蒸汽量预测

赛题描述

# ****1、赛题背景****

火力发电的基本原理是：燃料在燃烧时加热水生成蒸汽，蒸汽压力推动汽轮机旋转，然后汽轮机带动发电机旋转，产生电能。在这一系列的能量转化中，影响发电效率的核心是锅炉的燃烧效率，即燃料燃烧加热水产生高温高压蒸汽。锅炉的燃烧效率的影响因素很多，包括锅炉的可调参数，如燃烧给量，一二次风，引风，返料风，给水水量；以及锅炉的工况，比如锅炉床温、床压，炉膛温度、压力，过热器的温度等。

# ****2、赛题描述****

经脱敏后的锅炉传感器采集的数据（采集频率是分钟级别），根据锅炉的工况，预测产生的蒸汽量。

# ****3、数据说明****

数据分成训练数据（train.txt）和测试数据（test.txt），其中字段”V0”-“V37”，这38个字段是作为特征变量，”target”作为目标变量。选手利用训练数据训练出模型，预测测试数据的目标变量，排名结果依据预测结果的MSE（mean square error）。

# ****4、结果提交****

选手需要提交测试数据的预测结果（txt格式，只有1列预测结果）。

# ****5、结果评估****

预测结果以mean square error作为评判标准。

赛题分析

# 1、建模要求

本赛题是通过传感器采集到的数据，根据锅炉工况，来预测产生的蒸汽量，因此这是一个回归问题。

训练数据有标签，因此是一个监督学习数据集，可以考虑采用：线性回归，逻辑回归，决策树回归，SVM回归，随机森林回归。

评估方法：也就是损失函数的选择，MSE、MAE、Huber、log-coch、分数位。

# 2、数据分析

导入数据，首先查看数据特征的类型，整体查看数据分布。

数据观察与可视化

数据均为数值型，没有缺失值，同时对于异常值，这里没有考虑

# 3、数据预处理

为减小数据特征范围的影响，这里对数据进行规范化处理或标准化处理。

特征工程：1、去除无用特征；2、去除影响极小的特征；3、降维处理；

3、规范化或者标准化处理

# 4、模型尝试

尝试多种模型并评估模型的准确度

## 4.1 尝试模型

线性模型：ElasticNet, Lasso, BayesianRidge, LassoLarsIC,LinearRegression

支持向量机：SVR

决策树：DecisionTree

集成学习：RandomForestRegressor, GradientBoostingRegressor

## 4.2 模型评估方法

留出法

交叉验证：cross\_val\_score

k折交叉验证：KFold

自助法

## 4.3 借助函数来评估验证模型

Def -