CS6890: Fraud Analytics - Assignment 3
Identifying outliers in the data by using Variational Autoencoders

Akshay Santoshi - CS21BTECH11012
Nitya Bhamidipaty - CS21BTECH11041

April 2025

1 Introduction

In this assignment, we applied a Variational Autoencoder (VAE) to a dataset (data.csv) followed by clustering analysis to identify outliers. The dataset consists of 10 features across 1199 samples. The goal was to generate bad points by performing clustering on the latent representations to detect outliers using k-means on non-linear representations.

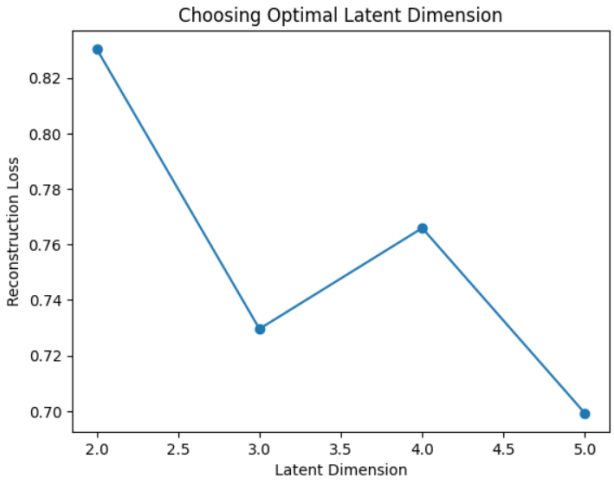
2 Approach

- First, we did data preprocessing, where we normalized the data. This is important for VAE training.
- Next, we defined a VAE model which consists of the encoder and decoder. For encoder, we used three linear layers, leakyRELU activations and batch normalization. We used similar architecture for decoder.
- For latent dimension selection, we tested multiple latent dimensions (2, 3, 4, 5) to determine the optimal one. The choice of latent dimension was taken to be 5 after observing the plot of latent dimension vs reconstruction loss. We used this value to train the VAE.
- Synthetic data was generated from sampling. This was evaluated by comparing feature distributions (KDE plots) and correlation heatmaps to assess similarity.
- Next, we applied K-means clustering on the latent representations obtained from encoder. We used elbow method to determine the optimal value of k (number of clusters).

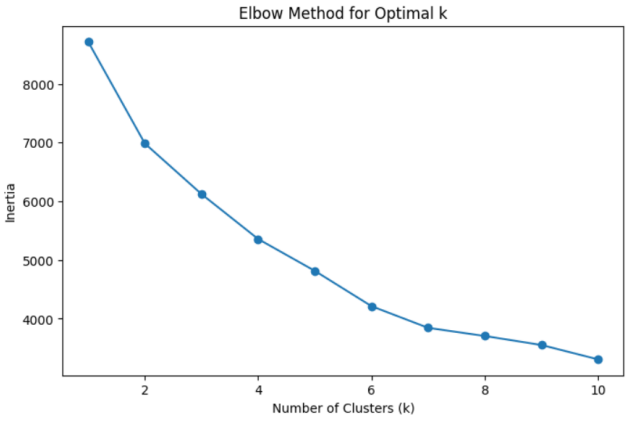
- For outliers detection, we considered small clusters which are clusters with fewer than 50 points as outliers. Additionally, the boundary points of big clusters, which are points beyond 95 percent of distance from the cluster center were identified as outliers.

3 Results

- We chose latent dimension for VAE as 5, after testing for loss in dimensons: [2, 3, 4, 5].



- Using elbow method to find the optimal number of clusters for kmeans, we got k = 6.



