Variational Autoencoder (VAE) for High-Dimensional Anomaly Detection

---

1. Dataset Description

A synthetic high-dimensional dataset was generated with 20 features:

• 5000 normal samples from N(0,1)

• 500 anomalies from N(5,3)

Data was scaled using StandardScaler and reduced to 10 PCA components.

2. VAE Architecture

Encoder:

- Dense(32) → Dense(16)

- Outputs: z_mean, z_log_var, sampled z using reparameterization.

Decoder:

- Dense(16) → Dense(32) → output layer.

Loss:

total_loss = reconstruction_loss(MSE) + beta * KL_divergence

3. Hyperparameter Tuning

Grid search:

• latent_dim: 2, 4

• beta: 0.5, 1.0

• learning_rate: 0.001, 0.0005

Best parameters:

latent_dim = 4

beta = 0.5

learning_rate = 0.001

4. Training & Outputs

Final VAE trained on normal samples.

Saved files:

• vae_encoder.h5

• vae_decoder.h5

5. Anomaly Detection Results

Reconstruction error used as anomaly score.

Test set performance:

• ROC-AUC = 1.000

• Precision = 1.000

• Recall = 1.000

• F1 Score = 1.000

Optimal threshold (Youden): 8.076

6. Latent Space Parameters

Exported z_mean and z_log_var values for 50 test samples.

7. Visualization

• Score distribution showed clear separation between normal & anomaly.

• ROC curve showed perfect discrimination.

8. Conclusion

The VAE accurately modeled normal behavior and detected anomalies with perfect performance on synthetic data. All required project deliverables were completed successfully.