

DEEP LEARNING FOR SPEECH & LANGUAGE

Winter Seminar UPC TelecomBCN, 24 - 31 January 2017

Day 1 Lecture 1

Welcome

Instructors



Antonio
Bonafonte



J. Adrián Rodríguez
Fonollosa



Marta R.
Costa-jussà



Javier
Hernando



Santiago
Pascual



Elisa
Sayrol



Xavier
Giró

Organizers



Image Processing Group
Signal Theory and Communications Department



UNIVERSITAT POLITÈCNICA
DE CATALUNYA
BARCELONATECH

+ info: [TelecomBCN.DeepLearning.Barcelona](https://www.telecombcn.com/deeplearning-barcelona)

[\[course site\]](#)

Instructors

Marta Ruiz Costa-jussà

Ramón y Cajal Researcher at Universitat Politècnica de Catalunya (UPC)



Web: <https://www.costa-jussa.com>



Instructors

Antonio Bonafonte



Telecommunication Engineer and PhD UPC

Working on TTS since 1998

Consultor Bell Labs, NJ, 1999

Cofounder of VERBIO (spoken language technology), 1999

Lecturing (now): audiovisual bachelor: speech and audio processing, and multimedia coding.

Instructors

Jose Adrián Rodríguez Fonollosa

Professor



- M.S. and Ph.D. in electrical engineering (UPC, 1986 and 1989).
- Visiting professor, 1992, Signal and Image Processing Institute, University of Southern California.
- Marconi Young Scientist Award (1992)
- Cofounder (1999) of VERBIO (spoken language technology).
- Director of the Center for Language and Speech Technologies and Applications (TALP) 2006-2019.
- +150 scientific papers in statistical signal analysis and their application in communication systems, speech processing and machine translation
- 1st prize winner of GE Flight Quest 2.
- Highest global rank achieved on Kaggle: 3rd
- [Futur](#), [Google Scholar](#), [Linkedin](#), [Kaggle](#)

Instructors

Javier Hernando



1988 Telecommunication Engineer, UPC

1991 Professor ETSETB, now Full Professor

1993 Doctor pel TSC Extraordinary Award

2002 Consultor @ Panasonic, Santa Barbara, California

2003 Coordinator European Master of Research on IT (MERIT)

2009 Consultor @ Herta Security, S.L.

2016 Director TALP Research Center

Instructors

Santi Pascual

PhD Candidate

TALP Research Center, Universitat Politècnica de Catalunya



Telefónica

Instructors

Elisa Sayrol

Associate Professor at Universitat Politècnica de Catalunya (UPC)



Web: <https://imatge.upc.edu/web/people/elisa-sayrol>



UNIVERSITAT POLITÈCNICA DE CATALUNYA
BARCELONATECH

Department of Signal Theory
and Communications

Image Processing Group

Instructors

Xavier Giro-i-Nieto

Associate Professor at Universitat Politècnica de Catalunya (UPC)

- Web: <https://imatge.upc.edu/web/people/xavier-giro>



UNIVERSITAT POLITÈCNICA DE CATALUNYA
BARCELONATECH

Department of Signal Theory
and Communications

Image Processing Group



SONY



COLUMBIA UNIVERSITY
IN THE CITY OF NEW YORK

Teaching assistants

Carlos Escolano

Pre-doctoral researcher at Universitat Politècnica de Catalunya (UPC)

Ahmad Mel

Pre-doctoral researcher at Universitat Politècnica de Catalunya (UPC)

Manel Baradad

Master MET student at Universitat Politècnica de Catalunya (UPC)

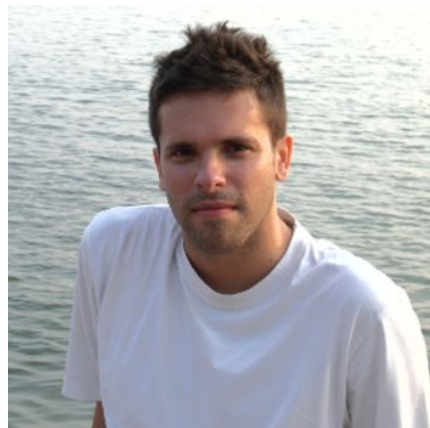
Albert Aparicio

BSc student at Universitat Politècnica de Catalunya (UPC)

Invited Talks

Joan Serra

Telefonica



Title: Facts and myths about deep learning.

Abstract: Deep learning has revolutionized the traditional machine learning pipeline, with impressive results in domains such as computer vision, speech analysis, or natural language processing. The concept has gone beyond research/application environments, and permeated into the mass media, news blogs, job offers, startup investors, or big company executives' meetings. But what is behind deep learning? Why has it become so mainstream? What can we expect from it? In this talk, I will highlight a number of facts and myths that will provide a shallow answer to the previous questions. While doing that, I will also highlight various applications we have worked on at our lab. Overall, the talk wants to place a series of basic concepts, while giving ground for reflection or discussion on the topic.

