

COMPUTATIONAL INTELLIGENCE

AUTOENCODERS
for feature extraction



Adrian Horzyk

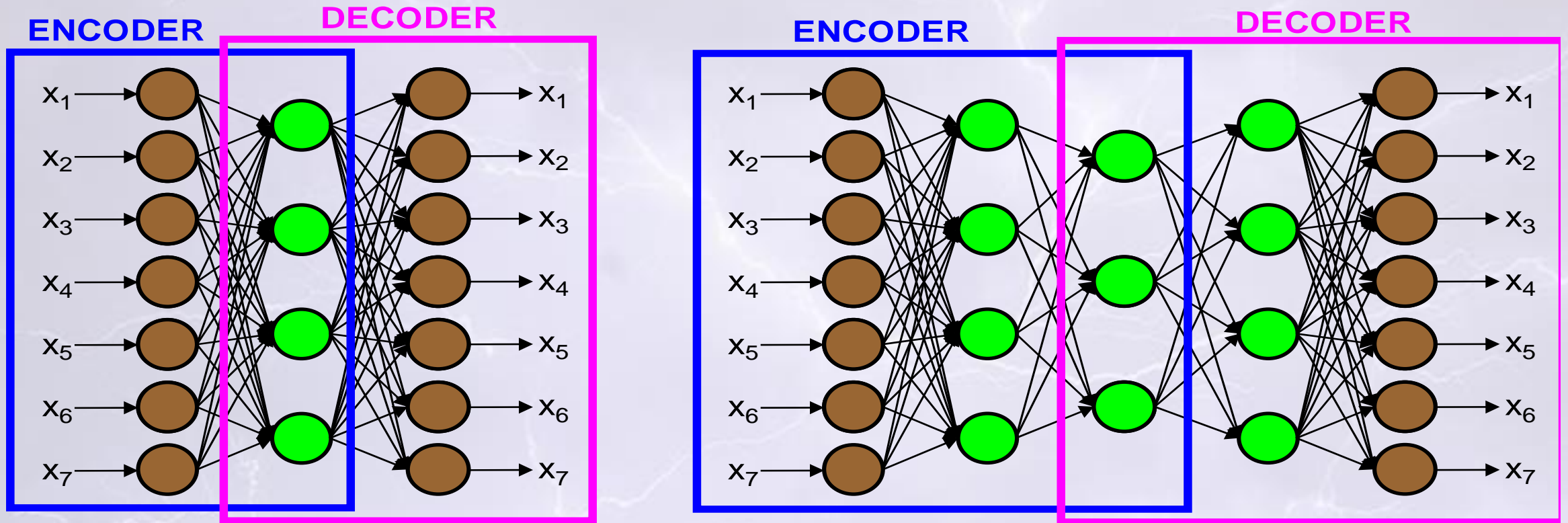


Autoencoders

Autoencoder is a kind of artificial neural networks which is trained to represent a set of training data in an unsupervised manner using a reduced dimensionality and gets the same output data as input ones.

The reduced dimensionality is used to find out frequent combinations which constitute complex data features which can be used in various classifiers.

Autoencoders consist of encoders and decoders:





Types of Autoencoders



We can distinguish a few types of autoencoders:

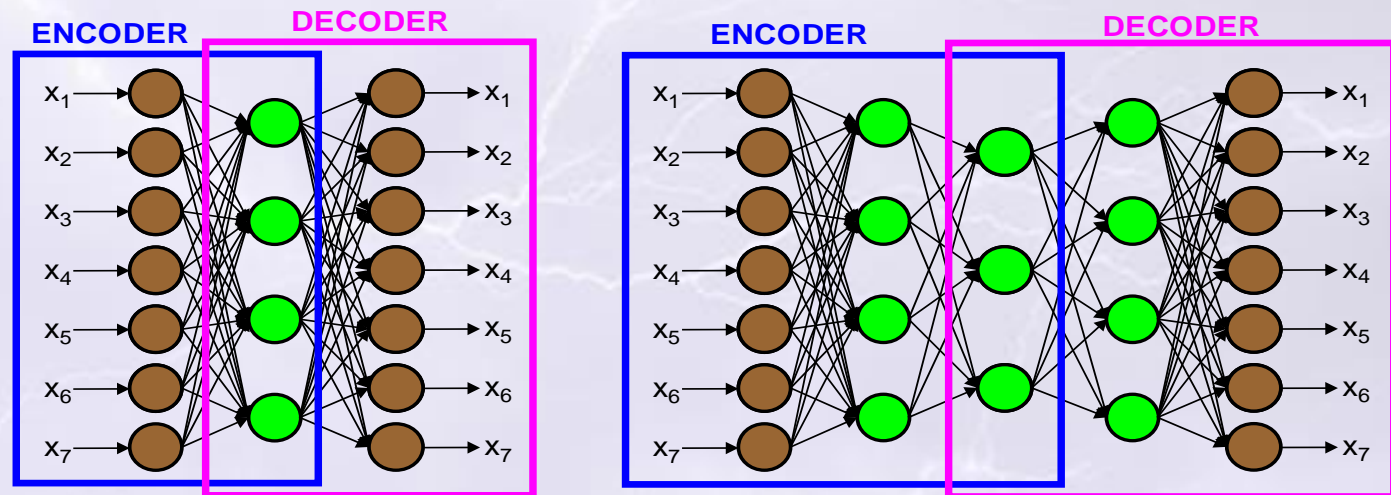
- ✓ **Undercomplete Autoencoders** – are defined to represent data in undercomplete way, i.e. the outputs do not reproduce inputs precisely in order to allow for generalization, feature extraction, data distribution, and correction of outliers. Training of such autoencoders aims to minimize the **loss function** defining the differences between outputs and inputs. When the autoencoders are linear, they work similarly to PCA (Principal Components Analysis), so they can replace such kind of preprocessing algorithms (PCA or ICA).
- ✓ **Autoencoders with Regularization** – use the complexity of the modeled distribution of the data to select an adequate dimension and capacity of encoders and decoders. They use a loss function to be resistant to noise and missing data, and learn correct data distribution. These autoencoders can be non-linear and overcomplete as well.
- ✓ **Sparse Autoencoders** – are autoencoders which are used for other computational tasks, e.g. for classification, where we need to represent frequent features more than find a perfect identity function. In this approach, representation of rare features is penalized. This leads to a sparse representation of inputs and useful feature extraction as a preparation phase for classification.
- ✓ **Anomaly Detection Autoencoders** – are autoencoders which are used to detect rare features that stand for various anomalies in data and can identify outliers.
- ✓ **Denoising Autoencoders (DAE)** – try to find a function which returns correct output for noised, corrupted or incomplete inputs. They have to recover the original undistorted inputs on their outputs.

