SDS7102: Linear Models and Extensions

Qiang Sun & Eric Moulines

Fall, 2025

E-mail: qiang.sun@mbzuai.ac.ae
Web: **https://sites.google.com/view/qsun**
E-mail: eric.moulines@mbzuai.ac.ae
Web: **https://mbzuai.ac.ae/eric-moulines**
Class Hours: T 12:00-13:20 @ CR3, Th 11:00-12:20 @ CR1
Office Hours: Th 12:30 – 1:30

TA: Ding Bai
E-mail: ding.bai@mbzuai.ac.ae
Lab Room: CR1
Lab Hours: F 9:00-10:50

Course Description

A central goal in statistics is to use data to build models that allow us to make inferences about the underlying data-generating processes or predict future observations. Although real-world problems are often complex, the linear model frequently provides a good approximation to the true data-generating process. Moreover, linear models possess elegant algebraic and geometric properties and often admit explicit formulas, offering deep insights into various aspects of modern machine learning. In our experience, the insights gained from linear models are broadly applicable, with only rare exceptions.

Computing

Python Programming Language.

Reference Textbooks

Lecture notes will be provided on the course website. Our lecture notes closely follow the following book.

- Peng Ding, Linear model and extensions. Chapman & Hall.

Other references include:

- Annette Dobson, Adrian Barnett, An introduction to generalized Linear models.

  - A classic book, focused on GLMs, which covers standard topics in an elementary way. On certain subjects, it offers further insights that may be of interest to students who wish to go deeper.

- Alan Agresti, Foundations of Linear and generalized linear models.
  - A bible of the field, covering all the classics.

Prerequisites/Corequisites

This course assumes that the audience has basic training in linear algebra, probability theory, and statistical inference.

Topics

The followings are potential topics and might change subject to the instructor's view on the progress of the course:

- Introduction
- Multivariate linear regression
- Statistical inferecne
- Model fitting, checking, and misspecfication
- Overfitting and explicit regularization
- Overparameterized models and implicit regularization
- Generalized linear models

Assignment Policy

We will only take .tex and .pdf (along with related materials), python notebook for assignments. Associated and reproducible code, if any, must be attached. Late submissions will NOT be accepted.

If plagiarisms are found, they will be reported. Both (or multiple) copies of the assignments will be given zero grades.

Exam details and dates

- Midterm: TBD @ week 8.
- Final: TBD @ week 16.