

IASC-LARS WEBINAR ON COMPUTATIONAL STATISTICS AND DATA SCIENCE



IASC- LARS SCHOOL ON COMPUTATIONAL
STATISTICS AND DATA SCIENCE

MACHINE LEARNING FOR DATA SCIENCE

RODRIGO SALAS, MARVIN QUERALES, AYLEEN BERTINI
UNIVERSIDAD DE VALPARAÍSO, CHILE



JAVIER LINKOLK LÓPEZ
UNIVERSIDAD PERUANA UNIÓN, PERÚ



SEPTEMBER 27-30, 2021

INTERNATIONAL ORGANIZING COMMITTEE

ALBA MARTÍNEZ-RUIZ, MONDAY ADENOMON, PAULO CANAS RODRIGUES
HAN-MING (HANK) WU, SARA TASKINEN, CRISTIAN GATU
LUIS FIRINGUETTI-LIMONE, DAVID MUÑOZ-NEGRON

IASC-LARS Webinar on Computational Statistics and Data Science

MACHINE LEARNING FOR DATA SCIENCE

PROGRAM



IASC- LARS SCHOOL ON COMPUTATIONAL STATISTICS AND DATA SCIENCE

The Latin American Regional Section of the International Association for Statistical Computing (IASC-LARS), the IASC-LARS School on Computational Statistics and Data Science, the International Association for Statistical Computing (IASC), and the International Statistical Institute (ISI) are pleased to invite postgraduate and undergraduate students to attend the IASC-LARS Webinar Course “Machine Learning for Data Science”. The course will be taught by Professor Rodrigo Salas and Collaborators Marvin Querales and Ayleen Bertini from the Universidad de Valparaíso, Chile, and Javier Linkolk López from the Universidad Peruana Unión, Perú, September 27-30, 2021. The course will be given in an English/Spanish hybrid format.

The IASC-LARS Courses aim (1) to spread the knowledge base and advances in Statistical Computing in Latin America and the world, (2) to provide an overview of the state-of-the-art of the ongoing research in computational statistics, (3) to provide an overall perspective of the application of computational statistics in data science problems, (4) to present applications where computational statistics have been crucial to solve problems in real-life applications, and (5) to increase the number of researchers and practitioners in computational statistics and data science.

IASC-LARS Webinar on Computational Statistics and Data Science

MACHINE LEARNING FOR DATA SCIENCE

PROGRAM

Machine learning is a research and development field that has seen tremendous growth in recent times. On the one hand, many applications have appeared in multiple areas, such as engineering and healthcare, to name a few. On the other hand, these models have been widely studied in their theoretical aspects to know their properties and limitations.

Technological and computational development has meant that large volumes of data are being collected at every moment. A new discipline known as Data Science has emerged, generating new and innovative solutions from data. In this sense, learning machines are one of the main tools used by the data scientist.

In this webinar, we will study the theoretical foundations of learning machines in conjunction with some of the more popular methods. Hands-On type activities will complement the teaching using the Python language with its toolboxes. Additionally, machine learning models will be applied in problems related to the environment and health.

INSTRUCTORS: Rodrigo Salas (Universidad de Valparaíso, Chile).
Javier-Linkolk López (Universidad Peruana Unión, Perú).
Marvin Querales (Universidad de Valparaíso, Chile).
Ayleen Bertini (Universidad de Valparaíso, Chile).

AGENDA

(Local time: Valparaiso – Chile, GMT -4 hours)

Monday – September 27

session 1: 9:00-10:00

session 2: 10:15-12:00

- Introduction to Python
 - History of Python
 - Installing Python: Anaconda, Miniconda, Google Colab.
 - Programming Language
- Data Analysis with Python
 - Toolbox for Data Science: PANDAS, Scikit-Learn, Scipy, Numpy, Statmodels.
 - Hands On: “*Statistical Data Analysis with Python*”

Tuesday – September 28

session 1: 9:00-10:00

session 2: 10:15-11:15

session 3: 11:30-13:30

- Introduction to Data Science
- Unsupervised Machine Learning Methods
 - Clustering Analysis
 - Principal Component Analysis
- Explainable Machine Learning Models
 - Logistic Regression
 - Decision Trees
 - Feature Evaluation
- Hands on: *“Explainable Machine Learning models for Health data”*

