

# 共享单车

K a g g l e 实战案例

网易云课堂 | © 微专业

# 简介



01

给出美国首都华盛顿**DC**的两年之中，每个月前**19**天每个小时的共享单车使用情况

02

预测每个月剩下的每天里每个小时的共享单车需求量

03

自动调参和模型堆砌

# 课程安排

01 数据读取

02 特征工程

03 挑选模型

04 搭建模型



# 01 读取数据

## Pandas

- 读取训练集和测试集
- 没什么特别好说的

- RMSLE

$$\sqrt{\frac{1}{n} \sum_{i=1}^n (\log(p_i + 1) - \log(a_i + 1))^2}$$

Pandas



## 02 特征工程

### Pandas

- 时间特征
- 也没什么特别好说的

Pandas



## 03 挑选模型

### Scikit Learn

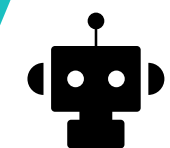
- Lasso Regression
- Ridge Regression
- XGBoosting Regressor
- Random Forest Regressor
- support vector regression



## 03 挑选模型

- 手动调参：
  - 欠拟合和过拟合之间的调试
- 自动调参：
  - Grid Search : [0, 2, 4, 6, 8]
  - Random Search : [1.3, 6.5, 3.2, 7.5, 0.1]

## 03 挑选模型（自动调参）

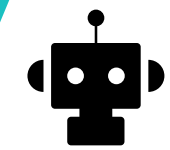


### Lasso 调参

- Alpha: 惩罚因子，越大越防过拟合
- ...



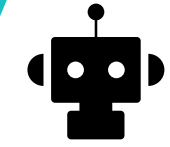
## 03 挑选模型（自动调参）



### Ridge 调参

- Alpha: 惩罚因子，越大越防过拟合
- Solver: 求解器，效果对不同形状数据不同
- ...

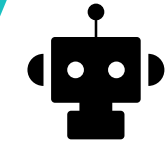
## 03 挑选模型（自动调参）



### XGB Regressor 调参

- 'objective': 'reg:linear'
- Max\_depth: 最大深度, 越深越拟合
- Reg\_alpha: 惩罚因子, 越小越拟合
- Learning\_rate: 学习率, 越大学的越快但是可能错过错过最优点, 越大越拟合
- ...

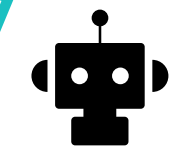
## 03 挑选模型（自动调参）



### Random Forest Regressor 调参

- Max\_depth: 最大深度, 越深越拟合
- Max features: 最多特征数, 越大越拟合
- n\_estimators: 有多少颗树
- ...

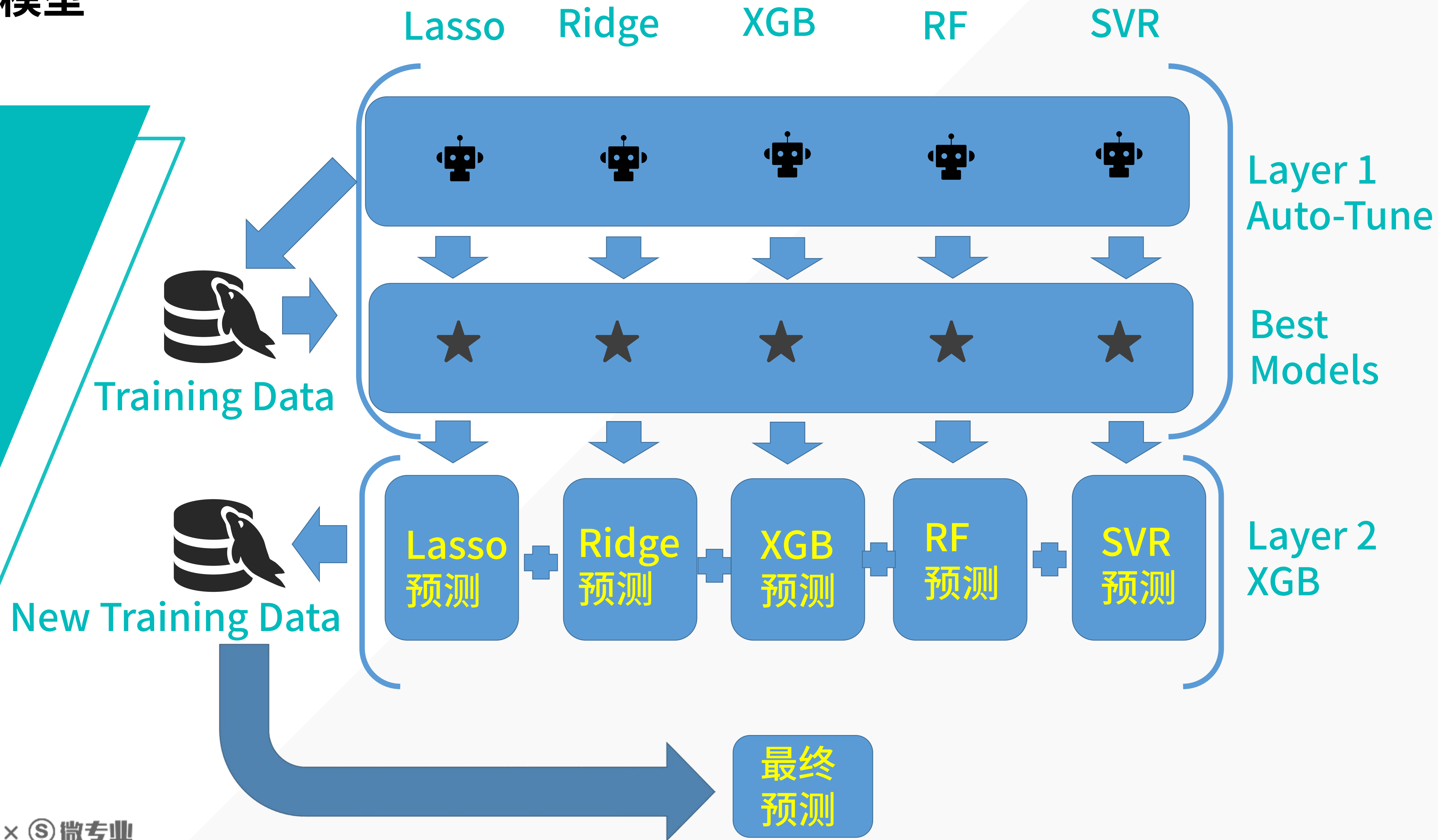
## 03 挑选模型（自动调参）



### Support Vector Regression 调参

- Kernel: 核函数,将低维数据映射到高维从而达到线性可分
- Gamma: Kernel的系数, 越小表示每个点影响范围更大
- C: 惩罚系数, 越大越拟合
- ...

## 04 搭建模型





# THANK YOU