Training of Autoencoders



Autoencoders are trained in an unsupervised way using the algorithm typically used for supervised learning, e.g. backpropagation. This is because we use the outputs which are the same as the inputs.

- ✓ Assume that we have a set of unlabeled training examples $\{x_1, x_2, x_3, \dots\}$, where $x_i \in \mathbb{R}^n$.
- ✓ An autoencoder uses outputs defined as $y_i = x_i$ where y_i is an expected output value.
- ✓ Autoencoders can learn to extract features similarly as Convolutional Neural Networks (CNN) do.
- ✓ The training capabilities of autoencoders are associated with the number of encoding and decoding layers. When autoencoders have more than single encoding and decoding layers, we call them **deep autoencoders**. Deep autoencoders usually have a better compression ratio than flat autoencoders.
- ✓ Deep autoencoders can be constructed from flat autoencoders trained subsequently and separately.
- ✓ Autoencoders are usually trained using the backpropagation algorithm, however, we can also use other algorithms, e.g. the recirculation algorithm.



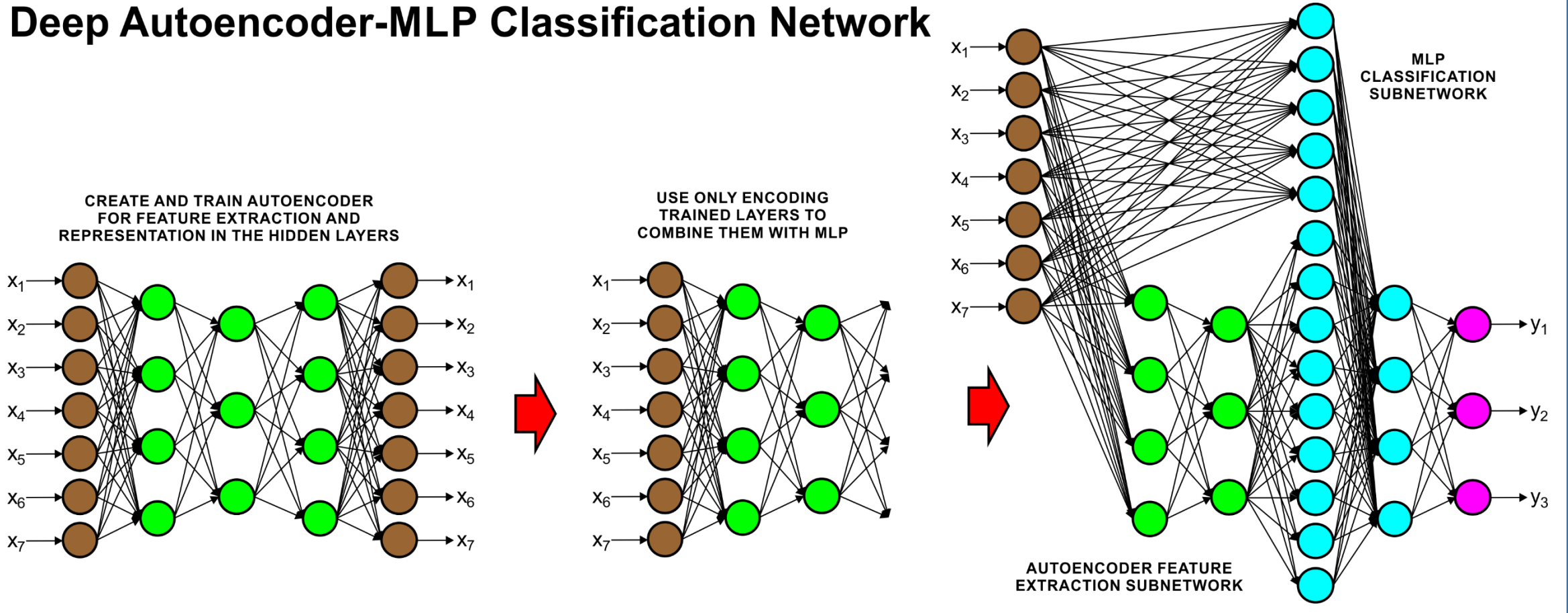


Combining Autoencoders with MLPs



Sparse Autoencoders are often trained to be combined with other types of artificial neural networks, e.g. MLPs. This is because they can preprocess raw input data and extract useful features for other networks:

Deep Autoencoder-MLP Classification Network




One of our goals during laboratory classes will be to implement such a combination of an autoencoder and MLP.



BIBLIOGRAPHY AND REFERENCES



1. Ian Goodfellow, Yoshua Bengio, Aaron Courville, Deep Learning, MIT Press, 2016, ISBN 978-1-59327-741-3 or PWN 2018.
2. Stanford University Tutorial of Unsupervised Learning used to Autoencoders:
<http://ufldl.stanford.edu/tutorial/unsupervised/Autoencoders/>



ACADEMIC WEBSITE - ADRIAN HORZYK, PhD, DSc.
AGH University of Science and Technology in Cracow, Poland
Faculty of Electrical Engineering, Automatics, Computer Science and Biomedical Engineering
Department of Biocybernetics and Biomedical Engineering, Field of Biocybernetics

Dossier Research Publications Courses Graduates Consultations Contact

LECTURES
(will be renewed and expanded during the semester)

- Introduction to Artificial and Computational Intelligence
- Artificial Neural Networks, Multilayer Perceptron MLP, and Backpropagation BP
- Radial Basis Function Networks RBFN
- Unsupervised Training and Self Organizing Maps SOM
- Recurrent Neural Networks**
- Introduction of Final Projects and Description of Requirements
- Associative Neural Graphs and Associative Structures
- Deep Associative Semantic Neural Graphs DASNG
