Module introduction

DSTA

Data Science: Techniques and Applications (DSTA)

A new[ish] module

- Designed for MSc Data Science students
- Contents carefully dovetail with existing MSc DS modules

. . .

we are in this together

Main topics

- Python modules for DS, `Numpy`, `Pandas`, `Matplotlib` and `Scikit-learn`.
- The geometric view of multidimensional data (Lin. Algebra)
- The data-as-network view (graph algo.)
- the text-as-data view (Information Theory)

. . .

- a motivating DS problem: sports ranking
- selected topics: matrix *slicing:* finding latent dimensions to datasets.
- [discontinued: handling exponentially-distributed data]

Important aspects

The inevitable *overfitting* to MSc DS may make this module less appealing to MSc ACT students than the title suggests

. . .

Machine Learning topics are selected so as not to overlap the essential Machine Learning and Applied ML modules.

The table of contents is varied and may appear syncretic wrt. traditional, textbook-based modules..

. . .

This is a *final* module in charge of synthetising a large, fast-moving area.

. . .

so, expect a more seminarial (than textbook) approach.

Organization

In lab

- Alessandro Provetti
- Abul Hasan
- Paschalis Lagias
- Alberto Matuozzo

Online

Standard Moodle/Teams provision

Updated at-a-glance programme on the satellite site:

http://www.dcs.bbk.ac.uk/~ale/dsta

https://ale66.github.io/dsta/

All-in-one downloadable materials:

https://github.com/ale66/dsta/

Please check your email and MyBirkbeck calendar for (unlikely) time/place amendments.

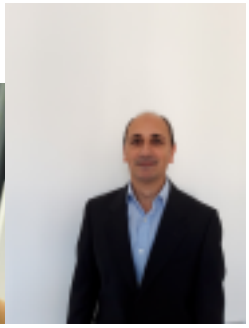
Please see your

- PG admins, Matt Spence and Yeti Wan, for support in navigating school: cs-pg@bbk.ac.uk


also see your

- TAs, Abul, Paschalis and Alberto, for support with lab experience





and finally see your

- module coordinator, Alessandro, about the study materials and general *'meaning of life'* questions

Too many e-mails

Your feedback and questions are always welcome.

Help avoiding inbox overflow by please contacting us via MS Teams or Moodle class forum

Assessment

Past Marking trends

DSTA vs. ML in 2019:

Coursework

For the first time, coursework will take the form of an in-class, multiple-choice test via Codio/ Moodle.

On Week 6 (May 28th) at 20:00 for about 30/40 minutes.

10 questions on the 10 foundational units: from 2.b to 4.c (see the online calendar).

In view of the novelty of the approach, marks will be awarded generously, but always at the discretion of the School's Exam board.

Final test

For the first time, the final test will take the form of an in-class, multiple-choice test via Codio/ Moodle.

On the final Week 11 (July 2nd) at 19:00 for about 60/80 minutes.

20 questions on the 20 advanced units: from 4.d to 10.c but excluding the notebooks on 7.d, 8.d, 9.d and 10.d (see the online calendar).

In view of the novelty of the approach ...