- In big clusters, we got the boundary points as shown in the table below:

```
=== Boundary Points from big clusters ===
          cov1      cov2      cov3      cov4      cov5      cov6      cov7  sal_pur_rat  igst_itc_tot_itc_rat  lib_igst_itc_rat
2     0.947603  0.455667  0.001743  0.128610 -0.004054 -0.162069  0.960601    -0.030209              1.535697         -0.054215
5     0.595378 -0.531958  0.679654 -0.126799  0.455487  0.432046  0.988092    -0.029813              0.768742         -0.054167
13    0.825371 -0.248244  0.220238  0.128478  0.383093  0.364443  0.951265    -0.030247              1.512327         -0.054221
18    0.948575  0.464917  0.366662  0.515506  0.064824 -0.192308  0.997203    -0.031555              1.838233         -0.054304
25    0.310202  0.714827  0.999397  0.450586  0.999196  0.951493  0.994530     3.957174             -0.567733         -0.042841
45    0.188614  0.031166  0.832167  0.342503 -0.153480  0.824574  0.152444    -0.013936             -0.296467         -0.052397
53    0.999971  0.801451  0.450078  0.296673 -0.671271  0.221535  0.935767    -0.032791              0.313554         -0.054196
60    0.995746  0.943895 -0.294339 -0.409825 -0.546041  0.567263  0.850460    -0.032687              0.143176         -0.054144
65    0.989749  0.939645  0.973857  0.944348  0.450608 -0.020708  0.958587     0.022532              0.704163         -0.053252
72    0.998746  0.212017 -0.063191  0.101606  0.328125  0.497050  0.994448    -0.032126              1.445437         -0.054300
102   0.999988  0.998002 -0.263647 -0.279467  0.000000  0.490375  0.356098    -0.033177             -1.066401          7.486296
106  -0.133945  0.283608  0.970914 -0.171278  0.000000  0.976835  0.905625    -0.030864             -0.375026         -0.053857
116   0.996294  0.936644  0.942416  0.892207  0.416943  0.004257  0.940807    -0.021392              0.283056         -0.053187
124   0.946262  0.741639 -0.122708 -0.149332 -0.450775  0.507202  0.880061    -0.031580              0.924464         -0.054280
138   0.999163  0.999993  0.743275  0.750927  0.000000 -0.459238  0.880631    -0.030943              1.583726         -0.054258
160   0.999347  0.969439  0.084767  0.151435  0.000000 -0.091304  0.004396    -0.032430             -1.063559          0.109207
161   0.999999  0.973877  0.270032  0.223560  0.000000  0.066139 -0.225197    -0.030177             -0.962634         -0.047378
168   0.907334  0.986871  0.403606 -0.018537  0.000000  0.079338  0.249307    -0.032239             -1.000277         -0.046231
177   0.913044  0.993652  0.495064  0.595524  0.661698  0.707592  0.993330    -0.032280              1.866162         -0.054325
201   0.033564 -0.146953 -0.142238 -0.215256 -0.133407  0.653039  0.547611    -0.029229             -0.849317         -0.052171
202   1.000000  0.903017 -0.665221 -0.674896  0.000000  0.947314  0.591421    -0.032109             -1.066436         33.188277
213   0.968639  0.987501  0.821807  0.740876  0.000000 -0.164718  0.517192    -0.031127             -0.692171         -0.052769
225   0.729880  0.987338  0.502494  0.427280  0.000000 -0.412246  0.998549    -0.032672              2.109923         -0.054356
233   0.998990  0.725474 -0.333254 -0.354830 -0.183836  0.131956  0.509740    -0.032411             -0.429788         -0.053754
248   0.999991  0.900182 -0.793428 -0.777158  0.000000  0.656264  0.093283    -0.033191             -0.852278         -0.052728
249   1.000000  0.856922  0.823258  0.497245  0.000000  0.591521  0.021841    -0.029940             -1.066299          4.959007
251   0.999778  0.938575  0.407715  0.469158 -0.719622 -0.682734  0.989535    -0.033599              1.397712         -0.054371
278   0.809370  0.913733  0.500696  0.563106  0.557942  0.270214  0.999991    -0.034272              2.164212         -0.054435
296   0.999995  0.394505  0.000000  0.000000  0.000000 -0.458461  0.068400    -0.031696             -0.943973         -0.049939
308  -0.166308  0.537869  0.177641  0.320768 -0.275142  0.393301  0.936062     0.011202              1.120954         -0.054245
356   0.999983  1.000000  0.878975  0.878308  0.000000 -0.233162  0.645662    -0.006128             -0.522472         -0.047875
