ANX-PR/CL/001-01

LEARNING GUIDE

SUBJECT

93000819 - Deep Learning For Acoustic Signal Processing

DEGREE PROGRAMME

09AQ - Master Universitario En Ingenieria De Telecomunicacion

ACADEMIC YEAR & SEMESTER

2023/24 - Semester 2

Index

Learning guide

1. Description

1.1. Subject details

Name of the subject	93000819 - Deep Learning For Acoustic Signal Processing
No of credits	6 ECTS
Type	Optional
Academic year ot the programme	Second year
Semester of tuition	Semester 4
Tuition period	February-June
Tuition languages	English
Degree programme	09AQ - Master Universitario en Ingenieria de Telecomunicacion
Centre	09 - Escuela Tecnica Superior De Ingenieros De Telecomunicacion
Academic year	2023-24

2. Faculty

2.1. Faculty members with subject teaching role

Name and surname	Office/Room	Email	Tutoring hours *
Eduardo Lopez Gonzalo	C-330	eduardo.lopez@upm.es	Sin horario. Appointment arranged by email
Luis Alfonso Hernandez Gomez (Subject coordinator)	C-330	luisalfonso.hernandez@upm.es	Sin horario. Appointment arranged by email

Mateo Jose Camara Largo	C-301	mateo.camara@upm.es	Sin horario. Appointment arranged by email
Juan Ignacio Godino Llorente	C-312	ignacio.godino@upm.es	Sin horario. Appointment arranged by email

* The tutoring schedule is indicative and subject to possible changes. Please check tutoring times with the faculty member in charge.

3. Prior knowledge recommended to take the subject

3.1. Recommended (passed) subjects

- Aprendizaje Predictivo Y Descriptivo

- Laboratorio De TÉcnicas De Aprendizaje AutomÁtico

3.2. Other recommended learning outcomes

- Signal Processing, Speech and Audio Processing

- It is important to have some previous knowledge of Machine Learning and Deep Learning tecniques

- Previous exposure to a programming language, such as MATLAB, R or Python

4. Skills and learning outcomes *

4.1. Skills to be learned

CE1 - Capacidad para aplicar métodos de la teoría de la información, la modulación adaptativa y codificación de canal, así como técnicas avanzadas de procesado digital de señal a los sistemas de comunicaciones y audiovisuales.

CG1 - Poseer y comprender conocimientos que aporten una base u oportunidad de ser originales en el desarrollo y/o aplicación de ideas, a menudo en un contexto de investigación.

CG2 - Que los estudiantes sepan aplicar los conocimientos adquiridos y su capacidad de resolución de problemas en entornos nuevos o poco conocidos dentro de contextos más amplios (o multidisciplinares) relacionados con su área de estudio.

CG4 - Que los estudiantes sepan comunicar sus conclusiones ?y los conocimientos y razones últimas que las sustentan? a públicos especializados y no especializados de un modo claro y sin ambigüedades.

CG5 - Que los estudiantes posean las habilidades de aprendizaje que les permitan continuar estudiando de un modo que habrá de ser en gran medida autodirigido o autónomo.

CT1 - Capacidad para comprender los contenidos de clases magistrales, conferencias y seminarios en lengua inglesa.

CT3 - Capacidad para adoptar soluciones creativas que satisfagan adecuadamente las diferentes necesidades planteadas.

CT4 - Capacidad para trabajar de forma efectiva como individuo, organizando y planificando su propio trabajo, de forma independiente o como miembro de un equipo.

CT5 - Capacidad para gestionar la información, identificando las fuentes necesarias, los principales tipos de documentos técnicos y científicos, de una manera adecuada y eficiente.

4.2. Learning outcomes

RA295 - Competence on technologies for extracting knowledge from a variety of acoustic signals combining signal processing and machine learning

RA10 - Saber realizar una presentación de carácter técnico, ante una audiencia de pares, que describa el trabajo realizado y sus resultados, de forma clara y bien estructurada, en el tiempo establecido, y usando un lenguaje preciso

* The Learning Guides should reflect the Skills and Learning Outcomes in the same way as indicated in the Degree Verification Memory. For this reason, they have not been translated into English and appear in Spanish.

5. Brief description of the subject and syllabus

5.1. Brief description of the subject

This course will present state-of-the-art technologies for processing, extracting knowledge and generating acoustic signals. The course will consider a broad range of applications: speech, music, audio, environmental acoustics, forensics acoustics, biomedical applications, bots and human-machine interaction, speech-to-speech translation, assistive technologies etc.

Together with the basic concepts of different acoustic fields main emphasis in the course will be put on the combination of signal processing and machine learning, particularly deep neural networks, for generating audio signals (speech and music) and extracting relevant knowledge from them (recognition of speech, speaker, language, emotions, musical genre, acoustic events etc.). Most of the activities in the course will be practical using latest technologies such Tensorflow/Keras and PyTorch for machine learning and deep learning applied on acoustic signals.

