2}} e^{-\frac{1}{2}\left(\frac{x-\mu_i}{\sigma_i}\right)^2}$$

- Where f_1, f_2 ... f_k are k underlying distributions in the mixture model f.

Multivariate Gaussian Mixture Models

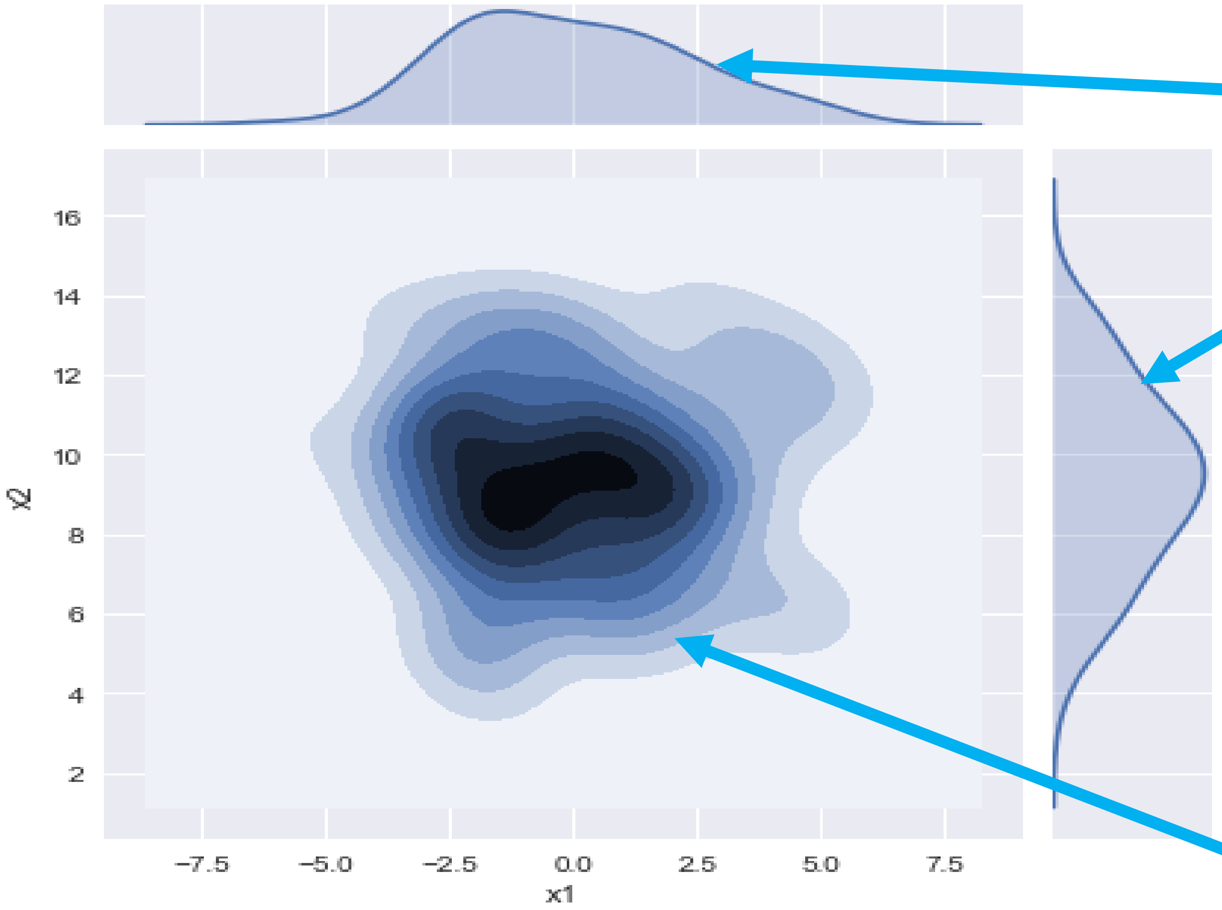
- A multivariate Gaussian mixture model consist of a mixture of multivariate normal distributions.
- A multivariate normal distribution describes the joint distribution of several random variables in a high dimensional space.
- A multivariate normal distribution is parametrized by:
  - a vector of means (mean vector) of the random variables, μ.
  - a variance-covariance matrix for the random variables.
- The multivariate normal distribution is also called the multinormal distribution.

