Conditional Variational Autoencoders with Fuzzy Inference

Yury Gurov[1[0000−0002−7033−9996]] and Danil Khilkov[1[0000−0001−9284−6924]]

NIIAS Institute of Informatization, Automation and Communication in Railway Transport, Russia, Moscow 109029 Nizhegorodskaya str., 27 bldg. 1 info@vniias.ru
www.vniias.ru/

Abstract. We present an approach to constructing Conditional Variational Autoencoders (C-VAE) models with fuzzy inference during classification. This preserves disentangling capabilities of VAE and at the same time performs latent space clusterization. Fuzzy C-VAE model provides useful features for anomaly detection, utilizing partially labeled datasets and controlled generation of new samples.

Keywords: fuzzy logic · deep learning · fuzzy inference · Conditional Variational Autoencoders · fuzzy cvae · neuro-fuzzy

1 Introduction

Hybrid neuro-fuzzy systems has a long time history and is still an active research area [?]. Main feature that attract attemtion to neuro-fuzzy systems is possibility to combine the power of neural network with advantages of fuzzy logic, such a human-like reasoning. At this work we propose an approach to constructing Conditional Variational Autoencoders (C-VAE) [?,?,?] models with fuzzy inference during classification phase. This approach may preserve disentangling capabilities of VAE and at the same time performs latent space clusterization. Such fuzzy C-VAE model provides useful features for anomaly detection, utilization of partially labeled datasets and controlled generation of new samples.

The source code is available at GitHub repository (https://github.com/kenoma/pytorch-fuzzy).

2 Related work

In [?] attempt to apply fuzzy logic to the latent space of VAE was made with fuzzy c-mean clustering. Main drawback of this approach is that it requires prior knowledge about problem domain number of clusters and their iterpretations.

In [?] conditional VAE was modified in a way to process partially observed datasets. Authors proposed method that augments the conditional VAEs with a prior distribution for the missing covariates and estimates their posterior using amortised variational inference. At first sigh this approach has nothing to do with fuzzy logic, but it provides insigh into the problem of latent space clustering.

3 Methods

3.1 Variational Autoencoders

Variational inference is used to approximate a posterior distribution of a directed graphical model whose latent variables and parameters are intractable. The Variational Auto-Encoder (VAE) combines this approach with an autoencoder framework to learn the prior distribution of a latent space, $p_\theta(z)$, with parameters θ. The idea is that the prior distribution can then be sampled to produce a latent code, z, which is passed as input to the decoder to produce a sample output, $\tilde{x}$. VAEs consists of two NN for the probabilistic encoding and decoding process (see Figure 1a). As the true underlying distribution of the posterior is intractable and complex, a simple parametric surrogate distribution, $q_\Phi(z|x)$ (such as a Gaussian), with parameters Φ, is assumed to approximate the distribution and is optimized for best fit. The encoder network implicitly models the surrogate distribution, by mapping the distribution parameters, Φ, during the training process. The resulting model, $q_\Phi(z|x)$, is referred to as the recognition model. The optimization process of the recognition model revolves around minimizing the Kullback-Leibler (KL) divergence between the posterior and surrogate distributions. Once the latent prior distribution is learned, z can be sampled via the reparameterization trick. The (probabilistic) decoder network performs a mapping of the latent code to a structured sample output for each sample, thus producing a distribution of outputs, $p_\theta(x|z)$.

3.2 Conditional VAE

The C-VAE expands upon the framework of the VAE, by combining variational inference with a conditional directed graphical model. In the case of C-VAE, the objective is to learn a prior distribution of the latent space that is conditioned on an input variable y such that $p_\theta(z|y)$ (see Fig. **??**b). The conditioning of the distributions results in a prior that is modulated, by the input variable, creating a method to control modality of the output.

