

Deep Autoencoding Gaussian Mixture Models

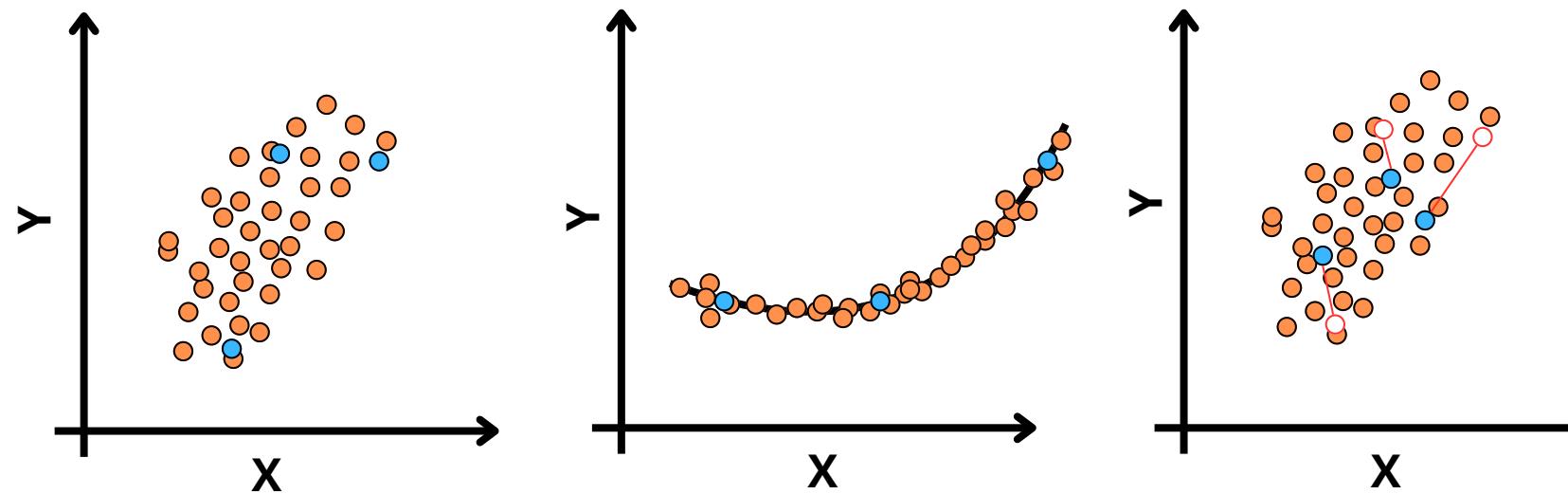
Unsupervised Learning project

Pietro Mihelj

Two types of anomaly behaviour

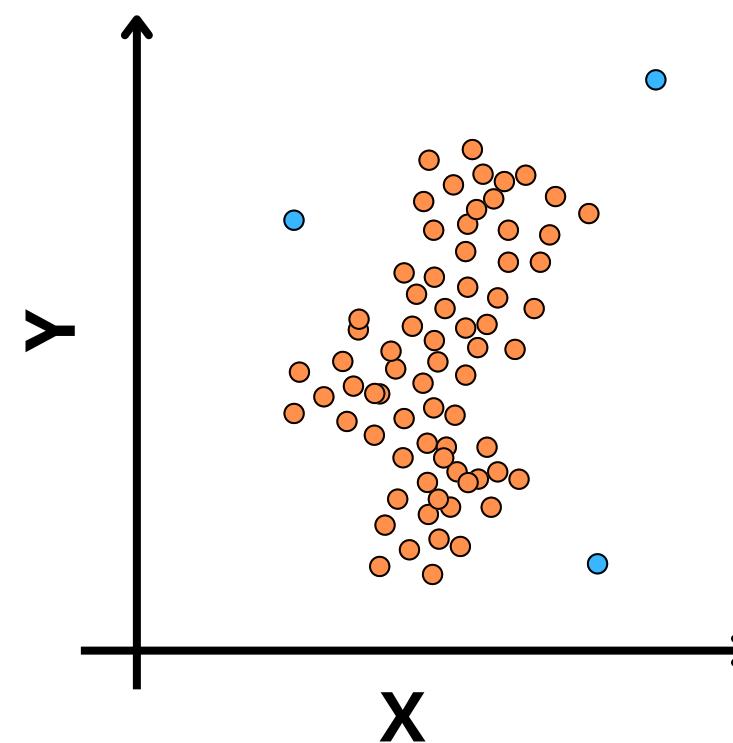
Projection anomalies

These types of anomalies are identified by the reconstruction