- Associative Pulsing Neural Networks
- Deep Learning Strategies and Convolutional Neural Networks
- Support Vector Machines SVM
- Fuzzy Logic and Neuro-Fuzzy Systems
- Motivated and Reinforcement Learning
- Linguistic, Semantic Memories, and Cognitive Neural Systems
- Psychological Aspects of Intelligence, Human Needs, and Personality
- Writing Journal Papers

COMPUTATIONAL INTELLIGENCE

This course includes 28 lectures, 14 laboratory classes, and 14 project classes.

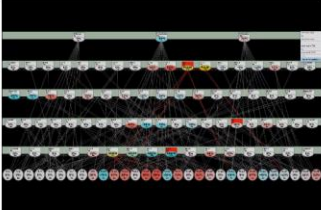

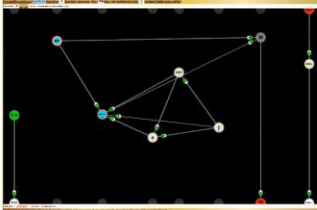
What is this course about?

This course is intended to give students a broad overview and deep knowledge about popular solutions and efficient neural network models as well as to learn how to construct and train intelligent learning systems in order to use them in everyday life and work. During the course we will deal with the popular and most efficient models and methods of neural networks, fuzzy systems and other learning systems that enable us to find specific highly generalizing models solving difficult tasks. We will also tackle with various CI and AI problems and work with various data and try to model their structures in such a way to optimize operations on them throughout making data available without necessity to search for them. This is a unique feature of associative structures and systems. These models and methods will allow us to form and represent knowledge in a modern and very efficient way which will enable us to mine it and automatically draw conclusions. You will be also able to understand solutions associated with various tasks of motivated learning and cognitive intelligence.

Lectures will be supplemented by laboratory and project classes during which you will train and adapt the solution learned during the lectures on various data. Your hard work and practice will enable you not only to obtain expert knowledge and skills but also to develop your own intelligent learning system implementing a few of the most popular and efficient CI methods.

Expected results of taking a part in this course:

- **Broad knowledge** of neural networks, associative and fuzzy systems as well as other intelligent learning systems.
- **Novel experience** and **broaden skills** in construction, adaptation and training of neural networks and fuzzy systems.
- **Ability to construct** intelligent learning systems of various kinds, especially deep learning solutions.
- **Good and modern practices** in modelling, construction, learning and generalization.
- **Own intelligent learning system** to use in your life or work.
- **Satisfaction** of enrollment to this course.



<http://home.agh.edu.pl/~horzyk/lectures/ahdydci.php>