SOSC 2019 Third International School on Open Science Cloud
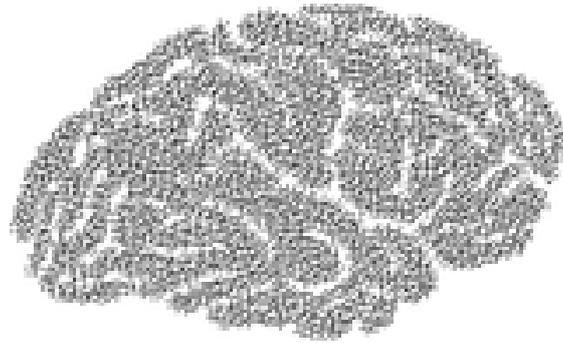
Report of Contributions

Contribution ID: **37**   Type: **not specified**

Introduction to Statistical Learning

Monday, 16 September 2019 09:15 (1 hour)

Statistical learning refers to a vast set of tools and techniques for modelling and understanding complex datasets. A great deal of interest on these aspects has recently surged in the statistical and computer science communities, and statistical learning has become a very hot field in many applications, such as marketing, finance, and other business disciplines. In this short introduction we review basic ideas of supervised and unsupervised learning, classification and regression, statistical models for inference and prediction, interpretability vs flexibility, bias-variance trade-off, overfitting.

Presenter: Prof. SCRUCCA, Luca (University of Perugia)

Session Classification: Lectures

Contribution ID: **38**   Type: **not specified**

Introduction to Intelligent Systems

Monday, 16 September 2019 10:15 (1 hour)

Intelligent systems are ubiquitous. Their growth and deployment are addressing industrial, scientific and societal challenges. A brief overview of the applications of intelligent systems and predictive models - with focus on current and future opportunities - is offered. A major focus is given on scientific challenges in medical sciences, chemistry, bioinformatics, geosciences and physics.

Presenter: Prof. BONACORSI, Daniele (University of Bologna)

Session Classification: Lectures

Contribution ID: **39**　　　　　　　　　　　　　　　　　　Type: **not specified**

Introduction to Intelligent Infrastructures

Monday, 16 September 2019 11:45 (1 hour)

What do we mean by "data and compute infrastructures" and to what extent they can be considered "Intelligent"? How and why are these terms changed over the past 20 years? This lecture will introduce the main results and challenges for efficient exploitation of data-driven research.

Presenter: Dr SALOMONI, Davide (INFN CNAF)

Session Classification: Lectures

Contribution ID: **40**   Type: **not specified**

Software ecosystem for Data Science

Monday, 16 September 2019 14:30 (2 hours)

Presenters: Prof. BONACORSI, Daniele (University of Bologna); Dr KUZNETSOV, Valentin (Cornell University)

Session Classification: Hands-on

Contribution ID: **41**

Type: **not specified**

Introduction to the mini-kaggle competition

Monday, 16 September 2019 17:00 (2 hours)

Presenters: Dr KUZNETSOV, Valentin (Cornell University); Prof. BONACORSI, Daniele (University of Bologna)

Session Classification: Hands-on

Contribution ID: **42**   Type: **not specified**

Machine/Deep Learning in HEP

Tuesday, 17 September 2019 10:00 (1 hour)

The speaker will give an overview of the utilisation of ML/DL techniques in High Energy Physics, with and without accelerators, giving (selected) examples. The talk will present the impact of such techniques in triggering, tracking and reconstruction, end-user data analysis, computing infrastructure optimization, etc. The talk will also discuss the transition from the use of field-specific knowledge for feature engineering to the use of full high-dimensional feature space to train cutting-edge ML algorithms, and will discuss possible evolutions of the role of ML/DL in HEP.

Summary

Presenter: Prof. BONACORSI, Daniele (University of Bologna)

Session Classification: Lectures

Contribution ID: **43**   Type: **not specified**

Data Science: state of the art

Tuesday, 17 September 2019 09:00 (1 hour)

The speaker will present the state of the art in data science, the role of (big) data in it, the skills needed to perform efficiently as a data scientist, and the key concepts around data, algorithms and techniques in the field.

Summary

Primary author: Dr KUZNETSOV, Valentin (Cornell University)

Presenter: Dr KUZNETSOV, Valentin (Cornell University)

Session Classification: Lectures

Contribution ID: 44
Type: **not specified**

Resampling Methods

Tuesday, 17 September 2019 11:30 (1 hour)

Choosing a proper level of complexity for a prediction rule (model selection) or evaluating its performance (model assessment) are two fundamental steps in any supervised statistical learning application. Both steps require reliable estimates of the expected prediction error. After providing some general definitions related to the prediction error in regression and classification, attention will be focused on estimating this error by re-using the observed data. In particular, methods based on validation sets, cross-validation and bootstrap will be illustrated. Advantages and disadvantages of each technique will be discussed.

Presenter: Prof. GALIMBERTI, Giuliano (University of Bologna)

Session Classification: Lectures

Contribution ID: **45**　　　　　　　　　　　　　　　　Type: **not specified**

Infrastructure as Code and Template-based orchestration in hybrid Clouds.

Thursday, 19 September 2019 15:30 (1 hour)

Summary

Presenter: TRACOLLI, Mirco (PG)

Session Classification: Lectures

Contribution ID: 46　　　　　　　　　　　　　　　　　　　Type: **not specified**

Containers, their composition and orchestration

Thursday, 19 September 2019 09:00 (1 hour)

Presenter:　Dr SALOMONI, Davide (INFN CNAF)

Session Classification:　Lectures

Contribution ID: 47
Type: **not specified**

Function as a Service and data streaming analytics in cloud infrastructures.

Thursday, 19 September 2019 14:30 (1 hour)

Presenter: Dr SPIGA, Daniele (INFN Perugia)

Session Classification: Lectures

Contribution ID: 48　　　　　　　　　　　　　　　　　Type: **not specified**

Intelligent Systems in Industry

Presenter: MAGUGLIANI, Fabrizio (E4 Computer Engineering)

Session Classification: Lectures

Contribution ID: 49

Type: **not specified**

Intelligent Systems in Industry

Session Classification: Lectures

Contribution ID: **50** Type: **not specified**

Wrap-up of the school and follow-up opportunities for students

Friday, 20 September 2019 11:30 (1 hour)

Session Classification: Lectures

Contribution ID: **51**   Type: **not specified**

Welcome message and Introduction to the School

Monday, 16 September 2019 09:00 (15 minutes)

Presenters: Dr MARON, Gaetano (INFN CNAF and Legnaro);   Dr SALOMONI, Davide (INFN CNAF)

Contribution ID: **52**   Type: **not specified**

Hands-on: FaaS, analytics and template-based orchestration.

Thursday, 19 September 2019 17:00 (2 hours)

Presenter: CIANGOTTINI, Diego (PG)

Session Classification: Hands-on

Contribution ID: **53**   Type: **not specified**

Hands-on: Mini kaggle intermediate wrap-up, and a new challenge

Presenter: Dr KUZNETSOV, Valentin (Cornell University (US))

Session Classification: Hands-on

Contribution ID: **54** Type: **not specified**

Hands-on: Mini-kaggle + image classification (Part I)

Wednesday, 18 September 2019 16:30 (2 hours)

Presenter: Dr KUZNETSOV, Valentin (Cornell University (US))

Session Classification: Hands-on

Contribution ID: **55** Type: **not specified**

(Mini-kaggle + image classification (Part II))

Friday, 20 September 2019 09:00 (2 hours)

Presenter: Dr KUZNETSOV, Valentin (Cornell University (US))

Session Classification: Hands-on