Multivariate Gaussian Mixture Models

Multivariate Normal Distribution:

- The multivariate normal distribution is a generalization of the univariate normal distribution to two or more random variables.
- The multivariate normal distribution is a distribution of a vector of random variables where each vector element or random variable has a univariate normal distribution.

Multivariate Gaussian Mixture Models

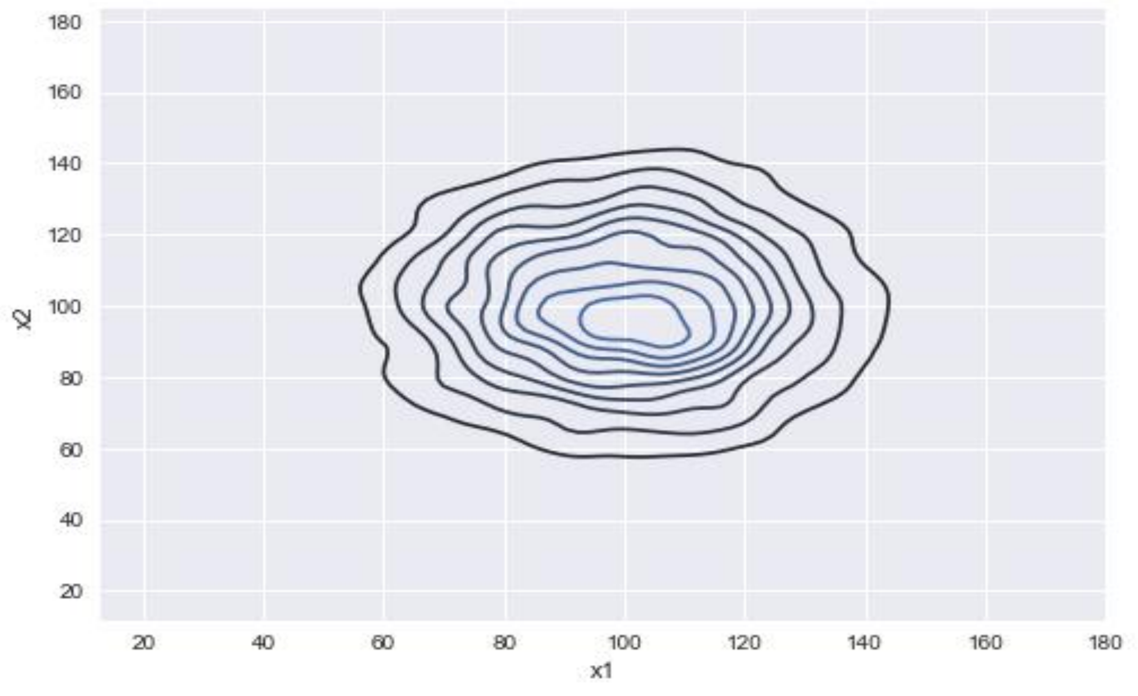


Multivariate Normal Distribution:
In a multivariate normal distribution, each random variable has a univariate normal distribution.