error generated when the data is projected onto the learned manifold of normal samples.



Density anomalies

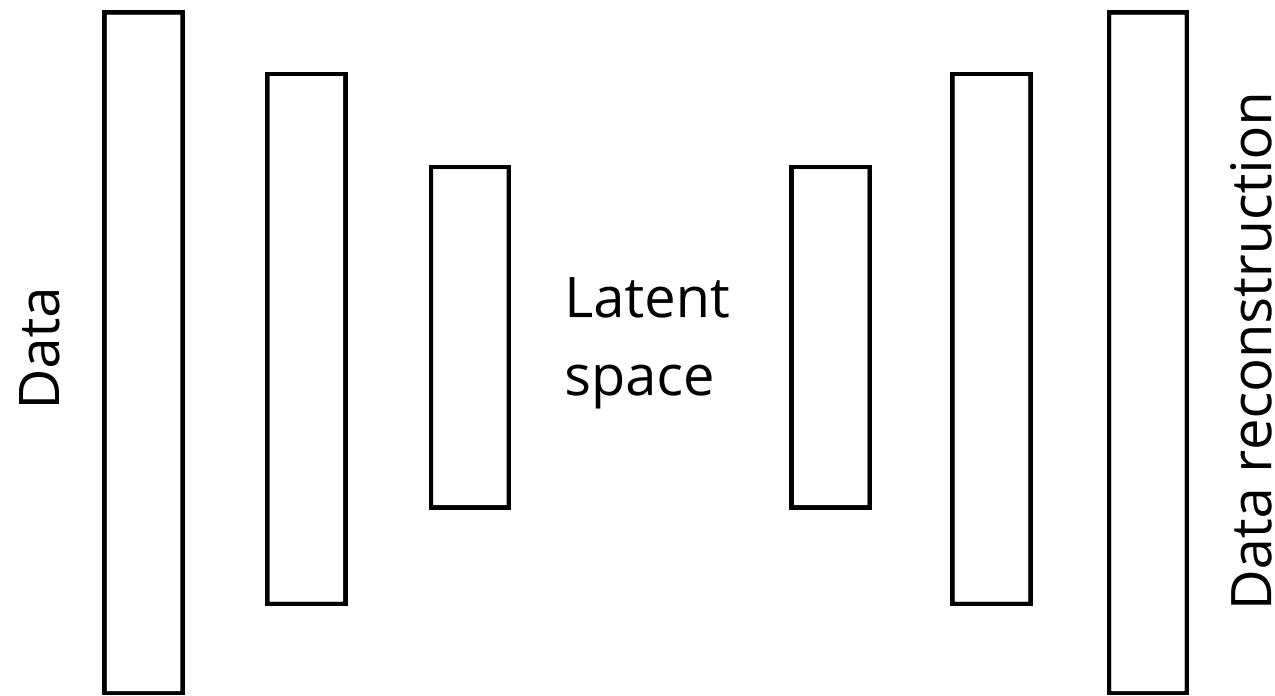
These anomalies represent statistical outliers, detected as data points residing outside the high-density clusters of the normal samples.



