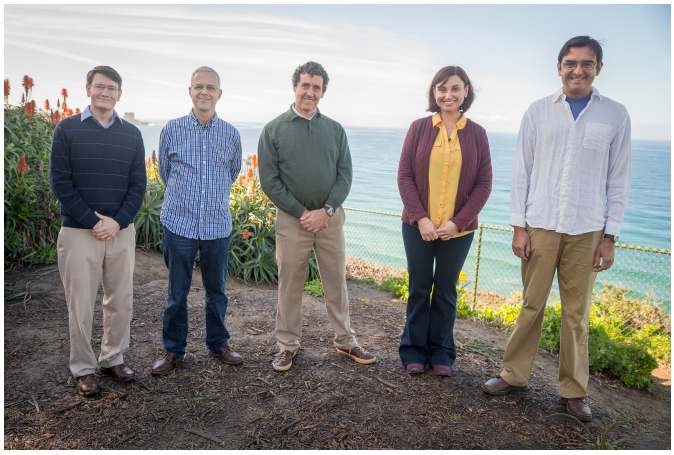
UC San Diego
UC San Diego News Center

February 28, 2017 | By Doug Ramsey

UC San Diego Launches Online Courses with edX to Advance Careers in Data Science



Instructors for the Data Science courses in the edX MicroMasters program (l-r): CSE's Leo Porter, ECE's Alon Orlitsky, CSE's Yoav Freund; SDSC's Ilkay Altintas; and CSE's Sanjoy Dasgupta. Altintas is also a lecturer in CSE, and Orlitsky has a secondary appointment in CSE. [Photo by Alex Matthews/Qualcomm Institute]

The University of California San Diego is launching an online series of courses in Data Science that will roll out through edX, the nonprofit online learning destination founded by Harvard and MIT. The four courses will be part of edX's MicroMasters® program, offering a credential for career advancement after successful completion of all four courses and the potential to accelerate a Master's degree.

"The MicroMasters program aims to bridge the gap between higher education and the workplace," said Karen Flammer, interim director of UC San Diego's Office of Online and Technology Enhanced Education. "The online program offers a modular credential with a pathway to credit and it is designed for learners looking for in-demand knowledge to advance their careers or to follow a path to an accelerated on-campus program."

The first 19 MicroMasters programs were launched in September 2016. Today edX announced a second wave of 16 programs, primarily in data science and computer science, created by UC San Diego and 11 other edX university partners. Of those, eight partners are offering MicroMasters programs for the first time, including UC San Diego, University of Pennsylvania, the University of Maryland system, University of British Columbia and RWTH Aachen (Germany).

UC San Diego's Data Science program grew out of a call for MicroMasters proposals fielded by former CSE professor Beth Simon (now in Education Studies). Simultaneously, edX CEO Anant Agarwal visited UC San Diego and met with CSE professor Yoav Freund, who took the lead on

coordinating faculty participation in the program along with San Diego Supercomputer Center (SDSC) chief data science officer Ilkay Altintas. The courses are based on the Master of Advanced Studies in Data Science and Engineering curriculum launched by CSE and SDSC in Fall 2014 (with Freund and Altintas among the MAS program's current co-directors).

"It is no surprise that we were asked to develop a series of courses in Data Science," said Dean Tullsen, chair of the Computer Science and Engineering (CSE) department. "UC San Diego has deep expertise in this area, and we recently launched an executive Master of Advanced Studies program In Data Science. Learners around the world will be able to learn critical skills needed to become a data scientist, taught by top UC San Diego instructors. They will learn how to gain new insights from big data by asking the right questions, manipulating data sets and visualizing their findings in compelling ways."

To prove that they've completed all four courses in the Data Science program, learners pay $350 per course to get a Verified Certificate. The four certificates become credentials of "real job relevancy" that edX believes will be viewed positively by employers and hiring executives as evidence of the job applicant's advanced knowledge of data science.

The combined certificates can also count toward 25 percent of the 400 credits required for graduation with a Master of Predictive Analytics degree from Australia's Curtin University, contingent on the learner applying and being accepted into Curtin's Master's program. The edX program anticipates that more universities will eventually recognize the value of the MicroMasters and provide course credit toward full-fledged Master's degree programs on other campuses in future.

