



14 – 18 AUGUST 2017 / UNIVERSITY OF COPENHAGEN

BIG DATA ANALYSIS

— TOOLS AND METHODS

Big Data is omnipresent from industries to government and is frequently considered a completely new approach to problem solving. While the possibilities are often exaggerated, Big Data does indeed introduce new opportunities and challenges. The ability to analyse and combine large data from different sources has obvious applications, nonetheless, the lack of quality in the data combined with a high variance means that conventional analysis often fails.

This course will bring you to the forefront of the newest tools and methods based on cutting edge research and experience.

- Statistical methods: Robust methods for very large datasets and data with very large variance and outliers
- Finding patterns and outliers in Big Data: Which methods can be used to identify sparse patterns in very large datasets, and how to identify data that does not follow the overall pattern for a dataset?
- Collecting data from instruments and devices: How to collect, store, and analyse data from a multitude of sources that produce data (i.e. Internet-of-Things)
- Systems for Big Data Analysis: Common systems for BDA; Hadoop, PyDisco, etc., and hardware systems design for efficient BDA

WHAT YOU WILL LEARN

By completing the course you will be able to set up basic Big Data Analysis end-to-end; from retrieving and cleaning the data, to establishing the information level and extracting patterns and finding outliers and to curate the necessary data.

You will get acquainted with a number of advanced tools like: Data cleaning, statistical methods for very large datasets, data stream analysis and finding patterns and outliers in Big Data, collecting data from instruments and devices (i.e. internet of things) and hardware systems design for efficient BDA.

COURSE CONTENT

We will use a few structured datasets consistently throughout the course, which illustrate the commerce and will be used to demonstrate the different steps in Big Data Analysis.

Core elements:

- Data cleaning: Detecting and correcting (or removing) corrupt or inaccurate records

Tools/methods introduced:

- Selected machine learning algorithms for large-scale data
- Random forests and large-scale exact nearest neighbour search
- Data curation: How to select data for long time curation, systems, techniques and standards for data curation?

We will be working with several programming tools, however all techniques that are covered are easily implemented with all standard data-analysis languages; Python, R, etc.

**”Very interesting, well organized, great teachers. I learned a lot!
Thank you for a great course!”**

*Agnethe Larsen, Head of Section,
Danish Business Authority*



”I found the course to be of very high quality theoretically.
The theoretical concepts covered were very detailed & relevant.”

Anup Singh, General Manager, Maersk Group, 2016.

PARTICIPANTS

The course is strictly focused on Big Data Analysis, thus a background in statistics and/or conventional data analysis is assumed. This course assumes an education at least at a Bachelor level and/or several years of data analysis experience.

COURSE DATES

5 days, 14 – 18 August 2017, 9:00 – 16:30 at the University of Copenhagen, Frederiksberg Campus.

COURSE DIRECTOR

Troels C. Petersen, Associate Professor, Particle Physics, Niels Bohr Institute, University of Copenhagen

OTHER COURSE TEACHERS

Brian Vinter, eScience, Niels Bohr Institute, University of Copenhagen

Joachim Mathiesen, Associate Professor, Biocomplexity, Niels Bohr Institute, University of Copenhagen

COURSE FEE

EUR 2,600/DKK 19,000 excl. Danish VAT. Fee includes teaching, course materials, all meals during the course.



”I really liked the atmosphere at the campus – this is fantastic – and it worked out to give an extra dimension on the course”

Former participant on a Copenhagen Summer University course

FOR MORE INFORMATION AND REGISTRATION:
copenhagensummeruniversity.ku.dk

OTHER RELATED COURSES:

21 - 25 August 2017, University of Copenhagen
Deep Learning for Data Analytics

14 - 18 August 2017, University of Copenhagen
Data Science with R

FOR MORE INFORMATION, SEE:
copenhagensummeruniversity.ku.dk