Two solutions

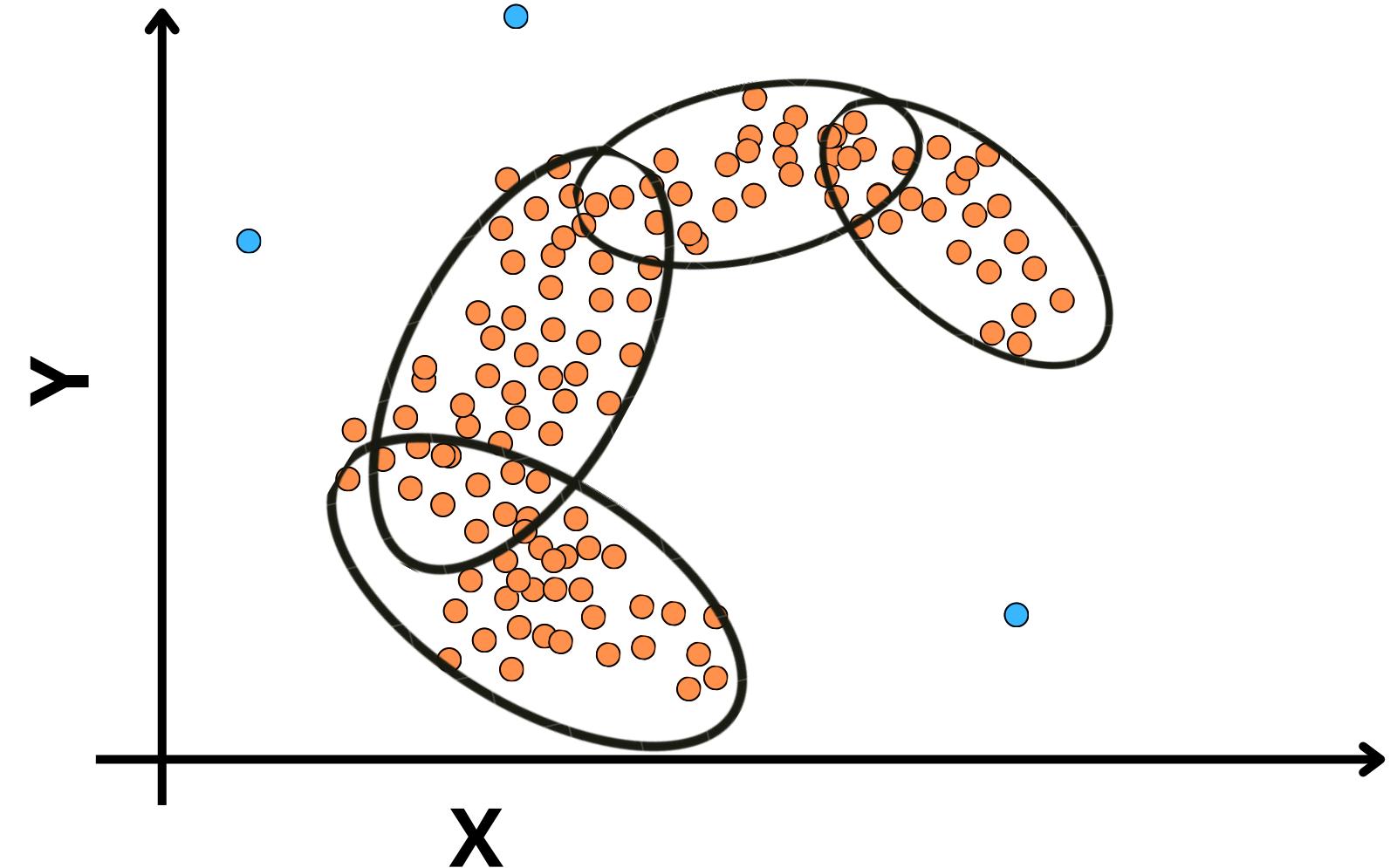
AUTOENCODER

This part of the net focuses on mapping data to the latent space to detect anomalous projections.