5.2. Syllabus

1. Fundamentals of Acoustics

    1.1. Reviewing basic concepts of sounds and the acoustic environment

    1.2. Overview of application fields: speech, music, audio, environmental acoustics, forensics acoustics, biomedical applications, bots and human-machine interaction, speech-to-speech translation, assistive technologies, etc.

2. Acoustic Signals and Systems

    2.1. Acoustic sources: signals characterization

    2.2. Time-frequency analysis

    2.3. Time-space analysis

    2.4. Physical and parametric acoustic models

3. Processing and extracting knowledge from acoustics

    3.1. Reviewing traditional feature extraction technique for acoustic signals

    3.2. Machine learning approaches in acoustics

4. Reviewing Machine learning

    4.1. Basic machine learning algorithms

    4.2. Introduction to machine learning Tools

5. Deep Neural Networks in acoustics

    5.1. Overview of applications and challenges of Deep Neural Networks in acoustics

    5.2. Practical Introduction to Deep Neural Networks using Keras & PyTorch

    5.3. Deep Learning applications

        5.3.1. Modeling and extracting acoustic features using Feed Forward and Convolutional Networks

        5.3.2. Combining Signal Processing and Deep Learning: Differentiable Digital Signal Processing

        5.3.3. Recurrent Networks: LSTM, Bi-directional LSTM, GRU, etc.

        5.3.4. Attention mechanism & Transformers

        5.3.5. Advanced Deep Learning technques in Audio

            5.3.5.1. Deep Generative Models

            5.3.5.2. Self-supervised Learning

5.3.5.3. Reinforcement Learning

5.3.6. Extracting knowledge: speech, natural language processing (NLP), speaker, language, musical genre, etc.

5.3.7. Audio synthesis: text-to-speech, music, audio effects, de-noising, sound source separation, ..

5.3.8. Interactive systems: bots, human-machine interaction, assistive technology, hearing aids,..

6. Schedule

6.1. Subject schedule*

Week	Classroom activities	Laboratory activities	Distant / On-line	Assessment activities
1	**Fundamentals of acoustics** Duration: 04:00 Lecture			
2	**Acoustics Signals & Systems** Duration: 04:00 Lecture			
3	**Reviewing Course Activities (Course Report)** Duration: 01:00 Cooperative activities	**Acoustics Signals & Systems** Duration: 03:00 Laboratory assignments		
4	**Processing and extracting knowledge from acoustics** Duration: 01:00 Lecture	**Machine Learning on Raw Audio** Duration: 01:00 Laboratory assignments **Machine learning on Audio Spectrum** Duration: 02:00 Laboratory assignments		
5	**Reviewing Course Activities (Course Report)** Duration: 02:00 Cooperative activities	**Reviewing Machine learning in Audio and Acoustics** Duration: 02:00 Laboratory assignments		
6	**Deep Learning in Acoustics** Duration: 02:00 Lecture	**Introduction to Deep Learning using Keras** Duration: 02:00 Laboratory assignments		
7		**FeedForward models on Raw and Spectral data** Duration: 04:00 Laboratory assignments		
8		**Introduction to Convolutional models** Duration: 04:00 Laboratory assignments		
9		**Understanding CNN models on Raw and Spectral data** Duration: 04:00 Laboratory assignments		
10	**Reviewing Course Activities (Course Report)** Duration: 02:00 Cooperative activities	**Differentiable Signal Processing and Deep Learning** Duration: 02:00 Laboratory assignments		
11		**Signal Procesing and Deep Learning examples: DDSP, Neural Synthesis** Duration: 04:00 Laboratory assignments		

12		**Introduction to Recurrent models: LSTM, GRU** Duration: 02:00 Laboratory assignments **Audio and Music Generation** Duration: 02:00 Laboratory assignments		
13		**Seq2Seq, Attention, Transformers** Duration: 02:00 Laboratory assignments **Attention & Transformers: Speech/Audio Recognition, Synthesis & Analysis** Duration: 02:00 Laboratory assignments		
14		**Review advanced applications** Duration: 04:00 Laboratory assignments		
15				
16				
17				**Course Report evaluation** Individual work Continuous assessment Not Presential Duration: 00:00 **Course Report evaluation** Individual work Final examination Not Presential Duration: 00:00

Depending on the programme study plan, total values will be calculated according to the ECTS credit unit as 26/27 hours of student face-to-face contact and independent study time.

* The schedule is based on an a priori planning of the subject; it might be modified during the academic year, especially considering the COVID19 evolution.

