Hierarchal Bayesian network for Group anomaly detection

In past two weeks, we worked mainly on understand the problem domain related to the area of finding group anomaly in a dataset and testing the proposed model on synthetic dataset.

In the first part we understand and study the Hierarchal Bayesian Model proposed in the paper (1). According to the model, the assumption is, each points in the distribution belongs to a group. Where each group are composed of topic proportion and each points is generated from a topic. So, the idea is to find the likelihood score of all groups, where a group with smaller likelihood is considered as anomalies. Then using the generative process proposed we generate synthetic dataset to compare the performance of the proposed method with baseline algorithm like GMM.

Then to perform this experiment we modified some part of the data generation process to allow us for flexible dimensional dataset. Then for each data generated we compute the anomaly score of the groups **Gm**, based on the inference model given below. To compare against GMM algorithm, since GMM is point wise anomaly detect, we first compute anomaly score for each point then we aggregate the score result of each group based on the membership of the point in the groups.

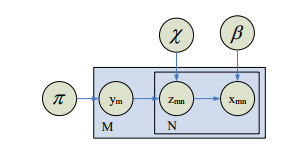


Figure 1 MGMM model



Figure 2 Group likelihood inference

The AUC result comparison shows the result from the group anomaly detection achieves better accuracy compared to point wise anomaly detector, i.e. GMM. But, the result are based on the synthetic dataset with known group membership of points.