FROM DETERMINISITC TO PROBABILISTIC AUTOENCODERS

A PREPRINT

Feng Geng
shouldsee@qq.com *

Hua Zong Ban

June 15, 2022

ABSTRACT

Since the famous paper on denoising autoencoder (DAE), AutoEncoder mainly referes to usage of neural networks to extract meaningful hidden representations, with its performance measured with resconstruction error or generative capabilities, enchanceable with denoising or entropic regularisiation. Here we show modern autoencoder objectives connect to traditional approximation objectives like PCA through a simple framework, which unifys the determinisitc and probabilistic views.

1 Introduction

Ideas of autoencoders trace back to resricted boltzmann machines. Later research showed denoising objectives prevent overfitting when latent space is bigger than the original space in feedforward neural nets. The addition of noise in latent space is further formalised under the framework of variational inference. The core idea is to approximate the intractable posterior latent distribution with a neural network, so that we do not have to invert the decoder using Bayes theorem.

$$z = f(x) \tag{1}$$

$$L(\theta, \phi) = E_{z \sim q_\theta(z|x)}[p(x|z)] + KL(q_\theta(z|x)||p_\phi(z|x)) \tag{2}$$

On the other hands, non-probabilistic algorithms exist for producing linear and nonlinear embeddings under an error minimisation framework, such as principal components analyses (PCA), locally linear embeddings, Isomap, tSNE, UMAP. Many of the nonlinear embedding algorithms, however, do not include a decoding process, and is thus a one-way process constrained to preseve information. For these algorithms that encode high-dimensional information into low-dimensional vectors, it is not immediately obvious how to put them under the same hood as the probabilistic AEs.

1.1 Determinisitc autoencoders

We introduce deterministic autoencoders to keep the idea simple under an optimisation framework. For the determinisitc autoencoders, we ask the model to reconstruct the input under a certain distance functino as close as possible. decomposed into the sample-wise objective, it reads

$$\begin{aligned} L(m) &= \sum_b l(m, x_b) \\ &= \sum_b \max_{z_b} \left(-||g_m(z_b) - x_b||^2\right) \\ l(m, z_b, x_b) &= \left(-||g_m(z_b) - x_b||^2\right) \end{aligned}$$

*http://www.catsmile.info

note we denote autoencoding process as finding the z that optimise the reconstruction loss. To summarise the autoencoding process and model fitting in a single expression, consider the sum of all possible per-sample decoding errors.

$$
\begin{aligned}
L(m, Z) &= \sum_x l(m, z_b, x_b) \\
&\leq \sum_x \max_{z_b} l(m, z_b, x_b) \\
&\leq \max_m \sum_x \max_{z_b} l(m, z_b, x_b)
\end{aligned}
$$

where the max over m means fitting the model. Note however, this step-wise maximum is not equvalent to finding the gloabl optimum, since as model parameters moves, the local sample-wise maximum may change correspondingly.

$$
\max_{m,Z} L(m, Z)
$$

1.2 probabilistic Autoencoders

See Section 1.1. Fusce mauris. Vestibulum luctus nibh at lectus. Sed bibendum, nulla a faucibus semper, leo velit ultricies tellus, ac venenatis arcu wisi vel nisl. Vestibulum diam. Aliquam pellentesque, augue quis sagittis posuere, turpis lacus congue quam, in hendrerit risus eros eget felis. Maecenas eget erat in sapien mattis porttitor. Vestibulum porttitor. Nulla facilisi. Sed a turpis eu lacus commodo facilisis. Morbi fringilla, wisi in dignissim interdum, justo lectus sagittis dui, et vehicula libero dui cursus dui. Mauris tempor ligula sed lacus. Duis cursus enim ut augue. Cras ac magna. Cras nulla. Nulla egestas. Curabitur a leo. Quisque egestas wisi eget nunc. Nam feugiat lacus vel est. Curabitur consectetuer.

$$
\xi_{ij}(t) = P(x_t = i, x_{t+1} = j | y, v, w; \theta) = \frac{\alpha_i(t) a_{ij}^{w_t} \beta_j(t+1) b_j^{v_{t+1}}(y_{t+1})}{\sum_{i=1}^{N} \sum_{j=1}^{N} \alpha_i(t) a_{ij}^{w_t} \beta_j(t+1) b_j^{v_{t+1}}(y_{t+1})} \tag{3}
$$

1.2.1 Headings: third level

Suspendisse vel felis. Ut lorem lorem, interdum eu, tincidunt sit amet, laoreet vitae, arcu. Aenean faucibus pede eu ante. Praesent enim elit, rutrum at, molestie non, nonummy vel, nisl. Ut lectus eros, malesuada sit amet, fermentum eu, sodales cursus, magna. Donec eu purus. Quisque vehicula, urna sed ultricies auctor, pede lorem egestas dui, et convallis elit erat sed nulla. Donec luctus. Curabitur et nunc. Aliquam dolor odio, commodo pretium, ultricies non, pharetra in, velit. Integer arcu est, nonummy in, fermentum faucibus, egestas vel, odio.

