PS5842

Advanced Data Science Applications in Finance & Insurance

Overview

Yubo Wang

2023

COLUMBIA UNIVERSITY
IN THE CITY OF NEW YORK

Where to next?

- ,

COLUMBIA UNIVERSITY
IN THE CITY OF NEW YORK

Databases

- Relational database management system
- SQL
- Manage databases in terminal, jupyter and python
- Working with multiple databases and multiple tables simultaneously

Generalized Additive Models

- Nonlinear extensions of SLR

- Splines

- Generalized additive models

Deep Learning

- Feed forward neural networks

- Convolutional neural networks

- Recurrent neural networks

Linear Mixed Models

- Clustered data

- Repeated-measures data

- Longitudinal data

Bayesian Approach

- Linear regression

- Splines

- (Bayesian additive) regression trees

- Posterior sampling
  - Markov Chain Monte Carlo

Interpretive Machine Learning

- Model-Agnostic Methods
  - Partial dependence
    - Feature effect
  - Permutation feature importance
    - Importance as an increase in loss when features are permuted
  - Individual conditional expectation
    - How changing a feature changes the prediction
  - Shapley values
    - Fair attribution of the prediction to individual features
    - From coalitional game theory, how to fairly distribute payout among players

That was