Welcome to DSE230x

🔖 Bookmark this page

Welcome to Big Data Analytics using Spark!

We are delighted to welcome you to the fourth course in the MicroMasters in Data Science: Big Data Analytics using Spark.This course teaches you how to perform statistical analysis of very large datasets that do not fit on a single computer. You will learn some of the most popular tools for performing this type of analysis: apache spark, XGBoost and TensorFlow. You will learn how to use these tools through Jupyter Notebooks and experience the power of combining narrative, code and graphics to create convincing analytical documents.

Course Staff

Instructor

Yoav Freund, Professor of Electrical and Computer Engineering, UC San Diego

Learning objectives

This course has two main goals: The first is introduction to using large scale data analysis frameworks (Spark, XGBoost and TensorFlow). This includes the underlying computer architecture and the programming abstractions. The second is to combine methods from statistics and machine learning to perform large scale analysis, identify statistically significant pattern and visualize statistical summaries.

Course Outline

This is a ten-week course.

Topics

- Memory Hierarchy, latency vs. throughput.

- Spark Basics

- Dataframes and SQL

- PCA and weather analysis

- K-means and intrinsic dimensions

- Decision trees, boosting, and random forests

- Neural Networks and TensorFlow

总结：

Python notebooks

Jupyter notebooks are the foundation of this course. Most of the videos are overviews of a notebook. It is recommended that you follow along in the notebook while observing the video. The notebooks contain explanations, code and figures, as well as interactive widgets and pointers to additional material. It is recommended that questions to the forum refer to locations in the notebooks.

Time and grading policies for weekly assignments
Each assignment is due five weeks after it is given.
The worst two comprehension quiz scores will be dropped and the worst two programming assignment scores will be dropped. This means, for instance, that it is possible to obtain a full score while skipping any two of the programming assignments and any two of the comprehension quizzes.

Effort
The weekly effort for the course is intended to be roughly 10 - 20 hours.

Pace and deadlines
The course is instructor-paced. Every two weeks, two week's worth of the relevant material (videos and assignments) will be released, and will remain online until the end of the course. We encourage learners to keep current with the videos and assignments; however, as described above, there is a five week window for submitting each assignment.

总结:

Honor Code and Discussion Forum Etiquette

🔖 Bookmark this page

Honor Code

Beyond learning this important material, we hope that you will take the course seriously and respect your fellow classmates. Please read and abide by the edX honor code pledge. If you have any questions about what is permitted or not please let us know.

We value your feedback

This online course is new for us and for UC San Diego. We tried hard to provide you with high-quality education, but there will certainly be room for improvement. If there is anything, large or small, we can improve, please do let us know. We greatly appreciate your feedback and will take it very seriously.

Thank you!

Thank you very much for taking the course. We hope you will find it educational and useful.

Remember, when you are posting in discussion forums we expect you to observe common rules of etiquette (see a visual here). However, here's a few basic categories of etiquette to keep in mind.

- Respect the challenges of written communication (e.g., don't use all CAPS, avoid abbreviations, be careful with jokes, don't say things in writing you wouldn't say in person (e.g., call someone stupid), avoid posting when you are angry)

- Make your post valuable for others (e.g., keep it as short as possible, use the simplest language you can (not everyone speaks English as a first language), stay on topic, be accurate (or state that this is your opinion), before posting a question - check to see if someone else has already asked it, if many people respond to your post - provide a summary).

- Be respectful of others (e.g. respect the opinion of your classmates, be open to others with differing viewpoints, respect the differences in culture and experience among learners in the course).

Optional: Feedback

🔖 Bookmark this page

We care about your learning!

This online course is new for us and for UC San Diego. We believe we are providing you with world-class materials and resources to learn data science or we wouldn't have created the course. But, being experienced educators, we know things often go wrong in new courses. We value your patience as we resolve issues that come up and we promise we'll do our best to resolve them both fairly and quickly.

If you encounter problems, please post on the forums and fill in the surveys which occur periodically in the course. We take your feedback seriously and will use it to improve this, and future iterations, of this course.

This is all to say: Thank you for your patience and feedback!

总结:

What Are Discussions For?

We will use discussions to help you further think about and analyze the material taught in the lectures. Topics will vary depending on the week and you can post anonymously if you choose. You can use these discussions to express your ideas and interact with other learners taking this course! Participation in discussion forums is optional, but highly encouraged. It is to the benefit of all students the we have an active discussion board.

If you are a Verified Certificate learner, you also have access to a private Q&A discussion forum where you can interact with each other and ask questions of our teaching team. Although both forums will be monitored by our team of instructors, priority will be given to the Verified Certificate learner discussion.

Don't have a certificate? Click the green "Upgrade" button on the right , and then return to this section and you will complete a special introduction discussion with other Verified Certificate learners. If you aren't interested that's fine, too! Continue on and tell us about yourself and your goals for the course!

Installing Software for the course

详见下载下来的网页，里面有步骤说明

总结：