Wednesday – September 29

session 1: 9:00-10:00

session 2: 10:15-11:15

session 3: 11:30-13:30

- Fundamentals of Statistical Learning and Predictive Models.
- Supervised Machine Learning Methods
 - Linear Discriminant Analysis
 - Quadratic Discriminant Analysis
 - Random Forest, Gradient Boosting, XGBoost
 - Support Vector Machine
- Evaluation and Validation of Machine Learning Methods: Confusion Matrix, Performance Metrics, ROC Curve, Hold-one-out, Cross-Validation.
- Hands on: *“Time Series Forecasting of Pollution Data”*

Thursday – September 30

session 1: 9:00-10:00

session 2: 10:15-11:15

session 3: 11:30-13:30

- Artificial Neural Networks and Deep Learning
 - Multilayer Perceptron (MLP)
 - Recurrent Neural Networks (LSTM)
 - Convolutional Neural Networks (CNN)
- Neuro-fuzzy Inference Models: ANFIS, SONFIS, SINFIM
- Hands on: *“Deep Learning and Neuro-Fuzzy Inference framework for modeling Rainfall-Runoff”*.

Books

- M. J. Zaki, W. Meira Jr. “*Data Mining and Analysis. Fundamental Concepts and Algorithms*”. Cambridge University Press. 2014.
- C. Aggarwal. “*Data Mining. The textbook*”. Springer. 2015
- S. Raschka & V. Mirjalili. “*Python Machine Learning: Machine Learning and Deep Learning with Python, scikit-learn, and TensorFlow*”, 2nd Edition. 2017
- S. Masis. “*Interpretable Machine Learning with Python*”. Packt Publishing. 2021

Articles

- A. Encalada-Malca, J. Cochachi-Bustamante, P. Canas Rodrigues, R. Salas, J.L. López-Gonzales (2021). “A Spatio-Temporal Visualization Approach to the Exploration of PM10 Concentration Data in Metropolitan Lima”, *Atmosphere* 2021; 12(5):609, doi: 10.3390/atmos12050609, MDPI.
- Y. Morales, M. Querales, Harvey Rosas, H. Allende-Cid and R. Salas (2021) A Self-Identification Neuro-Fuzzy Inference framework for modeling Rainfall-Runoff in a Chilean watershed. *Journal of Hydrology*, Volume 594, pages 125910. doi: 10.1016/j.jhydrol.2020.125910.
- C. Parra, C. Ponce and R. Salas (2020). Evaluating the Performance of Explainable Machine Learning Models in Traffic Accidents Prediction in California. *Proceedings of the 2020 39th International Conference of the Chilean Computer Science Society (SCCC)*, IEEE Press. doi: 10.1109/SCCC51225.2020.9281196.
- A. Bertini, P. González, A. Holz, M.J. Varela, L. Sobrevia, R. Salas, Pardo (2020). “Identification of the main perinatal complications associated with pregestational obesity attended on the Aconcagua Health Service between the years 2015 – 2017, Chile”. XXXV Reunión Anual de la Sociedad Chilena de Ciencias Fisiológicas (SCHCF)
- H. Allende-Cid, R. Salas, A. Veloz, H. Allende, C. Moraga (2016) SONFIS: Structure Identification and Modeling with a Self-Organizing Neural Network. *International Journal of Computational Intelligence Systems*, vol 9, issue 4, pages 416-432. Atlantis Press.

Software

- Python 3.X using the framework of ANACONDA <https://www.anaconda.com/download/> or the framework of MINICONDA <https://docs.conda.io/en/latest/miniconda.html>
- Jupyter notebook (installed with anaconda or miniconda)
- Google Colab <https://colab.research.google.com/>
- Toolbox:
 - Numpy
 - Scipy
 - Matplotlib
 - Seaborn
 - Pandas
 - Scikit-Learn
 - Statmodels
 - Tensor-Flow
 - Keras

INSTRUCTORS



Rodrigo Salas Fuentes (Member IASC-ISI, and Board of Directors LARS) received the B.S. and MSc. degrees in Informatics Engineering and the Dr. Eng. degree in informatics from the Federico Santa María Technical University

(UTFSM) in Chile, in 2001, 2002 and 2010, respectively. From 2002 to 2004, he was a research assistant with the Informatics Department, UTFSM. Since 2004, has been with the Universidad de Valparaíso, where he is currently a Full Professor of the Biomedical Engineering School and teaches in Data Mining, Probability and Statistics, and Machine Learning. Dr. Salas is a main researcher of the Center of Research and Development in Health Engineering (CINGS-UV). His research interests include Artificial Intelligence, Machine Learning, Data Science, Decision Support Systems, Intelligent Systems and their applications to finance, air pollution, healthcare and medicine.



