Università degli Studi di Palermo

HR EXCELLENCE IN RESEARCH

# Advanced Deep Learning and Transformers

Teacher: Giansalvo Cirrincione

**Abstract**

Deep learning is the new frontier of artificial intelligence. It is based on very performing neural networks, thanks to the last software and hardware developments. In these last years, the transformer model has become the best tool. It is mainly used for advanced applications in natural language processing, like machine translation, named entity recognition, quiz answering and so on. Google is using it to enhance its search engine results. OpenAI has used transformers to create its famous GPT-2 and GPT-3 models, e.g., the famous ChatGPT, a chatbot launched in November 2022. It is also becoming more and more important in other fields of application, like computer vision (ViT), video processing, automatic speech recognition, forecasting (TFT), medical imaging, biological sequence analysis, and so on.

Transformers are a type of artificial neural network architecture which is based on self-attention for solving the problem of transduction or transformation of input sequences into output sequences which take context into account.

The course begins with a quick description of the basic ideas of neural networks and deep learning in such a way that no prerequisites are required. A brief description of backpropagation training with computational graphs is also given.

A quick description of convolutional (CNN) and recurrent (RNN) neural networks follows. CNN's are best suited for images, RNN's for sequences.

The basic ideas of the transformer (encoder and decoder) are illustrated and explained in detail, by means of the case study of machine translation. It follows the study of BERT and its variants. The most important transformer modifications are then briefly described (Longformer, Reformer, Performer, Big Bird, Talking Heads, and so on). Applications in several fields are then highlighted. In particular, the GPT family (GPT-1, GPT-2, GPT-3 and ChatGPT, GPT-4) is described in detail. The last part of the course is based on the interpretability and explainability of the transformers and presents the new ideas developed in the neural team of the DET department of the Polytechnic of Turin.

The course will be held:

Tu, 9th May 10-13
Th, 11th May 10-13
Tu, 16th May 10-13
We, 17th May 10-13
Th, 18th May 10-13
Tu, 23rd May 10-13
We, 24th May 10-13
Th, 25th May 10-13
We, 14th June 10-13

Università degli Studi di Palermo

HR EXCELLENCE IN RESEARCH

Th, 15th June 10-13
Fr, 16th June 10-13

## Program

1. Introduction to deep learning
    a. Basic ideas
    b. The artificial neuron
    c. The Multilayer Perceptron

2. Backpropagation with computational graphs
3. Deep learning techniques
    a. Training with adaptive gradient (Adam, Adagrad, RMSProp, and so on)
    b. Momentum
    c. Regularization (L1, L2, dropout)
    d. Hyperparameter setting.
    e. Normalization (batch and layer)
    f. Activation functions

4. Convolutional Neural Networks
    a. Basic ideas
    b. Architectures

5. Recurrent Neural Networks
    a. Basic Ideas
    b. Architectures
    c. Data fusion
    d. Attention

6. The transformer architecture: part 1
    a. The self-attention mechanism
    b. The encoder

7. The transformer architecture: part 2
    a. The decoder

8. BERT
    a. Basic ideas
    b. Variants
    c. Applications

9. Transformer modifications
    a. Modified multi-head attention.
        i. Longformer
        ii. Reformer

iii. Performer
iv. Big bird

b. Self-attention improvement (Talking heads)

10. Application-specific transformers
    a. Computer vision
    b. Forecasting

**Instructions for Participation**

- **Registration form:**
- https://forms.gle/2cM7U9YUgTV1LpHB7

- **Way to attend:**

In presence - Room T104 building 9 engineering department, UNIPA. Max 40 seats

On-Line – using the following Microsoft Teams channel:

Microsoft Teams meeting

**Join on your computer, mobile app or room device**

Click here to join the meeting

Meeting ID: 332 870 239 193

Passcode: ztacXU

Download Teams | Join on the web

**Or call in (audio only)**

+39 02 3056 2266,,805773162# Italy, Milano

Phone Conference ID: 805 773 162#

Find a local number | Reset PIN


Ricevi questa email in quanto sei stato invitato ad una attività/evento/meeting con l'Università di Palermo

Learn More | Meeting options | Legal