

MIS710 Machine Learning in Business

Lab 3: Linear Regression

Associate Professor Lemai Nguyen



Objective

- **Objective:** to learn and practise linear regression models with scikit-learn
- Download MIS710 Lab3.ipynb and the Housing3 dataset from the unit site in the CloudDeakin system
- Upload it to the MIS710 folder on your Google drive and open MIS710 Lab3.ipynb
- Follow the instructions

Lab 3: Task 1

- To predict medical insurance cost.
- The dataset was slightly modified to allow you deal with missing data. Download the modified data at the unit site.
- Data:
 - area
 - bedrooms
 - bathrooms
 - stories
 - mainroad
 - guestroom
 - basement
 - hotwaterheating
 - airconditioning
 - parking
 - prefarea
 - furnishingstatus
 - price
- Source: <https://www.kaggle.com/datasets/yasserh/housing-prices-dataset>

Lab 3: Task 2 Optional exercise

- To predict medical insurance cost.
- Data:
 - age: age of primary beneficiary
 - sex: insurance contractor gender, female, male
 - bmi: Body mass index, providing an understanding of body, weights that are relatively high or low relative to height,
 - objective index of body weight (kg / m^2) using the ratio of height to weight, ideally 18.5 to 24.9
 - children: Number of children covered by health insurance / Number of dependents
 - smoker: Smoking
 - region: the beneficiary's residential area in the US, northeast, southeast, southwest, northwest.
 - charges: Individual medical costs billed by health insurance
 - Source: <https://www.kaggle.com/datasets/mirichoi0218/insurance>