3.3 Fuzzy C-VAE

We propose C-VAE architecture where additional conditions are applied only to $/mu$ component in order to reorganize the latent space structure (see Fig**??**c). Reorganization achieved by using fuzzy term functions, where each term associated with sole condition i.e. label. Multidimentional Gaussian function is used to represent the fuzzy term function:

$$\nu(z, A_i) = e^{||[A_i . \tilde{z}]_{1 \cdots m}||^2},$$

where m is a latent space dimension size, i denotes term number, $\tilde{z} = [z_1, z_2, \cdots, z_m, 1]$ and A_i is transformation matrix in form

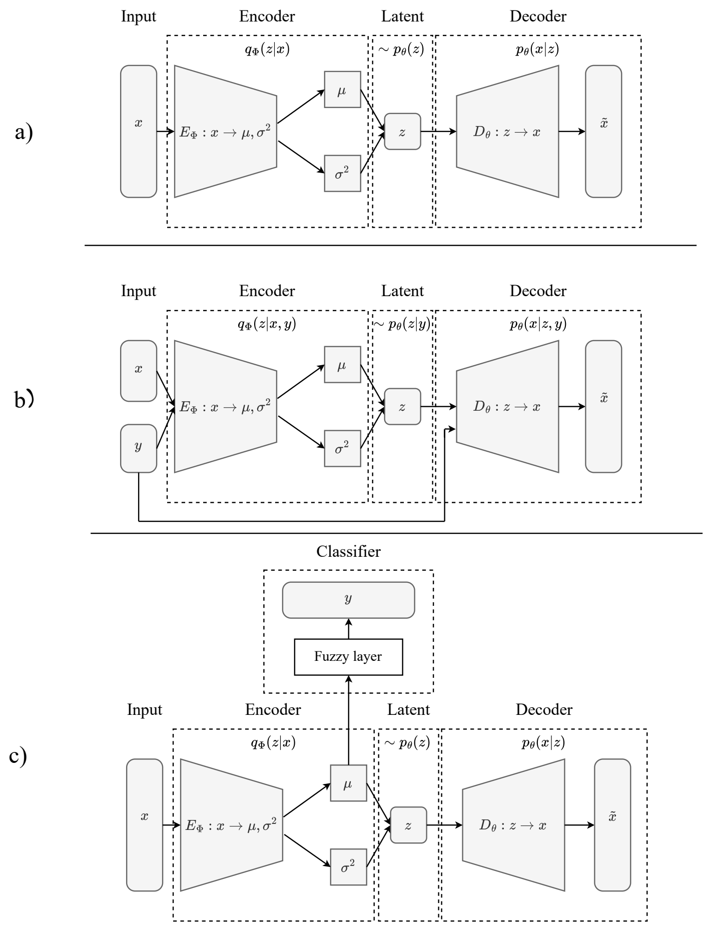

Fig. 1. Overview of the *a*) VAE *b*) CVAE by [?] and *c*) proposed fuzzy C-VAE model.

$$A_{(m+1)\times(m+1)} = \begin{bmatrix} s_1 & a_{12} & \cdots & a_{1m} & c_1 \\ a_{21} & s_2 & \cdots & a_{2m} & c_2 \\ \vdots & \vdots & \ddots & \vdots & c_3 \\ a_{m1} & a_{m2} & \cdots & s_m & c_m \\ 0 & 0 & \cdots & 0 & 1 \end{bmatrix}$$

with $c_{1\cdots m}$ centroid position, $s_{1\cdots m}$ scaling factor and $a_{1\cdots m,1\cdots m}$ as alignment coefficients. Such representation of multidimentional Gaussian term function easyly can be adopted for use in any modern machine learning framework. Set of $\nu(z, A_i)$ we call a fuzzy layer. Intuition behind fuzzy layer is that during training procedure every input vector z will be forced to group closer near centroid of correspoinding term function. Disentangling features of VAE combined with clustering possibilities of fuzzy layer provides a way to learn suprvised latent space which can be useful for anomaly detection and other tasks we discuss further.