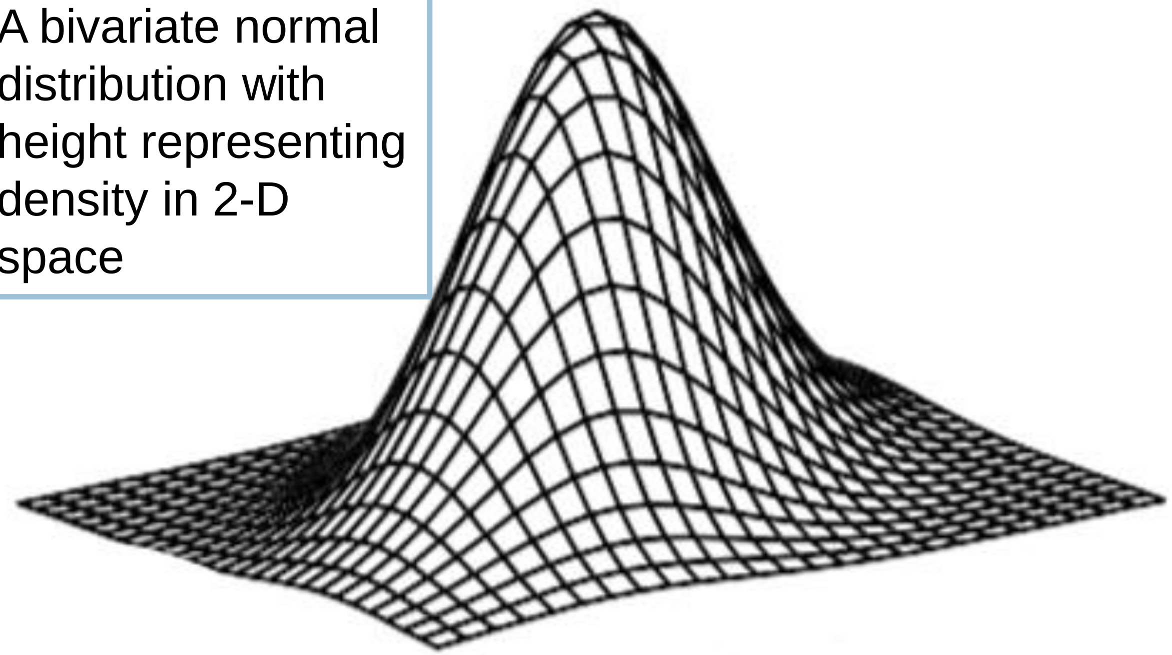
A density contour plot can also be used to represent multivariate distribution where each contour line represents points having the same density

Multivariate Gaussian Mixture Models

A bivariate normal distribution: a multivariate normal distribution in a 2-dimensional space (with two variables).

