Neural Networks
Lecture Notes
Autoencoders

Marcel van Gerven

1 Learning Goals

After studying the lecture material you should:

1. know what is meant by an autoencoder
2. be able to explain in words the different variants of existing autoencoders

2 Notes

2.1 Autoencoders

Autoencoders, also known as autoassociative networks or diabolo networks [Sch98] are standard MLPs that aim to reconstruct their input. They aim to find those $\mathbf{W}$ and $\mathbf{V}$ such that

$$\min_{\mathbf{W},\mathbf{V}} \frac{1}{2N} \sum ||\mathbf{x}^{(n)} - \mathbf{V}\mathbf{W}\mathbf{x}^{(n)}||^2 \,.$$

The rows of $\mathbf{W}$ are known as the *receptive fields* and the columns of $\mathbf{V}$ are known as the *projective fields*.

If we use as many hidden nodes as inputs (or more) then we are in the overcomplete regime. A trivial solution would then be to just copy the input. To prevent this trivial solution, people have used various approaches. One approach is to use fewer hidden nodes, such that we are in the undercomplete regime. People have also used sparse, denoising and contrastive autoencoders [BCV12]. The variational autoencoder is a recent variant which is a fully probabilistic model. We can also generate samples from this model by sampling from the hidden variables (latent space) $\mathbf{z}$ [KW13].

If we ignore the nonlinear functions $\mathbf{f}$ and $\mathbf{g}$ then we have a linear autoencoder for which the objective is:

$$\min_{\mathbf{W},\mathbf{V}} \frac{1}{2N} \sum ||\mathbf{x}^{(n)} - \mathbf{V}\mathbf{W}\mathbf{x}^{(n)}||^2 \,.$$

This model is analogous to performing a principal component analysis of the data.

3 Reading material

For background on autoencoders consult e.g. [VLBM08, RVM+11, Doe16].

References

[BCV12] Yoshua Bengio, Aaron Courville, and Pascal Vincent. Representation learning: A review and new perspectives. *arXiv Prepr. arXiv . . .* , (1993):1–34, 2012.

[Doe16] Carl Doersch. Tutorial on Variational Autoencoders. *arXiv*, pages 1–23, 2016.

[KW13] Diederik P Kingma and Max Welling. Stochastic Gradient VB and the Variational Auto-Encoder. pages 1–14, 2013.

[RVM$^+$11] Salah Rifai, Pascal Vincent, Xavier Muller, Xavier Glorot, and Yoshua Bengio. Contractive Auto-Encoders : Explicit Invariance During Feature Extraction. *ICML*, 85(1):833–840, 2011.

[Sch98] H Schwenk. The Diabolo Classifier. *Neural Comput.*, 10(8):2175–2200, 1998.

[VLBM08] P Vincent, H Larochelle, Y Bengio, and P.-A. Manzagol. Extracting and composing robust features with denoising autoencoders. In *Proc. 25th Int. Conf. Mach. Learn.*, volume 307, pages 1096–1103, Helsinki, Finland, 2008.