Lect.1: Artificial Intelligence and Deep Learning: The technological revolution and ethical challenges

Summary: We have been observing the spectacular progress of AI and machine learning for over a dozen years. It has been driven by improved neural network learning from large databases, the availability of these data from the internet and social networks and the use of fast Graphic Processor Units (GPUs) that perform calculations in parallel. Because such systems can autonomously acquire new knowledge from the observed environment, and even exceed human cognitive and perceptual abilities, they can be considered as systems endowed with artificial intelligence.

New methods and programs developed in this field led to breakthroughs in such fields as image recognition, speech recognition, computer vision, robotics, intelligent interfaces and a number of others that are revolutionizing today's homo sapiens environment. The talks reviews of the most important techniques to which we owe this technological breakthrough and a brief presentation of training with supervision and without supervision. The most frequently used models and architectures of neural networks, including their most useful applications, will be discussed.

The most useful results of intelligent systems are the complex decisions they are able to make. However, these decisions are made without any transparency. Therefore, the primary challenge for both designers and users of these systems is the assignment of responsibility. These systems, moreover, do not have an inherent or acquired code of ethical behaviour or awareness of legal norms. Further challenges inherent in autonomous systems relate to protection of privacy, user's safety or non-discrimination. These are the challenges of our day, which stand not only against the engineers and computer scientists creating these systems, but with entire societies and nations, which should use them for their broadly understood social good.

Lect. 2: Data Representation and Its Understanding in Deep Learning: Sparse Coding, Additive Features, Perceptrons and Constrained Autoencoders

Summary: Convoluted mappings needed for discriminative data representation in deep neural networks make their mappings less than transparent because of terms cancellations and complexity of representation. However, learning with meaningful constraints and especially with sparse coding allows for extraction of more meaningful and sparse discriminative features. These features can be understood as parts of original sets of objects and are generated through sparse basis vectors. Further, sparse basis functions (or receptive fields or filters) prove more useful when they can be decomposed, and then superimposed and reconstructed with as low a reconstruction error as possible.

Techniques discussed that meet the transparency criteria are (1) Nonnegative Matrix Factorization that reduces the number of basis vectors and allows for extraction of latent features that are additive and hence interpretable for humans. (2) A classic EBP architecture can also be trained under the constraints of non-negativity and sparseness. The resulting classifiers allow for identification of parts of the objects encoded as receptive fields developed by weights of hidden neurons. The results are illustrated with MNIST handwritten digits classifiers and Reuters-21578 text categorization. (3) A constrained learning of sparse non-negative weights in autoencoders also allows for discovery of additive latent factors. Our experiments with MNIST, ORL face and NORB object datasets compare the autoencoder accuracy for various training conditions. They indicate enhanced interpretability and insights through identification of parts of complex input objects traded-off for a small reduction of recognition accuracy or classification error.