MSc Urban Spatial Science

Data Science for Spatial Systems Course Outline

Huanfa Chen

Welcome Everyone

- In this course you’re going to learn how to:
  - Applied Machine Learning
  - Casual Inference
  - Applications in spatial/urban domains

- In the following Weeks you’ll be doing:
  - Python Programming | Data Handling | Data Analysis | Machine learning

The Teaching Team

Huanfa Chen

Ollie Ballinger

Jennie Williams

Xinglei Wang

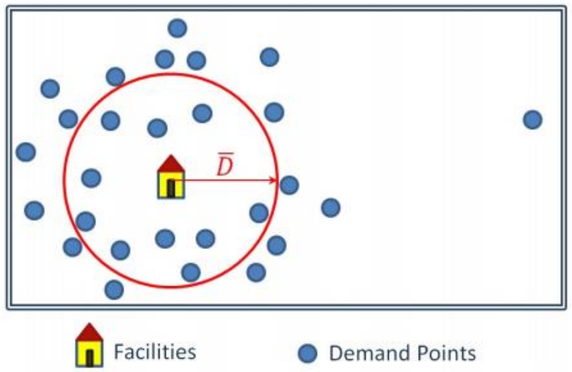
Francesco Terenzi

The Teaching Team

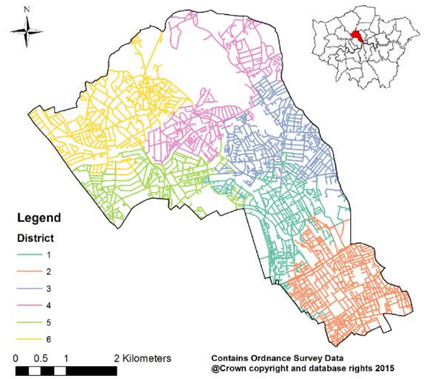
Huanfa Chen

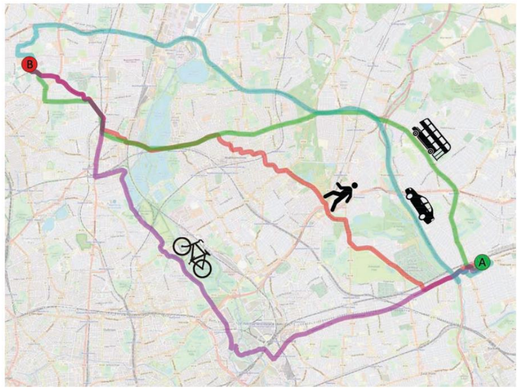
Lecturer in Spatial Data Science

Facility location selection



Police districting problem
（spatial clustering）





Machine learning for transport

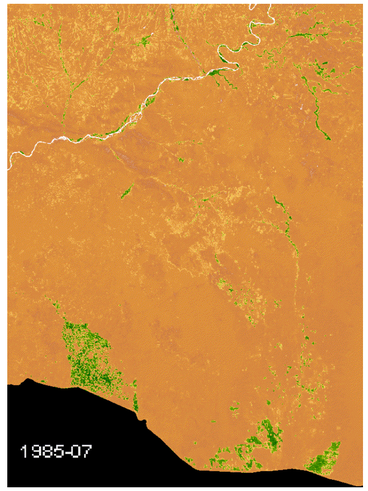
- Travel mode prediction (how people travel)
- Travel purpose estimation (why people travel)