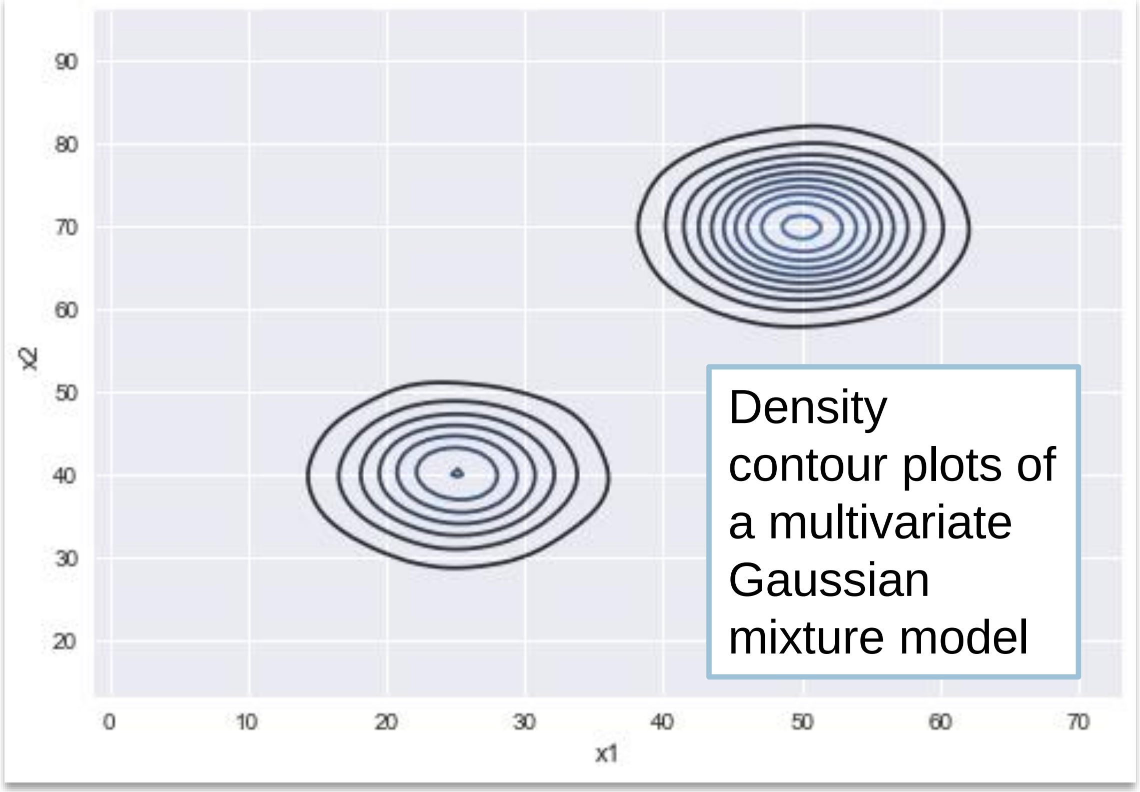


A bivariate normal distribution with height representing density in 2-D space

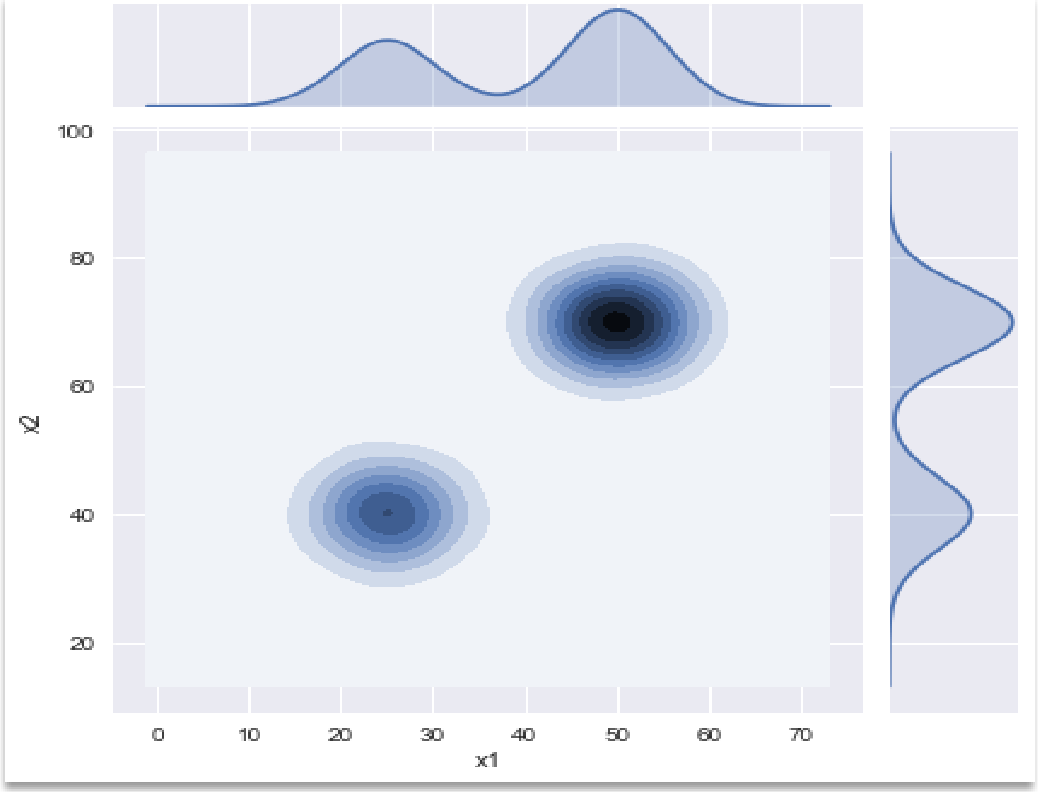


Multivariate Gaussian Mixture Models

Visualizing a multivariate Gaussian mixture model with two components in 2-D Space

