

## 作业要求：成年人死亡率预测

成年人死亡率指的是每一千人中 15 岁至 60 岁死亡的概率（数学期望）。这里我们给出了世界卫生组织（WHO）下属的全球卫生观察站（GHO）数据存储库跟踪的所有国家健康状况以及许多其他相关因素。要求利用训练数据建立回归模型，并预测成年人死亡率（Adult Mortality）。

训练数据（train\_data.csv）总共包含 2336 条记录，22 个字段，主要字段说明如下：

- Country: 国家
- Year: 年份
- Status: 发达国家或发展中国家
- Life expectancy: 预期寿命
- Infant deaths: 每千人口中的婴儿死亡人数
- Alcohol: 人均酒精消费量（以升纯酒精为单位）
- percentage expenditure: 卫生支出占人均国内生产总值的百分比
- Hepatitis B: 一岁儿童乙型肝炎免疫疫苗接种率
- Measles: 麻疹每 1000 人报告的病例数
- BMI: 所有人群平均 BMI 指数
- under-five deaths: 每千人口中五岁以下死亡人数
- Polio: 1 岁儿童脊髓灰质炎免疫覆盖率（%）
- Total expenditure: 政府卫生支出占政府总支出的百分比
- Diphtheria: 1 岁儿童白喉、破伤风类毒素和百日咳免疫接种率（%）
- HIV/AIDS: 每千名活产婴儿死于艾滋病毒/艾滋病（0-4 岁）
- GDP: 人均国内生产总值（美元）
- Population: 人口
- thinness 1-19 years: 10 至 19 岁儿童和青少年的消瘦流行率
- thinness 5-9 years: 5 至 9 岁儿童中的消瘦流行率
- Income composition of resources: 财力收入构成方面的人类发展指数（从 0 到 1）
- Schooling: 受教育年限
- Adult Mortality: 成人死亡率（每 1000 人中 15 至 60 岁死亡的概率）

测试数据(test\_data.csv)总共包含 592 条记录, 21 个字段, 和训练数据相比, 除了不包含 Adult Mortality 字段外, 其他完全相同。

具体要求如下：

1. 训练数据包含 2336 条记录和 22 个字段，对训练数据进行一定的可视化分析
2. 利用训练数据，选择合适的信息作为特征建立回归模型，并预测测试数据成年人死亡率
3. 利用 MO 平台进行模型性能评估，提交相关的**程序报告**和 **ipynb 代码文件**。

注意事项如下：

1. 推荐使用基于 Python 的 sklearn 库进行相关实验
2. 数据中可能会有一些字段的值存在缺失