Anomaly Detection

- it is the process of identifying patterns that do not conform the expected behavior.

Gaussian Mixtures

- A gaussian mixture is a probabilistic model that assumes that the instance were generated from a mixture of several Gaussian distributions whose patters are unknown.
- All the instances generated form a single Gaussian distribution from a cluster that looks like an ellipsoid. Each having a different shape, size, density and orientation.
- When you observe an instance, you know it was generated from one of the Gaussian distributions, but you are not told which one, and you do not know what the parameters of these distributions are.

There are several GMM variants , we are supposed to know in advance the number k of Gaussian distribution.
Process

1. **Choosing a Cluster**:
    - For each data point (or instance), we randomly pick one of several groups (clusters). Each group has a chance of being picked based on its weight (how important or likely it is to be chosen).
2. **Assigning the Instance**:
    - When we pick a cluster for a data point, we note which cluster it belongs to. For example, if we pick cluster 2, we say that this data point is in cluster 2.
3. **Generating Data Points**:
    - Now that we know which cluster the data point belongs to, we create its actual value by using a specific formula. This formula uses two things: a center point (mean) for the cluster and how spread out the points are around that center (covariance).

The class relies on the expectation maximization algorithm which has many similarities with k-means, it also initializes the cluster parameter randomly then it repeats the steps until convergence.First assigning instances to cluster and then updating the clusters. It is similar to the gernalization of k-means that also finds the shape, size and orientation as well as their relative weights. It uses soft clustering assignment.

It can also lead to poor convergence so it is advised to run multiple times.

Gaussian model is a generative model that allows to sample new instances form it.

- **Components**: Each Gaussian in the mixture has its own mean and covariance. The number of components (clusters) is a hyperparameter you define.
- **Weights**: Each Gaussian component has an associated weight (probability), indicating how much that component contributes to the overall mixture. These weights must sum to 1.

Hyperparameters:

1. **Number of Components (K)**:
   - This is the number of Gaussian distributions in the mixture. Choosing the right KKK is crucial because too few can lead to underfitting (not capturing the complexity of the data), while too many can cause overfitting (capturing noise).
2. **Covariance Type**:
   - This describes how the shape of the Gaussian distributions is defined. Common types include:
     - **Full**: Each component has its own full covariance matrix, allowing for an ellipsoidal shape. This is the most flexible option.
     - **Tied**: All components share the same covariance matrix.
     - **Diagonal**: Each component has its own diagonal covariance matrix, assuming features are independent.
     - **Spherical**: Each component has the same spherical covariance, meaning all features have the same variance and are equally important.
3. **Mean**:
   - The mean of each Gaussian component, which represents the center of that cluster.
4. **Covariance Matrix**:
   - This matrix determines the spread of the data points around the mean for each Gaussian. The shape and orientation of the covariance matrix affect how the clusters appear in the data.

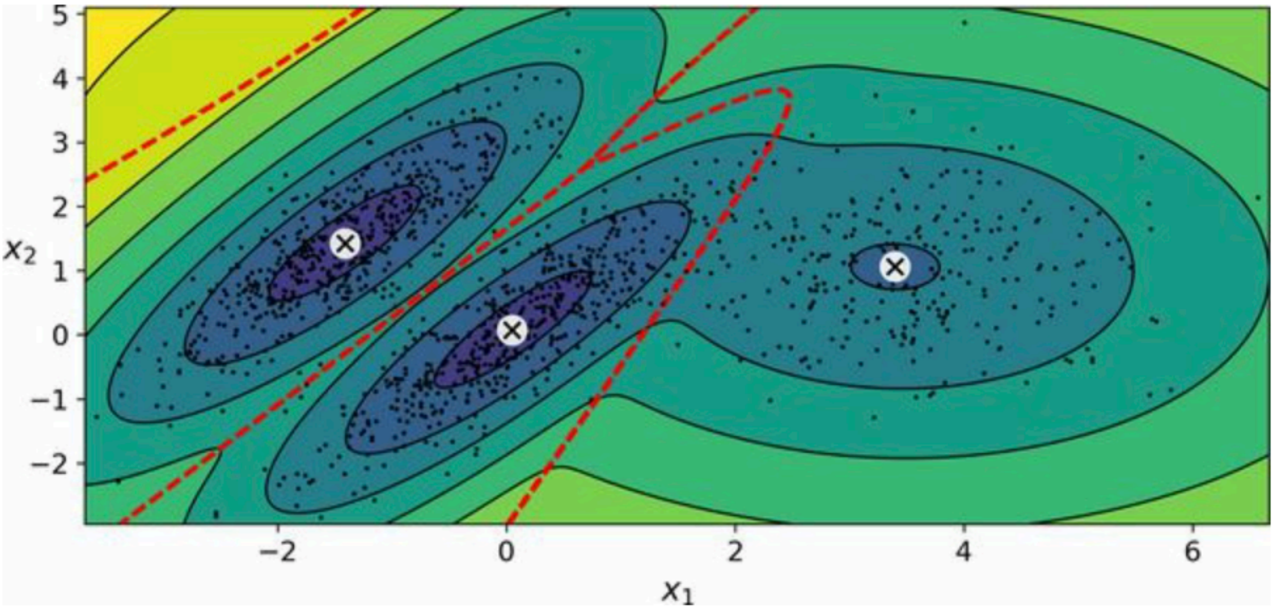

Figure 9-16. Cluster means, decision boundaries, and density contours of a trained Gaussian mixture model