Invited Talks

Jordi Pons



Title: Deep learning for Music Informatics Research

Abstract: A brief review of the state-of-the-art in music informatics research and deep learning reveals that such models achieved competitive results for several tasks in a relatively short amount of time. Due to these promising results, some researchers declare that is the time for a paradigm shift: from hand-crafted features and shallow classifiers to deep processing models. In the past, introducing machine learning for global modeling (ie. classification) resulted in a significant state-of-the-art advance. And now, some researchers think that another advance could be done by using data-driven feature extractors based on deep learning instead of using hand-crafted features. However, deep learning for music informatics research is still in its early ages - current systems are based on solutions proposed for computer vision or speech. We will present our work describing how to adapt these technologies for the music case.

Management

Instructor	Area
Xavier Giró	Coordination
Toni Bonafonte	Sponsors
Elisa Sayrol	Logistics
Santi Pascual	Web
Marta R. Costa-jussà	Lecture planning
Javier Hernando	Grading
José Adrián Rodríguez Fonollosa	Project

Deep Learning



What society thinks I do



What my friends think I do



What other computer scientists think I do



What mathematicians think I do



What I think I do

```
from theano import *
```

What I actually do

Motivation

The best job in the world: Data scientist.

25 Best Jobs in America

f 28k



in 9.4k

g+ 412

Want a new job? Glassdoor is here to help, identifying the 25 Best Jobs in America for 2016. The jobs that make this list have the highest overall Glassdoor Job Score, determined by combining three key factors – number of job openings, salary and career opportunities rating. These jobs stand out across all three categories.

United States ▼

2016 ▼

glassdoor.com

1



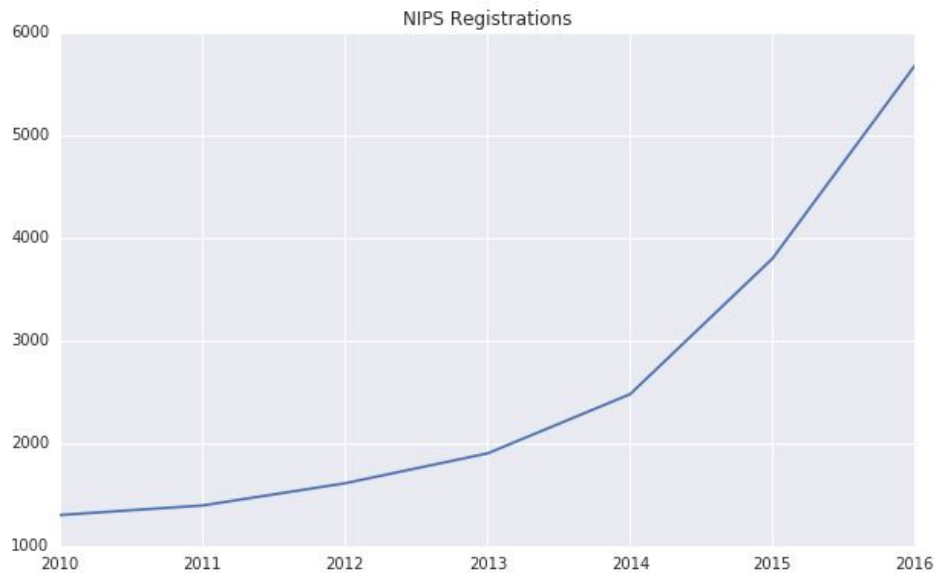
Data Scientist

Job Openings	1,736
Median Base Salary	\$116,840
Career Opportunity	4.1
Job Score	4.7

Source: [25 Best jobs in America](#) (Glassdoor)

Motivation

Nature, [“AI talent grab sparks excitement and concern”](#) (26/04/2016)



Motivation

The Economist, “[Artificial Intelligence has taken off](#)” (06/2016)



learning how to learn

Motivation

Nvidia, Deep Learning Applications, [NEWSLETTER](#)

[Lip Reading](#)

[“Voice Recognition Technology Almost as Accurate as Human”](#)

[“Emotionally Intelligent Speaker”](#)

Our Motivation

UPC, Deep Learning in Speech and Language, [BLOG](#)

Machines learn by themselves (with raw data)

less cleaning of data, less restrictions of languages, define architecture but not features...

	Tuesday 24	Wednesday 25
10:00	Welcome (M/X)	DNN/DBN (E)
10:20	Perceptron (S)	Recurrent (S)
10:40	Deep (E)	
11:00	Architectures I (X)	Word Embeddings (T)
11:20	Backprop (E)	Keras (JA)
11:40	Training (E)	
12:00	Keras (JA)	
12:20		
12:40		Architectures II (X)
13:00	Project (JA)	Project (JA)
13:20		
13:40		

	Thursday 26	Friday 27	Tuesday 31
10:00	Language Model (M)	Speaker Recognition II (J)	Project presentations
10:30	Speech Recognition I (JA)	Machine Translation II (M)	
11:00	Speaker Recognition I (J)	Industry talk: Joan Serrà	
11:30	Machine Translation I (M)		
12:00	Parametric Synthesis (T)	Waveform generation (T)	Break
12:30	Speech Recognition II (JA)	Multimodal (X)	Invited talk: Jordi Pons
13:00	Project (JA)	Project (JA)	
13:30			Closing

Course site

piazza

<http://piazza.com/upc/spring2017/etsetb230362/home>

Grading

	BSc	MSc
Lectures	70%	30%
Practical	20%	60%
Attendance	10%	10%

Grading: Online tests



WWW.PHDCOMICS.COM