ICME is offering 10 workshops this summer. Registration is required per workshop. Please see below for workshop descriptions and instructor biographies.

**Linear Algebra**

Matrix computations/linear algebra are the backbone of data science algorithms. If your linear algebra is rusty, or if you are not familiar with critical concepts (including orthogonal decompositions and least squares), then this workshop is for you. In the morning, we will cover basic ideas in linear algebra (matrix-vector manipulations, norms, subspaces and solving matrix-vector systems). In the afternoon, we will discuss more advanced concepts including QR decompositions and the SVD.

**About the Instructor:** Professor Margot Gerritsen is Senior Associate Dean for Educational Affairs in the School of Earth, Energy and Environmental Sciences and Co-Director of Women in Data Science (WiDS). From 2010 to 2018, she served as the Director of ICME. She received her Ph.D. in Scientific Computing and Computational Mathematics at Stanford in 1997. After five years as faculty member at the University of Auckland, she returned to Stanford in 2001. Her primary appointment is in Energy Resources Engineering. Margot specializes in computational modeling of fluid flow processes, with emphasis on reservoir simulation. She teaches several of the ICME core and service courses in numerical analysis and linear algebra, as well as courses in renewable energy and reservoir simulation.

**Introduction to Statistics**

Statistics is the science of learning from data. This workshop will help you to develop the skills you need to analyze data and to communicate your findings. There won't be many formulas in the workshop; rather, we will develop the key ideas of statistical thinking that are essential for learning from data.

We will discuss the main tools for descriptive statistics which are essential for exploring data, with an emphasis on visualizing information. We will explain the important ideas about
sampling and conducting experiments. Then we will look over some important rules of probability and discuss normal approximation and the central limit theorem. We will show you the important concepts and pitfalls of regression and how to do inference with confidence intervals and tests of hypotheses. You will learn how to analyze categorical data and discuss one-way analysis of variance. Finally, we will look at reproducibility, data snooping and the multiple testing fallacy, and how to account for multiple comparisons. These issues have become particularly important in the era of big data.

Broadly, there are three main reasons why statistical literacy is essential in data science: First, it provides the skills to assess whether the data are sufficient to answer the questions at hand. Second, it establishes a rigorous framework for quantifying uncertainty. And finally, it provides techniques for effectively communicating the findings of your analyses. This workshop equips you with the important tools in all of these areas. It is the statistical foundation on which the recent exciting advances in machine learning are built.

**About the Instructor:** Professor Guenther Walther studied mathematics, economics, and computer science at the University of Karlsruhe in Germany and received his Ph.D. in Statistics from UC Berkeley in 1994. His research has focused on statistical methodology for detection problems, shape-restricted inference, and mixture analysis, and on statistical problems in astrophysics and in flow cytometry.

He received a Terman fellowship, a NSF CAREER award, and the Distinguished Teaching Award of the Dean of Humanities and Sciences at Stanford. He has served on the editorial boards of the Journal of Computational and Graphical Statistics, the Journal of the Royal Statistical Society, the Annals of Statistics, the Annals of Applied Statistics, and Statistical Science. He was program co-chair of the 2006 Annual Meeting of the Institute of Mathematical Statistics and served on the executive committee of IMS from 1998 to 2012.

**Introduction to Programming in R**

This workshop is recommended for those who want to learn the basics of R programming in statistics, science, or engineering. The goal of this workshop is to familiarize participants with R’s tools for scientific computing and data analysis. Lectures will be interactive with a focus on learning by example, and assignments will be application-driven.

Example topics:
Introduction to Mathematical Optimization

Mathematical optimization underpins many applications in science and engineering, as it provides a set of formal tools to compute the ‘best’ action, design, control, or model from a set of possibilities. In data science and machine learning, mathematical optimization is the engine of model fitting. This workshop will provide an overview of the key elements of this topic (unconstrained, constrained, convex optimization, optimization for model fitting), and will have a practical focus, with participants formulating and solving optimization problems early and often using standard modeling languages and solvers. By introducing common models from machine learning and other fields, this workshop aims to make participants comfortable with optimization modeling so that they may use it for rapid prototyping and experimentation in their own work. Students should be comfortable with linear algebra, differential multivariable calculus, and basic probability and statistics. Experience with Python will be helpful, but not required.

Topics to be discussed in this workshop include:

- Formulating optimization problems
- Fundamentals of constrained and unconstrained optimization
- Convex optimization
- Optimization methods for model fitting in machine learning
- Optimization in Python using SciPy and CVXPY
- In-depth Jupyter Notebook examples from machine learning, statistics, and other fields
**About the Instructor:** Kevin Carlberg is a Distinguished Member of Technical Staff at Sandia National Laboratories in Livermore, CA. He received his PhD in Aeronautics and Astronautics from Stanford University in 2011 with a PhD minor in Computational and Mathematical Engineering, and was a President Harry S Truman Postdoctoral Fellow at Sandia from 2011 to 2014. Kevin leads a research group of PhD students, postdocs, and technical staff whose work combines concepts from machine learning, computational physics, and high-performance computing to drastically reduce the cost of simulating nonlinear dynamical systems at extreme scale.

