MOOC UGR Machine Learning and Big Data for Bioinformatics
GUIDE

DESCRIPTION

This course was designed by the University of Granada as a massive open online course (MOOC). The course will be taught in Spanish, English and is free of charge. This MOOC aims to offer a motivational, direct, and simple learning experience in terms of methodology, without affecting the rigor and depth of its content. To this end, the topics were designed and structured by a team of teachers, research staff, experts, and specialists in each of the areas covered. Our main aim was to offer a practical and applied vision of the tasks related to machine learning and its main techniques.

OBJETIVES

✓ GENERAL AIMS:

To provide a basic guide to tackle a specific scientific problem in biosciences and/or biohealth by leveraging the power of machine learning and big data techniques.

✓ SPECIFIC AIMS:

1. Introduce the course participants to the basic concepts of bioinformatics.

2. Provide knowledge about the basic steps involved in the analysis of bioinformatics data.

3. Tackle a biosciences problem by applying machine learning techniques, either by taking a supervised or unsupervised approach.

4. Tackle a big data problem with the use of machine learning techniques.

5. Learn about a graphical tool that can also be used for data analysis in this context.

MODALITY

The course is divided into 8 modules and will last for a total of 8 consecutive weeks. The modality is virtual, and the course will be imparted using an online methodology. There is no fixed timetable for participation in this MOOC and so each user can flexibly choose the most suitable time for them to complete the tasks set each week. However, we recommended trying to adhere to the proposed pace for the course so that you do not miss out on participation within its online community.

A new module will be become available on the Monday of each week, and on Friday of the same week, access will be granted to the corresponding module questionnaire. In addition, there are no deadlines for completing the questionnaires or participating in the forums and proposed activities. The only deadline is the end of the course itself. Similarly, the only requirement for participation is the availability of an internet connection.

CONTENTS AND TIMING

✓ CONTENTS

The knowledge required to meet the proposed objectives for this course will be provided through the contents comprising it. As shown in the timetable below, this MOOC has been divided into 8 modules, with each one corresponding to a week of work:

CONTENTS	
Module 1: What is bioinformatics? (Coordinators: *Carlos Cano, Coral del Val y Pedro Carmona*) - **Bioinformatics: what, why, and how?** - **Bioinformatics: applications in biosciences and biohealth** - **Precision medicine** (bioinformatics for the interpretation of cancers).	**Module 2: Bioinformatic analysis of an omics problem** (Coordinators: *Carlos Cano, Coral del Val y Pedro Carmona*) - **The problem: how do we obtain and prepare data?** - **Preprocessing and exploratory analysis** - **Functional analysis**

MACHINE LEARNING AND BIG DATA FOR BIOINFORMATICS

CONTENTS	
Module 3: Data science and machine learning (Coordinator: *Alberto Fernández Hilario*) - Data science and machine learning: what, why, and how? - Supervised learning - Unsupervised learning	**Module 4: Supervised learning: regression techniques** (Coordinator: *Rafael Alcalá Fernández*) - Regression: what, why, and how? - Standard regression methods - Machine learning methods for regression
Module 5: Supervised learning: classification techniques (Coordinator*: Alberto Fernández Hilario*) - Classification: what, why, and how? - Standard classification methods - Advanced classification methods	**Module 6: Unsupervised learning: clustering and association rules** (Coordinator*: Jesús Alcalá Fernández*) - Clustering and association rules: what, why, and how? - Clustering - Association rules
Module 7: Big Data (Coordinator: *Francisco Javier García Castellano*) - Introduction to big data - Introduction to Spark using Python - Machine learning in *Spark*: *MLlib* - Classification, regression, clustering, and association rules in Spark.	**Module 8: Graphic tools: the Konstanz Information Miner** (Coordinators*: María Martínez y José Manuel Soto*) - How to use KNIME: workflows - How to solve a problem with KNIME - Supervised learning: regression and classification methods - Unsupervised learning: clustering and association rules

✓ TIMETABLE (the calendar for this MOOC can be found inside the course learning area, on the top right-hand side of the platform, where you will also find all the important dates for its completion).

MACHINE LEARNING AND BIG DATA FOR BIOINFORMATICS

METHODOLOGY

Massive open online courses (MOOCs) are based on a non-attendance work methodology. As the acronym indicates, they are free, massive, open, and online training courses. This means that students must access the work platform via the internet, having previously registered and enrolled in the course. Given that the course is designed for anyone interested in the subject matter, registration is free of charge and can be done from anywhere and by anyone.

Each week a new topic will be dealt with in each module, which themselves each comprise several learning units consisting of a presentation video and assessable content, together with a series of activities that must be completed, as set out in the associated forums. The resources and materials made available to students were designed specifically for this course.

MATERIALS

- Multimedia videos made by UGR specialists in each subject.
- Content elaborated by each specialist which delves into each of the learning units.
- References and recommended bibliography.
- Glossary and other resources.

PROPOSED ACTIVITIES

- Creation of a profile on the platform.
- Participation in the forums.
- Watching videos.
- Reading the contents presented.
- Completion of the evaluation questionnaires for each module.
- Transversal participation and completion of the proposed activities

EVALUATION AND RECOGNITION

The Vice-rectorate of teaching at the UGR has recognized this MOOC, which has a total duration of 100 hours, for the award of 4 ECTS credits towards any of the degrees at the UGR. MOOC courses are framed within the paradigm of non-formal education and lifelong learning, and so their completion is voluntary.

Each of the 8 modules will be opened successively each week, without having to first pass the previous one, thereby allowing each student to continue progressing through the course at a pace suitable for them. In the middle of each week, the questionnaire corresponding to the ongoing module will become available, giving the course's students time to look at its contents and participate in the forums.

A badge related to the proposed syllabus will be available for each module in the course and can be obtained by viewing the materials in each module and successfully completing the questionnaire about its contents. The badges obtained for each module, as shown below, will be visible on each student's profile as they are attained.

Module 1

Module 2

Module 3

Module 4

Module 5

Module 6

Module 7

Module 8

END COURSE BADGE