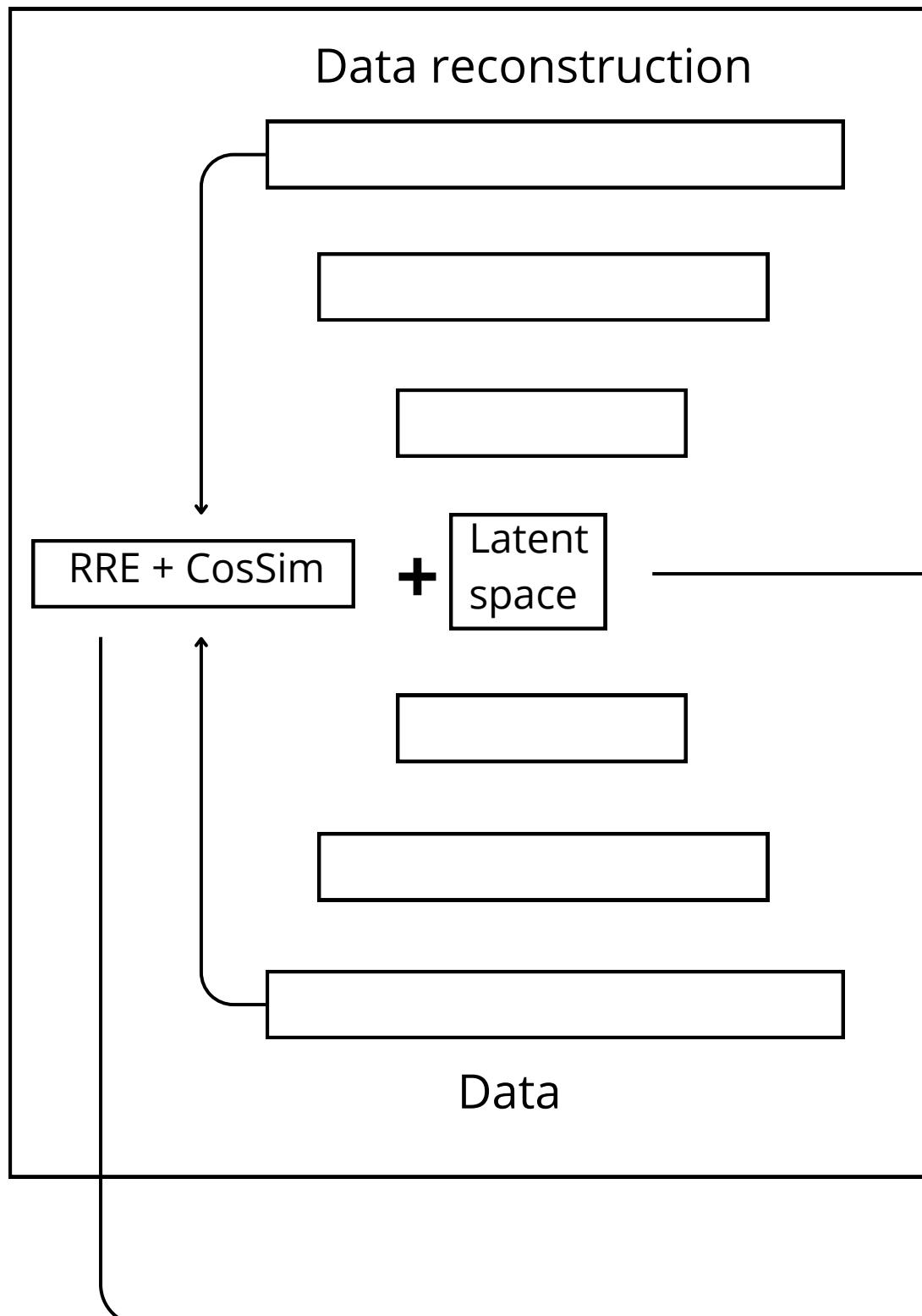
GMM

This part of the net focuses on learning the distribution of the data in the latent space to detect the density anomalies.