378   1.000000  0.985049 -0.818128 -0.839158  0.000000  0.943767  0.589556    -0.032756             -0.969197         -0.050059
383   0.983083  1.000000 -0.417261 -0.348643  0.000000  0.007628  0.013886    -0.034129             -0.977304         -0.050602
416   1.000000  0.937775  0.518851  0.498893  0.000000 -0.480075  0.078438    -0.031209             -1.064084          0.205998
462  -0.312219  0.993686  0.546496 -0.083316  0.000000  0.317602  0.633745    -0.009160             -0.790955         -0.052700
492   0.990779  0.999023 -0.743153 -0.792628 -0.401923  0.848454  0.894657    -0.033177              0.460462         -0.054247
495   0.937190  0.998131 -0.289857 -0.579922 -0.535282 -0.099784  0.999547    -0.033258              2.126933         -0.054385
544   0.993797  0.626927  0.533724  0.078995 -0.421030  0.968341  0.054833    -0.032082             -1.046099         -0.028653
556   0.936705  0.730814 -0.137128 -0.203968  0.000000 -0.043827  0.588258    -0.032106             -0.951176         -0.049528
558   0.998720  0.937802  0.852196  0.723002 -0.362788  0.554720  0.817692    -0.032296             -0.529022         -0.053589
559   0.999995  0.999789  0.820507  0.819809  0.000000  0.477295 -0.641931    -0.032637             -0.682291         -0.053340
564   0.999991  0.985618  0.521109  0.517412  0.000000 -0.264617  0.331915    -0.032833              0.632495         -0.054257
579   0.999665  0.999687  0.742898  0.736978  0.000000 -0.199314  0.293745    -0.032287             -0.136891         -0.053980
590   0.935332  0.994236  0.554837  0.554199  0.000000 -0.175750  0.487344    -0.032483             -0.991349         -0.048398
591   0.245619  0.999049  0.300700 -0.059870  0.000000  0.793884  0.786821    34.367195              0.391459         -0.054193
603   0.999075  0.998665  0.221312  0.216898 -0.709599  0.906405  0.960929    -0.032578              1.303369         -0.054316
606   0.752158  0.443502  0.000000  0.000000  0.000000  0.710299 -0.330908    -0.033128             -1.066153          1.340107
617   0.779312  0.924756  0.376765  0.024213  0.000000 -0.253367  0.938624    -0.033672              0.533353         -0.054281
634   1.000000  0.973775  0.000000  0.000000  0.000000  0.139212 -0.370801    -0.031962             -1.065460          0.476889
719   0.995573  1.000000 -0.499830 -0.469569  0.000000  0.817349 -0.859529    -0.033682             -1.061556          0.033165
727   0.999814  0.999847 -0.142689 -0.150315 -0.585397  0.786954  0.136990    -0.032751              0.239111         -0.054174
742   1.000000  0.758236  0.520414  0.052706 -0.473722  0.947093  0.709609    -0.032642             -0.647499         -0.053441
758   1.000000  1.000000 -0.119516 -0.119513  0.000000 -0.334804  0.855163    -0.032522             -0.620944         -0.053473
816   0.999953  0.999987 -0.254486 -0.257876  0.000000  0.992984 -0.714136    -0.032657             -0.529335         -0.053678
834   0.999861  0.998006 -0.625201 -0.644975  0.000000 -0.626728  0.143720    -0.032908              0.702145         -0.054270
859   0.988302  0.899342  0.507492  0.467459  0.000000 -0.511736  0.940332    -0.033288              0.163180         -0.054198
866   0.999428  0.822074  0.400275  0.325327  0.000000 -0.248904 -0.230404    -0.031192             -1.025085         -0.039506
874   0.996589  0.999892  0.902333  0.880036  0.000000  0.871109 -0.520048    -0.031821             -1.064662          0.270302
932   0.826183  1.000000  0.149461 -0.105047  0.000000  0.624052 -0.430079    -0.031100             -1.053800         -0.001910
940   1.000000  0.999355 -0.161472 -0.165730  0.000000 -0.123388 -0.077092    -0.035115             -0.988122         -0.052575
996   0.998851  0.959415  0.551929  0.546367  0.605376  0.495742  0.805618    -0.032820              1.435093         -0.054339
1024  0.996287  0.943707  0.637786  0.435098 -0.598050  0.853106  0.445108    -0.030675             -0.441364         -0.053432
1063  1.000000  0.991599  0.844219  0.823941  0.000000 -0.081792 -0.190562    -0.028441             -1.045474         -0.009374
```

Since we considered small clusters to be of size less than 50 points, we didn't get any outliers here as none of the 6 clusters have data points less than 50.