The UC San Diego MicroMasters program encompasses two sides of data science learning: the mathematical and the applied. Mathematical courses cover probability, statistics, and machine learning. The applied courses cover the use of specific toolkits and languages to delve into real-world data. They include languages such as Python, Numpy, Matplotlib, pandas and Scipy, as well as the Jupyter notebook environment and Apache Spark (e.g., to analyze data that does not fit within the memory of a single computer).

"Learners who take these courses will come away knowing how to collect, clean and analyze big data using popular, open-source software," said CSE's Tullsen. "They should also be able to perform large-scale data analysis and present their findings in a convincing, visual way."

The four courses to be introduced over the next year include (listed by starting date):

- ***June 1, 2017****:* Python for Data Science (200x) is being developed by SDSC chief data science officer and CSE lecturer Ilkay Altintas and CSE teaching professor Leo Porter. This course introduces learners to a collection of powerful, open-source tools to analyze data and conduct data science. “What makes this course particularly exciting is that you’ll learn these tools while solving compelling data science problems,” explained CSE’s Porter. “The course teaches students how to find answers within large datasets by using Python tools to import data, explore it, analyze it, learn from it, visualize it, and ultimately generate easily-sharable reports.”Added SDSC’s Altintas: “By learning these skills, learners can also become members of a worldwide community focused on building data science tools, exploring public datasets, and discussing evidence-based findings.”
- ***September 28, 2017***: Statistics and Probability in Data Science using Python (210x) will teach statistical and probabilistic approaches to understanding and gaining insights from data. After all, said instructor Alon Orlitsky, a professor in both CSE and Electrical and Computer Engineering (ECE), “the job of a data scientist is to glean knowledge from complex and noisy datasets; reasoning about uncertainty is inherent in the analysis of noisy data.” Concepts covered in 210x will include random variables, dependence, correlation, regression, PCA, entropy and MDL, and hands-on experience will involve applying theory to actual data using Jupyter notebooks. Orlitsky also directs the new Data Science Postdoctoral Fellows program, a partnership of CSE, ECE and the Qualcomm Institute.
- ***January 3, 2018***: Machine Learning for Data Science (220x) will explore machine learning’s role in data-driven modeling, prediction and decision-making. According to instructor and CSE professor Sanjoy Dasgupta, learners will come away from the course with a variety of supervised and unsupervised learning algorithms and the theory behind them. “Using real-world case studies,” noted Dasgupta, “they will learn how to classify images, identify salient topics in a corpus of documents, partition people according to personality profiles, and automatically capture the semantic structure of words and use it to categorize documents.” Armed with this knowledge, learners should be able to analyze many different types of data and build descriptive as well as predictive models.
- ***April 1, 2018****:* Big Data Analytics Using Spark (230x) will focus on analyzing very large datasets (Big Data) that typically require using a cluster of tens, hundreds or even thousands of computer processors. According to CSE professor and instructor Yoav Freund, “learners will come away knowing how to analyze massively large datasets using cloud computation. They will understand the source of bottlenecks in such computation and how to minimize them. Learning will be grounded on hands-on experience using Apache-Spark on their own computers and in the cloud.”

“We are honored to work with UC San Diego to launch a MicroMasters program in Data Science,” said MIT professor Anant Agarwal, CEO at edX. “It’s an exciting step toward furthering our shared mission to expand access to high-quality education. Signaling the next level of innovation in learning, MicroMasters programs are designed to meet the needs of both universities and employers, by providing learners with the in-demand knowledge and skills needed for success in today’s rapidly-evolving and tech-driven world.”

The Data Science program is one of four new edX MicroMasters in Data Science and Analytics. The others cover Business Analytics (Columbia University), Big Data (University of Adelaide), and Analytics (Georgia Tech). Strong demand from learners is expected, and for good reason. ‘Data scientist’ was rated the best job in the U.S. for 2017 according to a recent survey, with a median base salary of $110,000. There are also 215,000 open job positions in data science nationwide (as well as 45,000 other job openings seeking statisticians and business intelligence personnel).

MEDIA CONTACT

Doug Ramsey, , dramsey@ucsd.edu

UC San Diego’s Studio Ten 300 offers radio and television connections for media interviews with our faculty, which can be coordinated via studio@ucsd.edu. To connect with a UC San Diego faculty expert on relevant issues and trending news stories, visit https://ucsdnews.ucsd.edu/media-resources/faculty-experts.