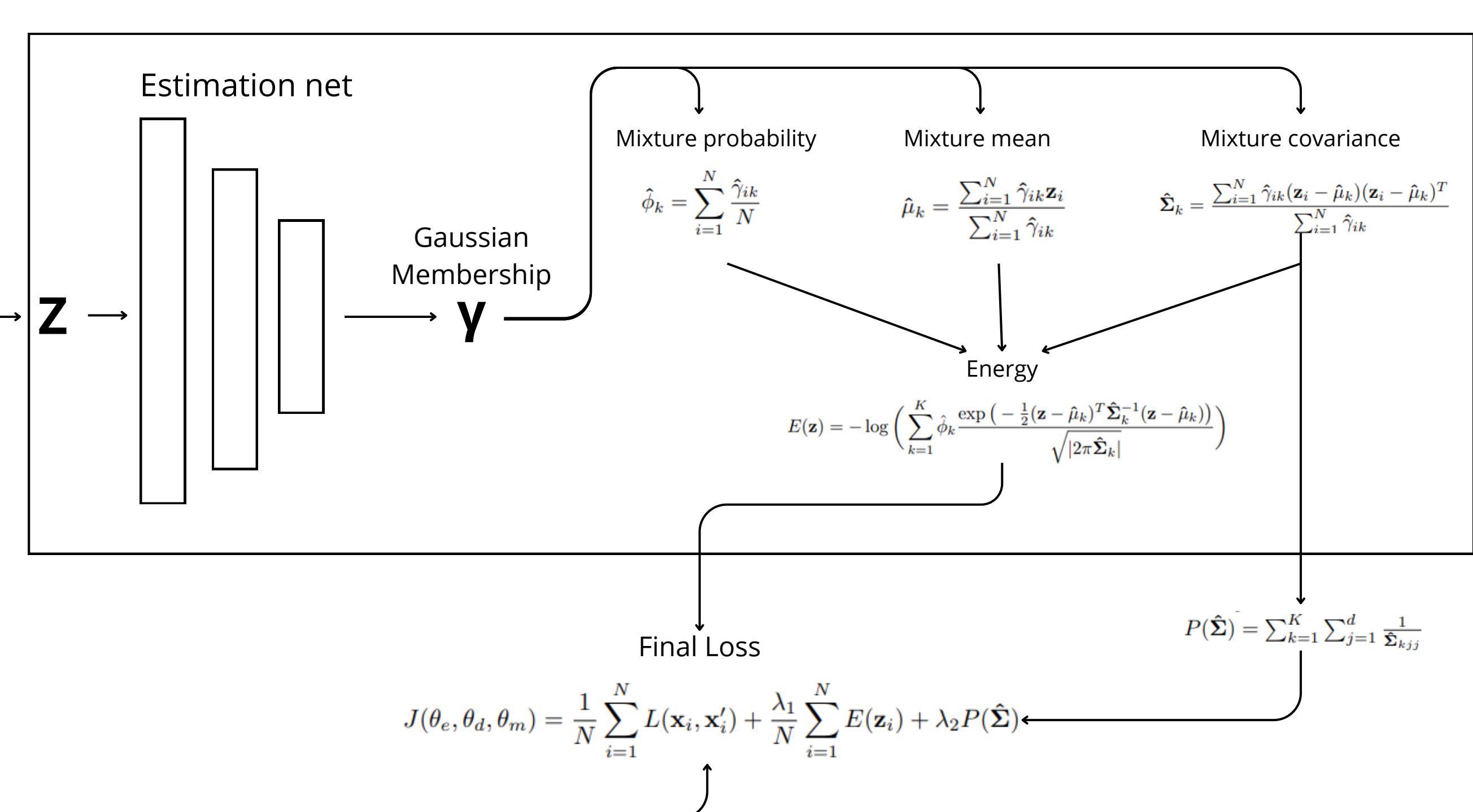


The joint optimization

AE



GMM



Results

Dataset

I used the KDD Cup dataset as a benchmark, which consists of network traffic data categorized into normal activities and various attack types.

Total number of data: 4.898.431

Number of normal data: 972.781

Number of anomaly data: 3.925.650

Confusion matrix comparison

Given the highly imbalanced nature of the dataset, I used a confusion matrix to evaluate the model's performance. This provides a detailed breakdown of false positives and false negatives.

