

CP3252 - Machine Learning Lab - Week 9

Date: 11.03.2025

1. Consider California Housing Prices dataset available at <https://www.kaggle.com/datasets/camnugent/california-housing-prices/>

Apply K-means clustering on the median income attribute to create economic segments in different regions of California.

2. Write the python code to implement Principal Component Analysis (PCA) and run the code for PCA on the flight data provided in the FlightDelay.csv file.
 - a) How many components did the PCA generate when you first run the analysis?
 - b) Provide a table of feature weights for each component.
 - c) How many components would you keep and why?
 - d) Using answers from (b) and (c) which features would you keep?