Paragraph Sed commodo posuere pede. Mauris ut est. Ut quis purus. Sed ac odio. Sed vehicula hendrerit sem. Duis non odio. Morbi ut dui. Sed accumsan risus eget odio. In hac habitasse platea dictumst. Pellentesque non elit. Fusce sed justo eu urna porta tincidunt. Mauris felis odio, sollicitudin sed, volutpat a, ornare ac, erat. Morbi quis dolor. Donec pellentesque, erat ac sagittis semper, nunc dui lobortis purus, quis congue purus metus ultricies tellus. Proin et quam. Class aptent taciti sociosqu ad litora torquent per conubia nostra, per inceptos hymenaeos. Praesent sapien turpis, fermentum vel, eleifend faucibus, vehicula eu, lacus.

2 Examples of citations, figures, tables, references

2.1 Citations

Citations use `natbib`. The documentation may be found at

`http://mirrors.ctan.org/macros/latex/contrib/natbib/natnotes.pdf`

Here is an example usage of the two main commands (`citet` and `citep`): Some people thought a thing [Kour and Saabne, 2014a, Hadash et al., 2018] but other people thought something else [Kour and Saabne, 2014b]. Many people have speculated that if we knew exactly why Kour and Saabne [2014b] thought this…

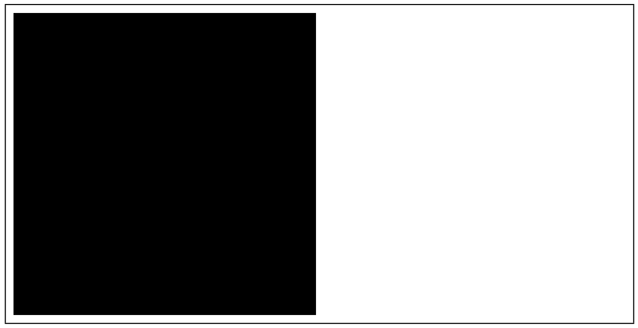

Figure 1: Sample figure caption.

Table 1: Sample table title

Part		
Name	Description	Size (μm)
Dendrite	Input terminal	$\sim$100
Axon	Output terminal	$\sim$10
Soma	Cell body	up to 10^6

2.2 Figures

Suspendisse vitae elit. Aliquam arcu neque, ornare in, ullamcorper quis, commodo eu, libero. Fusce sagittis erat at erat tristique mollis. Maecenas sapien libero, molestie et, lobortis in, sodales eget, dui. Morbi ultrices rutrum lorem. Nam elementum ullamcorper leo. Morbi dui. Aliquam sagittis. Nunc placerat. Pellentesque tristique sodales est. Maecenas imperdiet lacinia velit. Cras non urna. Morbi eros pede, suscipit ac, varius vel, egestas non, eros. Praesent malesuada, diam id pretium elementum, eros sem dictum tortor, vel consectetuer odio sem sed wisi.

See Figure 1. Here is how you add footnotes. [2] Sed feugiat. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Ut pellentesque augue sed urna. Vestibulum diam eros, fringilla et, consectetuer eu, nonummy id, sapien. Nullam at lectus. In sagittis ultrices mauris. Curabitur malesuada erat sit amet massa. Fusce blandit. Aliquam erat volutpat. Aliquam euismod. Aenean vel lectus. Nunc imperdiet justo nec dolor.

2.3 Tables

See awesome Table 1.

The documentation for `booktabs` (‘Publication quality tables in LaTeX’) is available from:

`https://www.ctan.org/pkg/booktabs`

2.4 Lists

- Lorem ipsum dolor sit amet
- consectetur adipiscing elit.
- Aliquam dignissim blandit est, in dictum tortor gravida eget. In ac rutrum magna.

References

George Kour and Raid Saabne. Real-time segmentation of on-line handwritten arabic script. In *Frontiers in Handwriting Recognition (ICFHR), 2014 14th International Conference on*, pages 417–422. IEEE, 2014a.

Guy Hadash, Einat Kermany, Boaz Carmeli, Ofer Lavi, George Kour, and Alon Jacovi. Estimate and replace: A novel approach to integrating deep neural networks with existing applications. *arXiv preprint arXiv:1804.09028*, 2018.

[2]Sample of the first footnote.

George Kour and Raid Saabne. Fast classification of handwritten on-line arabic characters. In *Soft Computing and Pattern Recognition (SoCPaR), 2014 6th International Conference of*, pages 312–318. IEEE, 2014b. doi:10.1109/SOCPAR.2014.7008025.