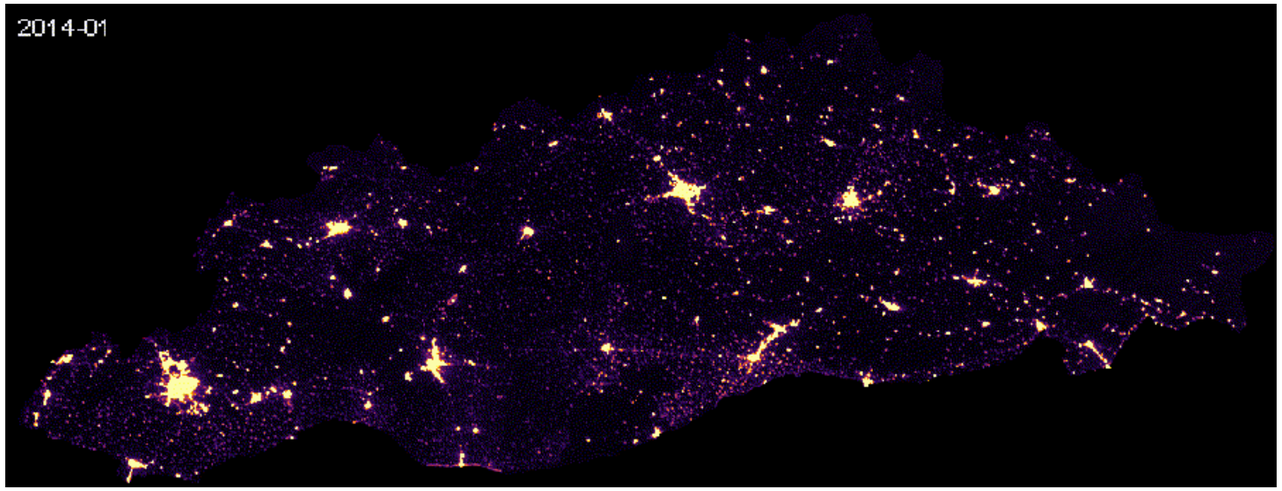
The Teaching Team

Ollie Ballinger

Lecturer in Geocomputation at CASA

- A quantitative social scientist by training
- Leveraging computational methods to understand the drivers of insurgent recruitment in Turkey.
- Using remote sensing and spatial data science to analyse the social world.





The Teaching Team

Jennie Williams

PhD candidate at CASA

- Natural language processing
- Word embeddings
- Clustering (t-SNE, UMAP)
- Current work: the application of data science techniques to the textual metadata of 500K PhD theses in EThOS

Contact: jennie.williams.20@ucl.ac.uk

The Teaching Team

Xinglei Wang

PhD student in SpaceTimeLab at Dept of Civil, Environmental and Geomatic Engineering

- Mobile phone location data mining
- Analysing human travel behaviour and mobility patterns.

Contact: xinglei.wang.21@ucl.ac.uk

The Teaching Team

Francesco Terenzi

Francesco is a PhD student in Ben Werner’s lab at the Barts Cancer Institute in London (Queen Mary University).

His interests lie in applying and developing computational methods to solve complex biological problems.

During his PhD, he combines mathematical modelling with Bayesian inference to understand cancer evolution.

Contact: f.terenzi@ucl.ac.uk

Moodle and Slack

Where are we going to be

- Moodle enrol key for CASA0006: **CASA2022**
- Slack: Join the Slack space at casa-students-2022.slack.com and the channel of **#0006-data-science-spatial-systems**.
- Please log into Slack using your UCL email address.

Week by week

What's going to happen when

- 10-week module
  - 15 credits course
- 2-hour lectures + 2-hour workshops
  - Workshop: quiz + Python notebook
  - This week workshop will break you in gently so may not take 2 hours
- Individual Assessment
  - Reproducible Python notebooks
- Keep checking Moodle and Slack
  - Heavy forum use to help you all work together

Lectures & Workshops

Where are we going to be

	Venue	**Time**	**Recorded?**	**Tap-in attendance required?**
Lecture	Birkbeck Gordon Sq (43) B04	Wed 2-4pm	Probably – will try in Week 1	Yes
Workshop	Bedford Way (26) LG04	Friday 1-3pm	No	Yes

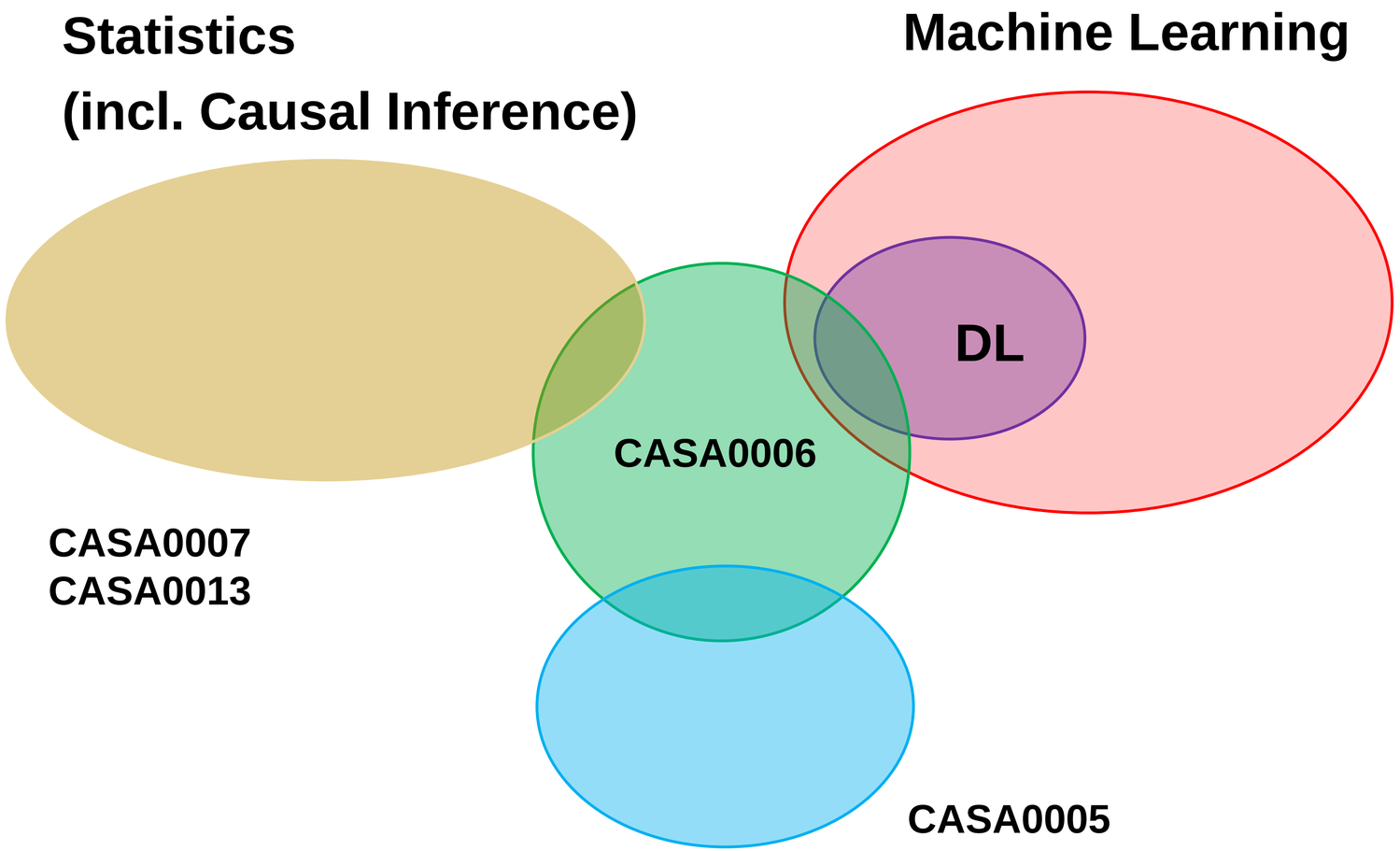
Please inform CASA-teaching@ucl.ac.uk if you can't attend a lecture or workshop.

