An Infomax approach to Autoencoder

Vincenzo Crescimanna, Bruce Graham

Research Programme in Contextual Learning in Humans and Machines,
Division of Computing Science and Mathematics, University of Stirling

UNIVERSITY of STIRLING

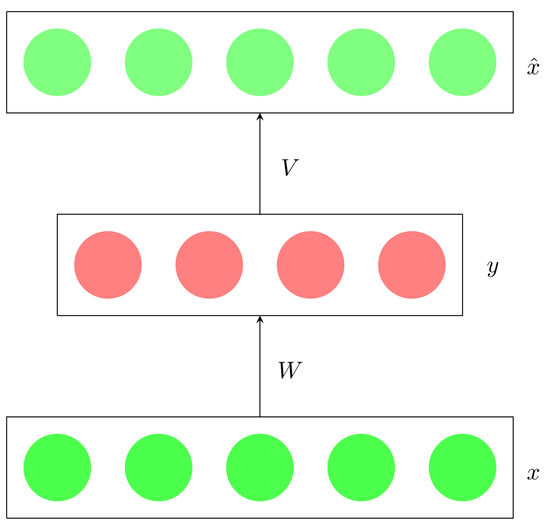
Objectives

The objective of a neural networks are:

- **Extract *good* features**, that are useful as input to a predictor
- **Learn a semantic map**: understand the relevant features associated to each task
- **Integrate the features**: fuse in a *clever* way the features extracted by each process

Introduction

A **good feature (representation)** is one that makes it easier to extract useful information when building classifier or other predictors.
The **autoencoder** (AE) model, is the only Machine Learning approach that learns the *direct encoding* (**filters**).



Autoencoder

Notation:

- x: input data
- $y = f(Wx)$: Extracted features
- W: filter matrix
- $\hat{x} = g(Vy)$: reconstructed input

Issue: Difficulty in establishing a clear objective or target function.
The greedy option, minimize the reconstruction loss:

$$\|x - \hat{x}\|$$

is similar to PCA, is not robust to noisy input and tends to learn the identity map.
The alternative: Denoising AE (DAE)
Idea: Train the neural network with noisy input and compute the reconstruction loss with respect to the original one.
Loss function:

$$\|x - \tilde{x}\|$$

where $\tilde{x} = g(f(Wh))$ with $h \sim \mathcal{N}(x, \sigma I)$

Infomax AE

Following Barlow's principle we propose as objective function for the neural network the **Infomax** approach:

$$\max_{\theta} I_\theta(x, y) \qquad (1)$$

The computation of (1) is intractable, but thanks to some approximations, (1) can be reformulated as the following minimization problem:

$$\min \|x - \hat{x}\| - \log |J_z f(z)| - H(z) \qquad (2)$$

with $z = Wx$ and H denotes the entropy function.

Results

- **Edge oriented filters**: The learnt filters, in the case of DAE and the proposed Infomax Autoencoder (IMAE), are edge oriented. Instead for the classic AE, it is impossible to recognize a shape.
- **Manifold learning** In the unsupervised accuracy test, IMAE over-performs DAE. This means that IMAE discovers the lower-dimensional manifold.

Important Result

Natural Patches

Autoencoder

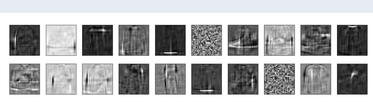
Denoising Autoencoder

IMAE

F-MNIST (500)

Autoencoder

Denoising Autoencoder



IMAE

Accuracy F-MNIST			
hidden units	AE (%)	DAE (%)	IMAE(%)
400 (u)	48.28	47.43	**48.43**
500 (u)	48.25	47.82	**50.05**
400 (s)	77.43	**82.7**	82.29

Related Works

The proposed model belongs in the class of **Regularized Autoencoder**. Like the Contractive Autoencoder (CAE) [1], the loss (2) aims to reduce the norm of the Jacobian J_z. Differently from CAE, eq. (2) maximizes, also, the entropy of the hidden layer.