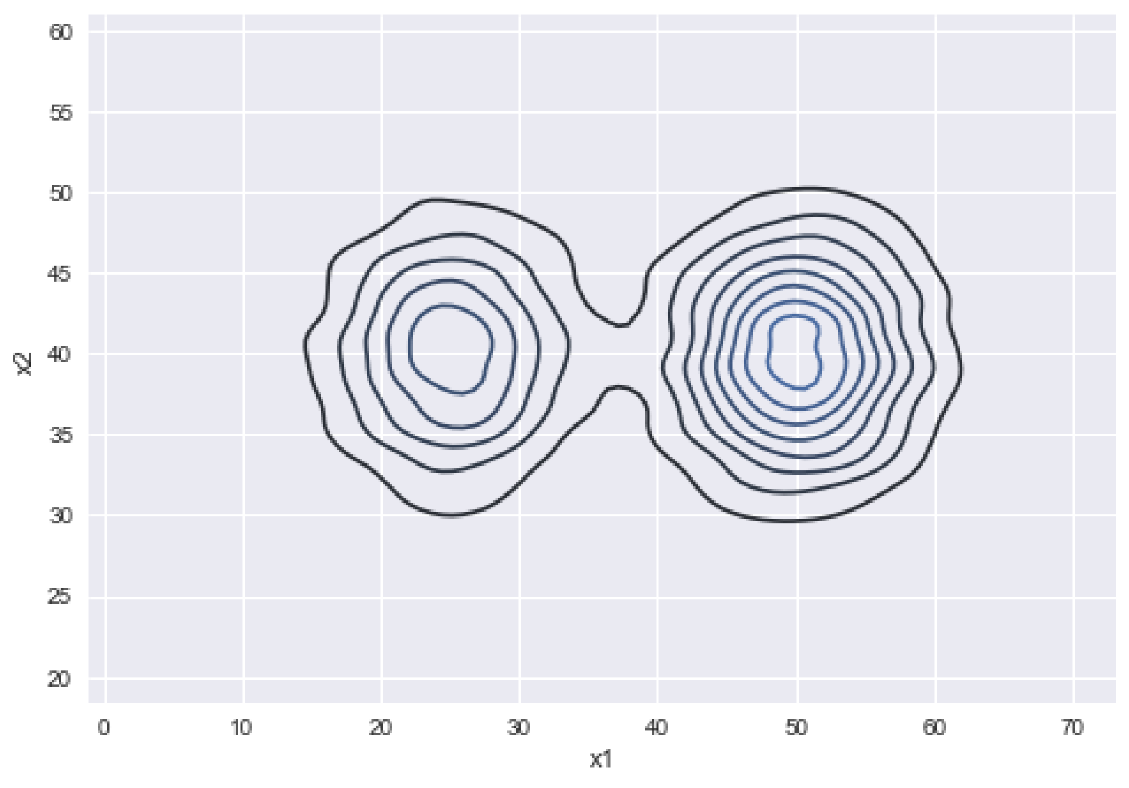


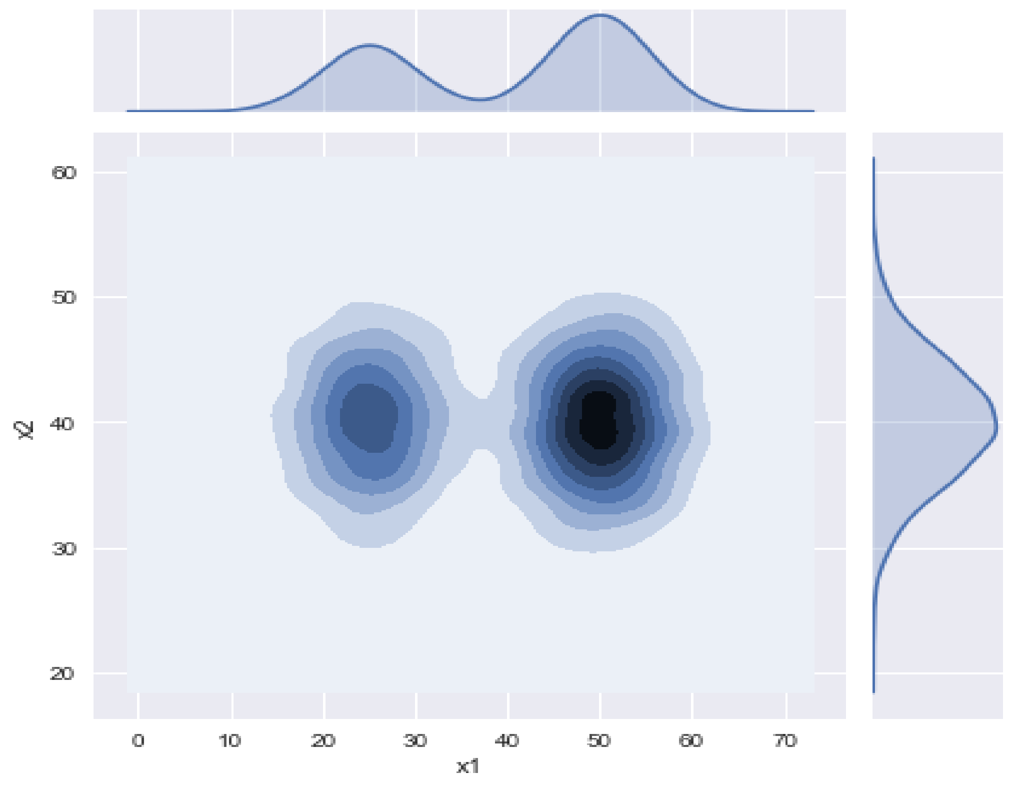
Density contour plots of a multivariate Gaussian mixture model



Multivariate Gaussian Mixture Models

Visualizing a multivariate Gaussian mixture model with two components in 2-D Space





Multivariate Gaussian Mixture Models

Multivariate Normal Distribution:

- A vector of random variables X = $\begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_d \end{bmatrix} = [X_1 \ X_2 \cdots X_d]^T$ with a mean vector $\mu = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_d \end{bmatrix} = [\mu_1 \, \mu_2 \cdots \mu_d]^T$ and covariance matrix Σ (d x d dimensions), follows a multivariate normal distribution if its density is:
- $f(\boldsymbol{x}) = p(\boldsymbol{x};\ \mu, \Sigma) = \dfrac{1}{\sqrt{(2\pi)^d |\Sigma|}} e^{-\frac{1}{2}(\boldsymbol{x}-\mu)^T \Sigma^{-1} (\boldsymbol{x}-\mu)}$

Multivariate Gaussian Mixture Models

Multivariate Normal Distribution:

$$f(\boldsymbol{x}) = p(\boldsymbol{x};\ \mu, \Sigma) = \frac{1}{\sqrt{(2\pi)^d |\Sigma|}} e^{-\frac{1}{2}(\boldsymbol{x}-\mu)^T \Sigma^{-1} (\boldsymbol{x}-\mu)}$$

- $\boldsymbol{x} = [x_1\ x_2 \cdots x_d]^T$ = a data point in a d-dimensional space, $x \in \mathbb{R}^{dx1}$
- $\mu\ = [\mu_1\, \mu_2 \cdots \mu_d]^T$ = mean vector; ; $\mu \in \mathbb{R}^{dx1}$
- Σ = covariance matrix where diagonal elements are variances of random variables while off-diagonal elements are covariances between variables.
- $\Sigma = \begin{bmatrix} \sigma_{11}^2 & \sigma_{12}^2 & \cdots & \sigma_{1d}^2 \\ \sigma_{21}^2 & \sigma_{22}^2 & \cdots & \sigma_{2d}^2 \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{d1}^2 & \sigma_1^2 & \cdots & \sigma_{dd}^2 \end{bmatrix}$ $\Sigma = \begin{bmatrix} \sigma_{11}^2 & 0 & \cdots & 0 \\ 0 & \sigma_{22}^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_{dd}^2 \end{bmatrix}$

Random variables are correlated

Typically used: random variables are assumed to be independent (no correlation).