Check the Moodle for more details.

CASA0006

1. Introduction to Module
2. Supervised Machine Learning
3. Tree-based Methods
4. Artificial Neural Networks
5. Analysis Workflow
6. Panel Regression
7. Difference in Difference
8. Regression Discontinuity
9. Dimensionality Reduction
10. Spatial Clustering

Why are we teaching this?



- Building on techniques from CASA0007/CASA0013/CASA0005
- Extracting knowledge and insights from spatial data
- Tell a story of your data

Pre-requisites

- Working knowledge of Python, Jupyter, Python notebooks, Markdown, Docker, Command line in Linux/Unix based OS, github, file formats
- A general understanding of statistics, linear regression
- A general understanding of GIS files, projection, UK census units
- No previous knowledge with machine learning is required

Course philosophy

- This is a practical and hands-on course.
- We will cover basic concepts and background theory relating to ML and causal inference,
  but not in great mathematical depth or rigor.
- Mainly focus on implementing and running ML/CI algorithms with examples in a spatial context, discussing the implications
- Using the variables and algorithms in the right way

DSSS Assessment

Technical Analysis and Visualisation Report

Individual Assessment

Module Assessment

- 100% mark from coursework
  - 100% - From Analysis / Python Notebook
- Individual Assessment
  - Working by yourself
- Max 2,000 word report with in-built analysis in form of Python Notebook
  - Deadline: 24th April 2023 at 5pm

Examples of Work

What we want you to strive to create

About this document

This is the assignment for the UCL CASA module Data Science for Spatial Systems (CASA0006)
This is a self contained Jupyter notebook with structure introduction, discussion and embeded code and bibliography.
Code is hosted here
Data is availble here

Multi-level segmentation of electricity consumption data:

the case of the Netherlands

This paper explores data from Dutch electricity providers and establishes provider-level traits based on clustering of cities and streets they supply to.
It highlights the varying customer and area profiles that each serves and serves to identify potential priorities

https://github.com/antoniosfiala/Electricity_clustering

7.0 | Discussion and conclusion

In conclusion, this paper has carried clustering of electricity usage data across the Netherlands with the aim of answering the following research questions with each addressed below:

```
RQ1 | What segments of energy use can be observed in the Dutch market?
RQ2 | Does the data suggest a relationship between the supplier and the segment it supplies?
RQ3 | How do clusters vary across geographical scales?
```

RQ3: Some variation can be observed across the levels, with clustering at city level yielding three clusters whereas street and postcode level analysis suggests four is the appropriate number.

RQ1: Along the measures available, categorise were observed that were chiefly defined by variation in smart-meter and low-tariff use that appear to correspond to the supplier they fall under.

RQ2: Building on RQ1 it appears that there may be a relationship between the supplier and the extent to which smart meters are used.

The implication of this is that both at institutional, governmental level and supplier level, there may be room to improve targetting to transition the remaining areas to smart metering.

https://github.com/antoniosfiala/Electricity_clustering

General Recap

- Going to be challenging but worthwhile
  - You'll learn real-world analytical skills and how to pull components together to make data shine.
- Learn how to work on a data science project
  - Skills that will be really useful in the coming years
- Present your work
  - Tell a story with your data