Machine Learning ICS-485

Assignment 1: Regression

1. Dataset

The life expectancy dataset has been collected from the World Health Organization (WHO) data repository and from United Nation (UN) website. Among all categories of health-related factors only those critical factors were chosen which are more representative. It considers data from year 2000 to 2015 for 193 countries, containing 22 columns and 2938 rows. All predicting variables was divided into several broad categories: Immunization related factors, Mortality factors, Economical factors and Social factors.

The dataset is made publicly available on Kaggle. The description and dataset can be accessed through the following link. https://www.kaggle.com/datasets/kumarajarshi/life-expectancy-who

2. Tasks

Your task in this assignment is to construct regression models using this data set to predict the life expectancy for any given country in a particular year. The evaluation of the models is performed as follows.

- Adopting 10-fold nested cross-validation
- Using MAE (Mean Absolute Error) and RMSE (Root Mean Square Error) to measure the performance of the models.

Analyze to answer the following questions.

1. Do various predicting factors which has been chosen initially really affect the Life expectancy? What are the predicting variables actually affecting the life expectancy?
2. Should a country having a lower life expectancy value (<65) increase its healthcare expenditure in order to improve its average lifespan?
3. How does Infant and Adult mortality rates affect life expectancy?
4. Does Life Expectancy have positive or negative correlation with eating habits, lifestyle, exercise, smoking, drinking alcohol etc.
5. What is the impact of schooling on the lifespan of humans?
6. Does Life Expectancy have positive or negative relationship with drinking alcohol?
7. Do densely populated countries tend to have lower life expectancy?
8. What is the impact of Immunization coverage on life Expectancy?

3. Submission

- Deadline: 25th - Feb - 2024
- Source code and a short report in pdf consisting the experimental setting, comparison between methods and answers for above questions, will be submitted through Blackboard.

Enjoy!