7. Activities and assessment criteria

7.1. Assessment activities

7.1.1. Assessment

Week	Description	Modality	Type	Duration	Weight	Minimum grade	Evaluated skills
17	Course Report evaluation	Individual work	No Presential	00:00	100%	5 / 10	CG4 CT3 CT4 CT1 CT5 CG2 CG5 CG1 CE1

7.1.2. Global examination

Week	Description	Modality	Type	Duration	Weight	Minimum grade	Evaluated skills
17	Course Report evaluation	Individual work	No Presential	00:00	100%	5 / 10	CT4 CG4 CT3 CT1 CT5 CG2 CG5 CG1 CE1

7.1.3. Referred (re-sit) examination

Description	Modality	Type	Duration	Weight	Minimum grade	Evaluated skills
Course Report evaluation	Individual work	Face-to-face	00:00	100%	5 / 10	CG4 CT3 CT4 CT1 CT5 CG2 CG5 CG1 CE1

7.2. Assessment criteria

Evaluation will assess if students have acquired all the competences of the subject. Thus, evaluation through extraordinary assessment will be carried out considering all the evaluation techniques used in ordinary evaluation (EX, ET, TG, etc.).

Progressive evaluation will be the preferred assessment method as it will be suited to the optimum learning process along the course. It will consist of the generation of a detailed Course Report to demonstrate students' abilities to apply Machine Learning and Deep Learning models on acoustic signals. The evaluation of the Course Report will represent 100% of final grade and it must be due by the final exam date.

The Course Report will also contain details and related experimental materials on specific final projects that could be developed in working teams. But each team member must prepare an individually report of these activities in her/his Course Report.

The Course Report will be worked on and continuously supervised during the course through several course assignments announced in Moodle. In that way students will receive updated feedback on their progress and continuous effort along the course. Students will be required to attend to specific presentations and final presentations to defend their work as it is going to be developed in their Report.

Global or final evaluation will also consist of the generation of a detailed Report to demonstrate students' abilities

to apply Machine Learning and Deep Learning models on acoustic signals. Specific experimental materials developed to address practical activities will be also submitted to be evaluated. Students will be required attend to a final presentation to defend their work.

Evaluation through extraordinary assessment will require the same process as the one described before for Global or final evaluation.

8. Teaching resources

8.1. Teaching resources for the subject

Name	Type	Notes
Handbook of signal processing in acoustics	Bibliography	Havelock, David, Sonoko Kuwano, and Michael Vorländer, eds. Handbook of signal processing in acoustics. Springer Science & Business Media, 2008.
Introduction to Statistical Learning	Bibliography	James, Gareth, et al. An introduction to statistical learning. Vol. 112. New York: springer, 2013.
Neural Networks and Deep Learning	Web resource	http://neuralnetworksanddeeplearning.com/
Intro Machine Learning Python	Web resource	https://www.dataquest.io/blog/machine-learning-python/
Librosa tutorial	Web resource	https://github.com/librosa/tutorial
Audio Analysis in Python	Web resource	https://github.com/tyiannak/pyAudioAnalysis/wiki
Audio Processing Using Deep Learning in MATLAB	Web resource	https://es.mathworks.com/help/deeplearning/audio-processing-using-deep-learning.html
Simple Audio Recognition	Web resource	https://www.tensorflow.org/versions/master/tutorials/audio_recognition

Tensorflow Intro	Web resource	https://www.tensorflow.org/get_started/
Keras Audio examples	Web resource	https://keras.io/examples/audio/
PyTorch Audio Tutorial	Web resource	https://pytorch.org/tutorials/beginner/audio_preprocessing_tutorial.html
Notes on Music Information Retrieval	Web resource	https://musicinformationretrieval.com/
DDSP: Differentiable Digital Signal Processing	Web resource	https://magenta.tensorflow.org/ddsp
Music and Art Using Machine Learning	Web resource	https://magenta.tensorflow.org/
Course Materialas at Moodle	Web resource	https://moodle.upm.es/titulaciones/oficiales/course/view.php?id=891

9. Other information

9.1. Other information about the subject

From a broad perspective, acquiring skills on technological fields as powerful as machine learning and deep learning will help to increase the number of youth who have relevant skills, including technical and vocational skills, for employment, decent jobs and entrepreneurship, thus fostering Goal 4.4 in Sustainable Development Goals (SDGs) 2030 United Nations Agenda.

More specifically, contents in the course related to music will remark the universal nature of music to connect people and to become an active a driver to improve our lives, our communities and our planet; this is the direction of initiatives such as http://MusicForSDGs.com.

Furthermore, the study of environmental acoustics and bioacoustics will help our students to learn new ways to use technology to contribute to sustainability and combat climate change as included in SDG Goals 13 and 14.