Javier Linkolk López Gonzales (Member IASC-ISI) received the B.S. degree in Statistical and Informatics Engineering from Universidad Peruana Unión (UPeU, Perú) and the M.Sc. degree in Metrology from Pontifical Catholic University of Rio de Janeiro (PUC-Rio, Brazil). Currently, He is a Ph.D.

candidate in Statistics of Universidad de Valparaíso (UV, Chile). His main research interests include pattern recognition in Machine Learning, air pollution with Deep Learning techniques, and Time Series with Singular Spectrum Analysis. He is an associate professor of Facultad de Ingeniería y Arquitectura, Universidad Peruana Unión.



Marvin Carrasquel (Member IASC-ISI) received his degree in Medical Technology from Universidad de Carabobo (Venezuela) in 2008 and the MSc. degree from Universidad Central de Venezuela in 2015. Recently, in 2021, he also received his Doctor

degree in Statistic from Universidad de Valparaíso (Chile). From 2010 to 2016, he was a Full Professor of the Health Sciences Faculty at Universidad de Carabobo, developing research in cardiovascular diseases, and applying statistical methods to biological sciences. From 2019 to 2020, he worked as part-time professor at Universidad Autónoma de Chile teaching Statistics and since 2021 he is working at Universidad de Valparaíso as Professor of Methodology and Statistics. His research interests include Data Science, Decision Support Systems, and their applications to healthcare and medicine.



Ayleen Bertini Rojas (Member IASC-ISI) Nutritionist and Dietitian graduated from the Universidad de Valparaíso, Chile. Certified in Food Psychology, a specialist in Chronic Diseases. Student of the Doctoral Program in Sciences and Engineering for Health at

the Universidad de Valparaíso. Research Assistant for the Laboratory for Research in Metabolic Diseases, Interdisciplinary Center for Research in Territorial Health of the Aconcagua Valley (CIISTe Aconcagua), Center for Biomedical Research, Universidad de Valparaíso, Chile. With experience in Interdisciplinary Research in Health and Data Science. Research interests: chronic diseases, diabetes, pregnancy, artificial intelligence for health.

IASC-LARS Webinar on Computational Statistics and Data Science

MACHINE LEARNING FOR DATA SCIENCE

REGISTRATION PROCEDURE

The IASC-LARS Webinar “Machine Learning for Data Science” will be held virtually using the platform [GoToWebinar](#) from September 27 to 30, 2021. The course will be given in an English/Spanish hybrid format. The deadline for registration is **September 10, 2021**. The course is free of charge.

Please complete the [IASC-LARS Webinar Registration Form](#) at <https://register.gotowebinar.com/register/8858121286906434319>

To become a IASC-LARS member, please complete the Membership Application Form at <https://www.isi-web.org/index.php/membership/individual-membership/iasc>.

All participants are expected to adhere to the ISI Community Principles and Conduct Policy (<https://www.isi-web.org/index.php/about-isi/policies/community-conduct>).

More information about GoToWebinar please visit <https://www.gotomeeting.com/webinar>. GoToWebinar application is also available for iOS, Android and Windows Phone: <https://support.goto.com/webinar/help/gotowebinar-for-mobile-devices-g2w050033>.

IASC-LARS Webinar on Computational Statistics and Data Science

MACHINE LEARNING FOR DATA SCIENCE GOTOWEBINAR PLATFORM

To attend this virtual course:

1. Please complete the IASC-LARS Webinar Registration Form at <https://register.gotowebinar.com/register/8858121286906434319> as soon as you decide to attend the course.
2. You will receive an email with an URL link to attend the webinar in the GTW platform.
3. Please, try to connect about 30 minutes before the start of the course.
4. We recommend joining via a high speed and wired connection and to use a USB headset for best sound quality.
5. Before to attend the course, please visit the webpage <https://www.gotomeeting.com/webinar/join-webinar> and see the video “GoToWebinar Attendee Quick Start”. For more information, you can also visit the YouTube channel <https://www.youtube.com/user/gotowebinar>.
6. If you have any questions or you need any help, please do not hesitate to contact us at iasc.lars.iasc@gmail.com.