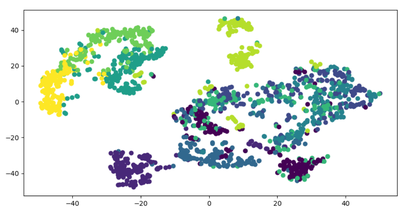
Why IMAE?

IMAE have some properties that DAE does not have:

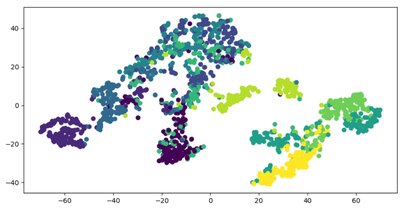
- Learn **robust representation**
- Discover **lower-dimensional manifolds** (see pic. below)
- **Independence** between features

Datasets

Natural Patches: Natural images are redundant, in this case the behaviour of the filters should be similar to the behaviour of the lower cells in visual cortex (V1 cells)
Fashion MNIST: Dataset of Zalando's article images where each image is associated to one of the 10 classes. It is suitable for classification task, and then to test the accuracy of the model, both in unsupervised and supervised way.



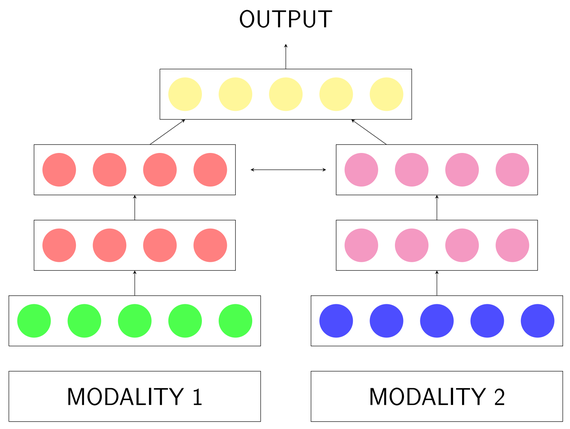
tSNE on IMAE



tSNE on DAE

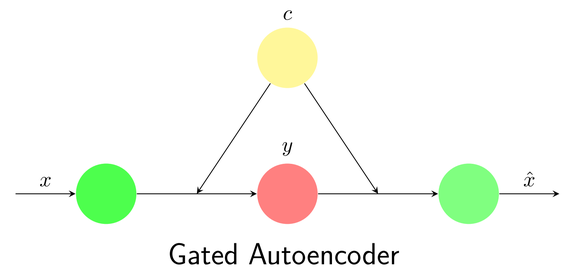
Future work

For complex tasks, a hierarchical neural network is not sufficient. It is necessary a neural network that extract different features for each local process.



Multimodal neural network

A feature, of a certain modality, is *r*elevant if the information contained is useful for the main task and this information is not contained in the other features, extracted by the other local processors.
Following [2] a possible model to obtain the relevant features is to use a **gated network**.



Gated Autoencoder

Where the context information *modulates* the weights of the parallel local process.
It is necessary to generalize (1) with an objective function that takes in consideration the context and have the property described above. A natural extension, proposed in [3], is the **Coherent Infomax**

$$I(x, y, c) = I(x, y) - I(x, y|c)$$

References

[1] S. Rifai, P. Vincent, X. Muller, X. Glorot, and Y. Bengio.
Contractive auto-encoders: Explicit invariance during feature extraction.
In *Proceedings 28th Int Conf on ML*, pages 833–840. Omnipress, 2011.

[2] A. Droniou, S. Ivaldi, and O. Sigaud.
Deep unsupervised network for multimodal perception, representation and classification.
Robotics and Autonomous Systems, 71:83–98, 2015.

[3] J. W Kay and WA Phillips.
Coherent infomax as a computational goal for neural systems.
Bulletin of mathematical biology, 73(2):344–372, 2011.