3.4 Learning Fuzzy C-VAE

To train fuzzy C-VAE we use the same loss function as in standard VAE with addition of fuzzy layer loss:

$$Loss = MSE(\tilde{x}, x) + KL(\mu, \log \sigma^2) + FZ(\tilde{y}, y),$$

where $MSE(\tilde{x}, x)$ is reconstruction loss, $KL(\mu, \log \sigma^2)$ is the KL-divergence (for more details see [**?**]) and $FZ(\tilde{y}, y)$ represents the mean squared error between the output and target conditional verctor.

Main drawback of fuzzy layer is that in hight dimensional cases it's hard to find good initial values for centroids and scaling factors mainly due to vanishing gradients. Is such a case it is possible to pass to fuzzy layer subsection of vector $/mu$ leaving remained part to be trained by VAE without any conditional restrictions.

4 Experiments

In this paper we would like to demonstrate ideas of fuzzy C-VAE on playground MNIST dataset. To make demostration fancy we provide additional label to samples of MNIST dataset. This label separates numbers with closed round loops in outline $0, 6, 8, 9$ from numbers without it $1, 2, 3, 4, 5, 7$. To make reasonable reconstruction loss we set size of the latent vector equal to 12 but only 2 first values of this vector are passed to fuzzy layer. The fuzzy CVAE model is trained using the Adam optimizer [**?**].

4.1 Latent space clusterization

On Fig

4.2 Learning on partially labeled dataset

4.3 Anomaly detection

5 Discussion

In this paper, we introduced a novel fuzzy inference layer to improve the performance of conditional VAEs. We achieve this by making trainable multidimentional representation of fuzzy term. The method that we proposed is applicable to a variety of conditional VAE models. The efficacy of our proposed method was demonstrated on MNIST dataset.

References

1. Bölat, K., Kumbasar, T.: Interpreting variational autoencoders with fuzzy logic: A step towards interpretable deep learning based fuzzy classifiers. In: 2020 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE). pp. 1–7 (July 2020). https://doi.org/10.1109/FUZZ48607.2020.9177631
2. de Campos Souza, P.V.: Fuzzy neural networks and neuro-fuzzy networks: A review the main techniques and applications used in the literature. Applied Soft Computing **92**, 106275 (2020). https://doi.org/https://doi.org/10.1016/j.asoc.2020.106275, https://www.sciencedirect.com/science/article/pii/S1568494620302155
3. Kingma, D.P., Ba, J.: Adam: A method for stochastic optimization (2017). https://doi.org/10.48550/arXiv.1412.6980
4. Kingma, D.P., Welling, M.: An introduction to variational autoencoders. Foundations and Trends® in Machine Learning **12**(4), 307–392 (2019). https://doi.org/10.1561/2200000056, http://dx.doi.org/10.1561/2200000056
5. Kingma, D.P., Welling, M.: Auto-encoding variational bayes (2022). https://doi.org/https://doi.org/10.48550/arXiv.1312.6114
6. Ramchandran, S., Tikhonov, G., Lönnroth, O., Tiikkainen, P., Lähdesmäki, H.: Learning conditional variational autoencoders with missing covariates. Pattern Recognition **147**, 110113 (2024). https://doi.org/https://doi.org/10.1016/j.patcog.2023.110113, https://www.sciencedirect.com/science/article/pii/S0031320323008105
7. Sohn, K., Yan, X., Lee, H.: Learning structured output representation using deep conditional generative models. In: Proceedings of the 28th International Conference on Neural Information Processing Systems - Volume 2. p. 3483–3491. NIPS'15, MIT Press, Cambridge, MA, USA (2015). https://doi.org/10.5555/2969442.2969628

fig2-a-eps-converted-to.pdf

fig2-b-eps-converted-to.pdf

Fig. 2. Latent space granulation for vanila VAE (top) and fuzzy C-VAE (bottom)