**Introduction to Python**

Introduction to Python will focus on scientific computing and data science. We will cover basic language concepts as well as specific tools for linear algebra and data science.

More precisely, the class will be divided into three parts:
- Python basics
- Linear algebra and scientific computing, with numpy & spicy
- Data Science, with pandas

The class is designed for people with no experience in Python, but who would like to start using it on some scientific application. However, some (past) programming experience with at least another language is strongly recommended. In particular, students are expected to be comfortable with basic notions such as variables, if-else, loops, lists, etc.

We will introduce each topic enough so that you can quickly start using Python for your own problems. The workshop will be interactive, with many examples (that the participants can play with during the session).

**About the Instructor:** Leopold Cambier is a fourth-year PhD student in ICME. He currently works with Professor Eric Darve on fast algorithms for linear systems, typically arising from the discretization of PDE’s. He also did two summer internships (summers 2016 and 2017) at NVidia, working on cuDNN and cuBLAS. Leopold obtained his Bachelor’s degree in Engineering and a Master’s degree in Mathematical Engineering from Université catholique de Louvain in Belgium. More info at [https://web.stanford.edu/~lcambier/](https://web.stanford.edu/~lcambier/)
**Intro to Machine Learning**

This workshop presents the basics behind understanding and using modern machine learning algorithms. We will discuss a framework for reasoning about when to apply various machine learning techniques, emphasizing questions of over-fitting/under-fitting, interpretability, supervised/unsupervised methods, and handling of missing data. The principles behind various algorithms—the why and how of using them—will be discussed, while some mathematical detail underlying the algorithms—including proofs—will not be discussed. Unsupervised machine learning algorithms presented will include k-means clustering, principal component analysis (PCA), multidimensional scaling (MDS), tSNE, and independent component analysis (ICA). Supervised machine learning algorithms presented will include support vector machines (SVM), lasso, elastic net, classification and regression trees (CART), boosting, bagging, and random forests. Imputation, regularization, and cross-validation concepts will also be covered. The R programming language will be used for occasional examples, though participants need not have prior exposure to R.

Prerequisite: undergraduate-level linear algebra and statistics; basic programming experience (R/Matlab/Python).

**About the Instructor:** Alexander Ioannidis earned his Ph.D. in Computational and Mathematical Engineering and Masters in Management Science and Engineering both at Stanford University. He is a postdoctoral fellow working on developing novel machine learning techniques for medical and genomic applications together with Carlos Bustamante, Chair of the Department of Biomedical Data Science at Stanford Medical School. Prior to Stanford he earned his bachelors in Chemistry and Physics from Harvard and a M.Phil from the University of Cambridge. He conducted research for several years on novel superconducting and quantum computing architectures at Northrop Grumman’s Advanced Technologies research center. In his free time he enjoys sailing.

**Introduction to Deep Learning**

Deep Learning is a rapidly expanding field with new applications found everyday. In this workshop we will cover the fundamentals of deep learning for the beginner. We will introduce the math behind training deep learning models: the backpropagation algorithm. Building conceptual understanding of the fundamentals of deep learning will be the focus of the first part of the workshop. We will then cover some of the popular architectures used in deep learning, such as Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), LSTMs, autoencoders and GANs. There will be a hands-on computing tutorial using Jupyter notebooks to build a basic image classification model via transfer learning. By the
end of the workshop, participants will have a firm understanding of the basic terminology and jargon of deep learning and will be prepared to dive into the plethora of online resources and literature available for each specific application area.

Prerequisites: Familiarity of basic concepts from linear algebra, such as vectors and matrices, as well as calculus concepts, such as differentiation. Familiarity with the python programming language and an ability to use Jupyter notebooks will be helpful for the hands-on sessions.

About the Instructor: Alexander Ioannidis earned his Ph.D. in Computational and Mathematical Engineering and Masters in Management Science and Engineering both at Stanford University. He is a postdoctoral fellow working on developing novel machine learning techniques for medical and genomic applications together with Carlos Bustamante, Chair of the Department of Biomedical Data Science at Stanford Medical School. Prior to Stanford he earned his bachelors in Chemistry and Physics from Harvard and a M.Phil from the University of Cambridge. He conducted research for several years on novel superconducting and quantum computing architectures at Northrop Grumman's Advanced Technologies research center. In his free time he enjoys sailing.

