Neural Discrete Representation Learning [1]
https://arxiv.org/abs/1711.00937

Matyáš Skalický (skalimat@fit.cvut.cz)

10. 3. 2019

1 Introduction

The paper [1] which I'm going to present on Tuesday describes the idea of using a variational autoencoder (VAE), but instead of having a continuous latent space, it has a discretisation step that creates discrete latent space embeddings. This makes a lot of sense, since in many real world scenarios, it simply does not make sense to interpolate between multiple categories. If we have a dog and a chair, it makes no sense to interpolate between the two categories.

Very often, you might see a shortcut VQ-VAE, which stands for Vector-Quantized Variational Autoencoder. In terms of the Variational Autoencoders, it basically means turning vector from continuous space into a vector of discrete values (finding a closest match from a codebook).

2 Prerequisites

TLDR: If you know ML basics, there is nothing that you have to study to prepare for my talk.

This paper is based on the model of Variational Autoencoder (VAE), which is built on the basis of the Autoencoder model. My presentation assumes that you are familiar with some of the machine learning vocabulary:

Feed Forward Neural Networks are the model, that fuels the current revolution in machine learning. If you are not familiar with them, see for example the chapter 10 of Daniel Shiffman's book - The Nature of Code.

Autoencoders are basically neural networks compromised of an encoder and decoder, that try to squish the input of a big dimension into a smaller "bottleneck" layer and then to reconstruct it back as the output layer. Real usecases include image compression, removing noise from images or even colouring black and white pictures. If you have not seen an autoencoder yet, you can have a look at one of the great online articles. But I suggest trying to build one yourself, it's pretty easy and fun in Keras.

Variational Autoencoders (VAE) combine the neural network architecture of autoencoders with a probabilistic model. This combination produces a powerful generative model, which is able to randomly sample from the latent space, or generate variations on an input image, from a continuous latent space. There are, again, many great tutorials on VAEs such as this, or this.

On Tuesday, Mato Choma is going to have a talk about the Tutorial on Variational Autoencoders [2] paper which covers the topic of variational autoencoders. Also, there is a talk from last week class by Ondra Biza, on the paper Variational Inference: A Review for Statisticians [3] which explains the math behind the variational inference.

3 Implementation

What I cannot create, I do not understand.
— Richard Feynman

If you are interested in a real working example of discrete representation learning, you can see one of the following implementations [1]:

PyTorch: https://github.com/nakosung/VQ-VAE
– the shortest implementation that I have stumbled on.

PyTorch: https://github.com/zalandoresearch/pytorch-vq-vae/blob/master/vq-vae.ipynb
– Jupyter notebook with some great description.

Chainer: https://github.com/dhgrs/chainer-VQ-VAE

TensorFlow: https://github.com/hiwonjoon/tf-vqvae

4 Further reading after talk

There is also a followup paper called Theory and Experiments on Vector Quantized Autoencoders [4] which investigates an alternate training technique for VQ-VAE, inspired by its connection to the Expectation Maximization algorithm.

[1] The original code by DeepMind is surprise, surprise, not published.

References

[1] Aaron van den Oord, Oriol Vinyals, et al. Neural discrete representation learning. In Advances in Neural Information Processing Systems, pages 6306–6315, 2017.

[2] Carl Doersch. Tutorial on variational autoencoders. arXiv preprint arXiv:1606.05908, 2016.

[3] David M Blei, Alp Kucukelbir, and Jon D McAuliffe. Variational inference: A review for statisticians. Journal of the American Statistical Association, 112(518):859–877, 2017.

[4] Aurko Roy, Ashish Vaswani, Arvind Neelakantan, and Niki Parmar. Theory and experiments on vector quantized autoencoders. arXiv preprint arXiv:1805.11063, 2018.