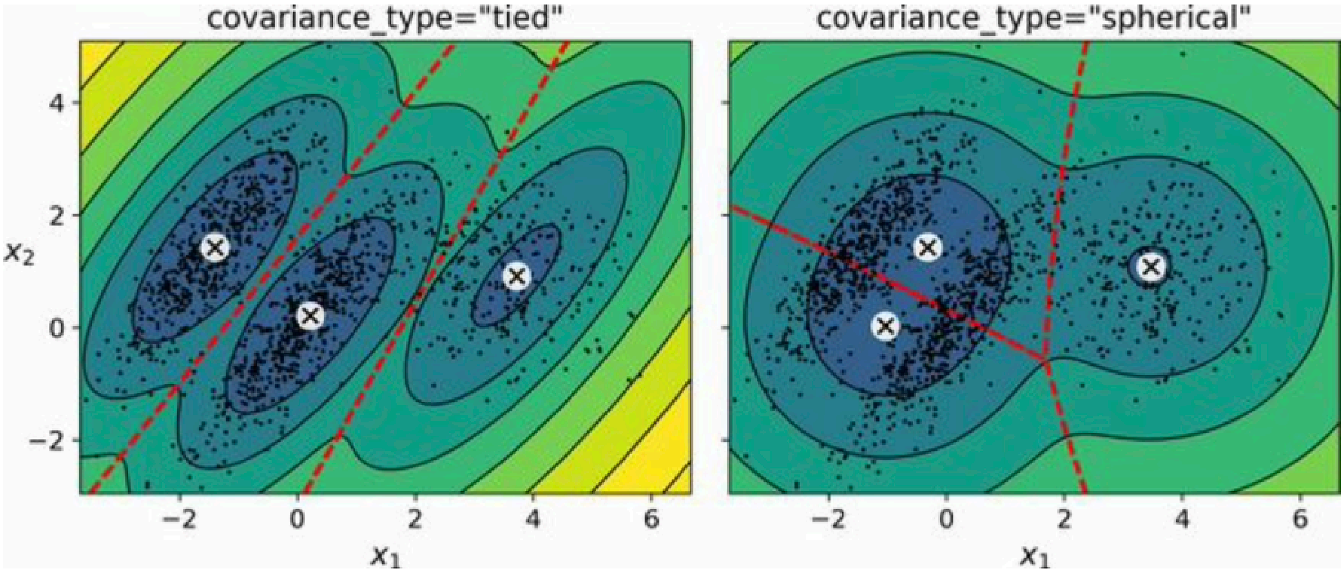

Figure 9-17. Gaussian mixtures for tied clusters (left) and spherical clusters (right)

To locate any anomaly , they are usually located in a low density region that can be considered an anomaly.

To find number of clusters:

- try to find the model that minimizes a theoritical information criteria such as the Bayesian information criterion OR Akaike information criterion.

  $$\text{AIC} = 2k - 2\ln(L)$$

  $$\text{BIC} = \ln(n)k - 2\ln(L)$$

  Bayesian Gaussian Mixture Models (BGMM) provide a more flexible approach to clustering by automatically adjusting the number of active clusters. Instead of manually searching for the optimal number of clusters, you can use the `BayesianGaussianMixture` class, which can assign weights close to zero for unnecessary clusters. You should set the number of clusters (`n_components`) to a value that is greater than the optimal number of clusters, based on your understanding of the problem. This way, the algorithm can eliminate the unnecessary clusters automatically. They don't do well with cluster of different shapes.

Other Algorithm include:

1. Isolation Forest

It's an algorithm used to detect anomalies (or outliers) in a dataset, particularly effective with high-dimensional data.

Process
*1. **Random Forest Creation**:

- An Isolation Forest builds a collection of decision trees (like a random forest). Each tree is created randomly.

2 . **Feature and Threshold Selection**:

- At each decision point (or node) in a tree:
- A feature (or variable) is randomly selected.
- A random threshold value is chosen from the minimum and maximum values of that feature to split the data into two groups.

3.**Isolation Process**:

- This process continues, creating branches until each data point is isolated. The goal is to split the dataset until individual points are separated from the rest.

4 **Anomaly Detection**:

- Anomalies are typically farther away from the majority of data points. Because of this, they usually require fewer splits to isolate than normal points.
- The algorithm measures how many splits it takes to isolate each point: fewer splits mean the point is more likely to be an anomaly.

2. Local Outlier factor

- It compares the density of instances around a given instance to the density around its neighbors. An anomaly is often more isolated than its k-nearest neighbors.

3. One class SVM

- One-Class SVM is primarily used for novelty detection, which means it identifies new or unusual instances compared to what it has seen before.

Process

1. **Mapping to High Dimensions**:
   - Similar to regular SVM (Support Vector Machine), One-Class SVM starts by transforming the data into a high-dimensional space. This helps in separating instances more effectively.
2. **Separation from the Origin**:
   - Instead of separating two classes (like in typical SVM), One-Class SVM tries to find a boundary that surrounds the instances (data points) while keeping them away from the origin (the center point of the space).
   - Essentially, it aims to create a “region” that contains most of the data points.
3. **Identifying Anomalies**:
   - If a new instance falls outside this defined region, it’s considered an anomaly. This means that the model identifies anything that doesn’t fit within the learned boundary as unusual or novel.

- One-Class SVM can struggle with very large datasets. It might be computationally intensive and slower to process when the amount of data is high.