About the Instructor: Anjan Dwaraknath is 5th year PhD student in the Institute for Computational and Mathematical Engineering. He is currently pursuing his research under Prof. Gunnar Carlsson on the potential of combining Deep Learning algorithms with Topological Data Analysis. Prior to Stanford he worked on derivative pricing models as an analyst in Goldman Sachs. He obtained his undergraduate degree in Engineering Physics from the Indian Institute of Technology, Madras.

Deep Learning for Natural Language Processing

This workshop will focus on practical applications and considerations of applying deep learning to Natural language processing (NLP). We will start by drawing inspiration from more traditional NLP approaches, and show how many modern deep learning-based algorithms have deep roots in traditional techniques, while showing how deep learning has enabled new improvements. This workshop will heavily focus on student’s understanding of problem templates in applied natural language processing, and about identifying application patterns.

We will have a practical focus, targeting algorithms, and problem templates which are able to be deployed and used today. We will cover the different components that go into deep
learning systems, including word vector representations (word2vec, GloVe), contextual representations (ELMo, BERT), and general model components such as convolutional layers, Transformers, and others. We will also cover introductory material in applications such as classification, intent understanding, and others.

We will be using the Keras library for a practical session where we will implement select models, and thus, some experience with both Python and Machine Learning is required. We recommend taking the Machine Learning and Deep Learning ICME workshops for a better understanding of the material included in this NLP workshop.

**About the Instructor:** Luke de Oliveira currently leads engineering for the AI platform team at Twilio as part of Twilio AI. Previously, Luke was Founder & CEO of Vai Technologies, which was acquired by Twilio in 2018. Luke is also a visiting researcher at LBNL where he works on generative modeling in the natural sciences, and is a published author at the intersection of Deep Learning and Physics. In addition, Luke serves as an advisor to a variety of technology startups. Luke earned his M.S. in Computational and Mathematical Engineering from Stanford University and his B.S. in Applied Mathematics from Yale University.

---

**Introduction to High Performance Computing**

In the past 50 years, supercomputers have achieved what was once considered only possible in Sci-Fi movies. The key to the tremendous success of supercomputers has been a combination of outstanding architectures plus software that uses all the available resources and makes parallelization possible. This secret sauce has led to different implementations across fields. A mechanical engineer would use MPI and OpenMP to have a balance between computations and memory load to deal with millions of nodes in physical simulations, whereas a data scientist would use MapReduce and Spark to have an adaptable and resilient algorithm for the challenges of big data. This workshop explores the key features of these two approaches, explaining their underground philosophy and how they use the architecture. The final goal is to give the student a taste of the different programming paradigms and the tools to decide which is the best approach.

**About the Instructor:** Cindy Orozco is a 4th year PhD student at ICME working with Professor Lexing Ying in matrix and tensor optimization applied to computer vision and quantum physics. She did her undergrad in Mathematics and Civil Engineering in Universidad de Los Andes, Colombia, and a masters in Applied Mathematics at King Abdullah University of
Science and Technology, Saudi Arabia. Her sweet spot is the intersection between simulation, physical sciences and pure math, producing algorithms to model physical phenomena based on abstract mathematical structures. Her interest in High Performance Computing comes after an internship in Hewlett Packard Labs estimating the performance of future supercomputers and recognizing the gap between the optimal usage of these machines and the typical practices among researchers.

Interactive Data Visualization in D3

This workshop introduces D3, a powerful tool for creating interactive data visualizations on the web. This workshop is geared toward people in the industry and academia who want to better communicate their projects and research through visualizations on the web. Attendees will be introduced to the basics of web design. They will learn how to input CSV data on D3 and use it to create simple interactions with graphical elements on the screen. At the end, they will be familiar with some of the standard D3 libraries. These tools will be the basis for more complex visualizations, like those that appear on d3js.org.

This workshop will cover the basics of D3: inputting data, creating scales and axes, and adding transitions and interactivity, as well as some of the most used libraries: stack, cluster and force layouts. We will make use of HTML, CSS and JavaScript, which will be introduced at the beginning. A basic programming background in a language like Python, JavaScript, C++ or R is highly recommended.

About the Instructor: Sergio Camelo Gomez is a 3rd Year PhD Student at Stanford ICME, where he works on creating technological tools for the sustainable sourcing of food in emerging markets, leveraging on optimization and machine learning. At Stanford, he teaches ICME’s Data Visualization class (CME151A). Sergio also spends his summers working with the Data Science team at Airbnb or the Locations and Place Data teams at Facebook NY. Originally from Colombia, Sergio graduated with a Bachelor's and Master's in Mathematics from Universidad de los Andes, and worked for two years as a Data Scientist in Quantil Matematicas Aplicadas. To learn more about Sergio, visit his webpage: http://web.stanford.edu/~camelo/