Multivariate Gaussian Mixture Models

How to Derive the Multivariate Normal Density

- The multivariate normal density function $f(\boldsymbol{x})$ of the multivariate normal distribution is a joint normal density function of the random variables in the multivariate space:
- $f(\boldsymbol{x}) = f(x_1, x_2, \ldots x_d) = f(x_1) * f(x_2) * \cdots * f(x_d)$

$$= \frac{1}{\sqrt{2\pi\sigma_1^2}} e^{-\frac{1}{2}\left(\frac{x_1-\mu_1}{\sigma_1}\right)^2} * \frac{1}{\sqrt{2\pi\sigma_2^2}} e^{-\frac{1}{2}\left(\frac{x_2-\mu_2}{\sigma_2}\right)^2} * \cdots * \frac{1}{\sqrt{2\pi\sigma_d^2}} e^{-\frac{1}{2}\left(\frac{x_d-\mu_d}{\sigma_d}\right)^2}$$

$$= \left(\frac{1}{\sqrt{2\pi\sigma_1^2}}\right) * \left(\frac{1}{\sqrt{2\pi\sigma_2^2}}\right) * \cdots * \left(\frac{1}{\sqrt{2\pi\sigma_d^2}}\right) * e^{\sum_{i=1}^{d} -\frac{1}{2}\left(\frac{x_i-\mu_i}{\sigma_i}\right)^2}$$

Multivariate Gaussian Mixture Models

How to Derive the Multivariate Normal Density

- $f(\boldsymbol{x}) = (\frac{1}{2\pi})^{1/d} * (\sigma_1^2 * \sigma_2^2 * \cdots \sigma_d^2)^{-\frac{1}{2}} * e^{\sum_{i=1}^{d} -\frac{1}{2}\left(\frac{x_i - \mu_i}{\sigma_i^2}\right)^2}$
- $f(\boldsymbol{x}) = (\frac{1}{2\pi})^{1/d} * |\Sigma|^{-\frac{1}{2}} * e^{-\frac{1}{2}(\boldsymbol{x}-\mu)^T \Sigma^{-1}(\boldsymbol{x}-\mu)}$
- $f(\boldsymbol{x}) = p(\boldsymbol{x};\ \mu, \Sigma) = \frac{1}{\sqrt{(2\pi)^d |\Sigma|}} e^{-\frac{1}{2}(\boldsymbol{x}-\mu)^T \Sigma^{-1}(\boldsymbol{x}-\mu)}$

$\boldsymbol{x} = [x_1\ x_2 \cdots x_d]^T$ = data point; $x \in \mathbb{R}^{dx1}$
$\mu\ = [\mu_1\ \mu_2 \cdots \mu_d]^T$ = mean vector; ; $\mu \in \mathbb{R}^{dx1}$

Σ^{-1} = inverse of the covariance matrix.

$|\Sigma|$ = determinant of the covariance matrix.

Multivariate Gaussian Mixture Models

- A multivariate Gaussian mixture model f made up of k-components, $f_1, f_2,.., f_k$ is given by:
  - $f(\boldsymbol{x}) = \sum_{i=1}^{k} \lambda_i f_i(\boldsymbol{x};\ \mu_i, \Sigma_i)$
  - $$f(\boldsymbol{x}) = \sum_{i=1}^{k} \lambda_i \frac{1}{\sqrt{(2\pi)^d |\Sigma_i|}} e^{-\frac{1}{2}(\boldsymbol{x}-\mu_i)^T {\Sigma_i}^{-1}(\boldsymbol{x}-\mu_i)}$$

$\boldsymbol{x} = [x_1\ x_2 \cdots x_d]^T$ = data point; $\boldsymbol{x} \in \mathbb{R}^{dx1}$
$\mu_i = [\mu_{\mathrm{i}1}\ \mu_{i2} \cdots \mu_{id}]^T$ = mean vector of ith component; ; $\mu_i \in \mathbb{R}^{dx1}$

- Σ_i = covariance matrix of the ith component.
- μ_i = mean vector of the ith multivariate component.
- λ_i = weight of each component
- Note that each component is a multivariate normal distribution.