

Contents

1	Introduction	1
1.1	Hypothesis	3
1.2	General Objective	4
1.3	Specific Objectives	4
2	Generative-Inference models: Background	5
2.1	Generative models	5
2.1.1	Generative adversarial networks	5
2.2	Variational autoencoders	6
2.3	Evaluation metrics	7
2.3.1	Inception Score (IS)	8
2.3.2	Fréchet inception distance (FID)	8
3	Generative-Inference models: Foundations	10
4	Generative-Inference models: A matching joint distributions perspective	16
4.1	Matching $q(x, z)$ and $p(x, z)$ adversarially	16
4.2	Minimizing $D_{KL}(q(x, z) p(x, z))$ by decomposing it	18
4.3	Minimizing $D_{KL}(p(x, z) q(x, z))$ by decomposing it	18
4.4	Which method should I use to match $q(x, z)$ and $p(x, z)$?	19
5	Generative-Inference models: Representation learning perspective through mutual information.	21
5.1	Detailed Analysis	25
5.1.1	Generative capabilities for models that bound $\mathcal{I}_q(x, z)$	26
5.2	Related methods	27
5.3	Discussion	28
6	Generative-Inference models: Empirical analysis and proposed models	29
6.1	Practical considerations of generative-inference’s loss functions	29
6.2	Wasserstein Variational Autoencoders and others	30
6.3	Experiments	31
6.3.1	Dataset	31
6.3.2	Evaluation metrics	32
6.3.3	Architecture details	32
6.4	Results	35

6.4.1	Datasets with low complexity	35
6.4.2	Changing the number of latent dimensions	40
6.4.3	Datasets with higher complexity	43
6.4.4	Discussion	45
7	Generative-Inference models: Extending the graphical model to three variables	46
7.1	Variational Deep Embedding (VaDE)	48
7.1.1	Deriving the loss function of VaDE using Jensen’s inequality	48
7.1.2	Loss function of VaDE from a matching joint distributions perspective	49
7.1.3	Loss function of VaDE from a mutual information perspective	50
7.2	Matching priors and conditional for clustering	51
7.2.1	Model definition	51
7.2.2	Loss function and optimization of MPCC from a matching joint distributions perspective	52
7.2.3	Loss function of MPCC from a mutual information perspective	55
7.2.4	Related methods	56
8	Experiments and results for matching priors and conditional for clustering	58
8.1	Experimental setup	58
8.1.1	Datasets	58
8.1.2	Evaluation metrics	58
8.1.3	Empirical details	59
8.1.4	Architecture details	59
8.2	Results	62
8.2.1	Ablation study	62
8.2.2	Comparison between GMM Prior and Normal Prior	63
8.2.3	Generation quality of MPCC	63
8.2.4	Clustering results	64
8.3	Discussion	64
9	An astronomical application with Variational Deep Embedding	67
9.1	Related work	68
9.2	Astronomical datasets	69
9.3	A decoder for variable length and irregular sample time series	70
9.4	Anomaly detection with LC-VaDE	72
9.4.1	Graphical model	72
9.4.2	Loss function of LC-VaDE	73
9.4.3	Anomaly detection score	74
9.5	Experiments	75
9.5.1	Metrics	75
9.5.2	Architecture details	75
9.5.3	Results	77
9.5.4	Discussion	79
10	Conclusions	85
10.1	Future work	86