

# 线性回归模型代码实现

Numpy、pandas、scikit-learn

小胖



# 目录

## ONE 生成训练数据

Numpy、pandas、matplotlib

## TWO 模型实现

scikit-learn

## THREE 两种线性回归对比

OLS v.s. LAD

# 生成训练数据

Numpy、pandas、matplotlib

生产记事本

日期	玩偶个数	成本	第几天
04/01	10	7.7	1
04/02	10	9.87	2
04/03	11	10.87	3
04/04	12	12.18	4
04/05	13	11.43	5
04/06	14	13.36	6
04/07	15	15.15	7
04/08	16	16.73	8
04/09	17	17.4	9
...	...	...	...

x

y

$$y = x + \varepsilon$$

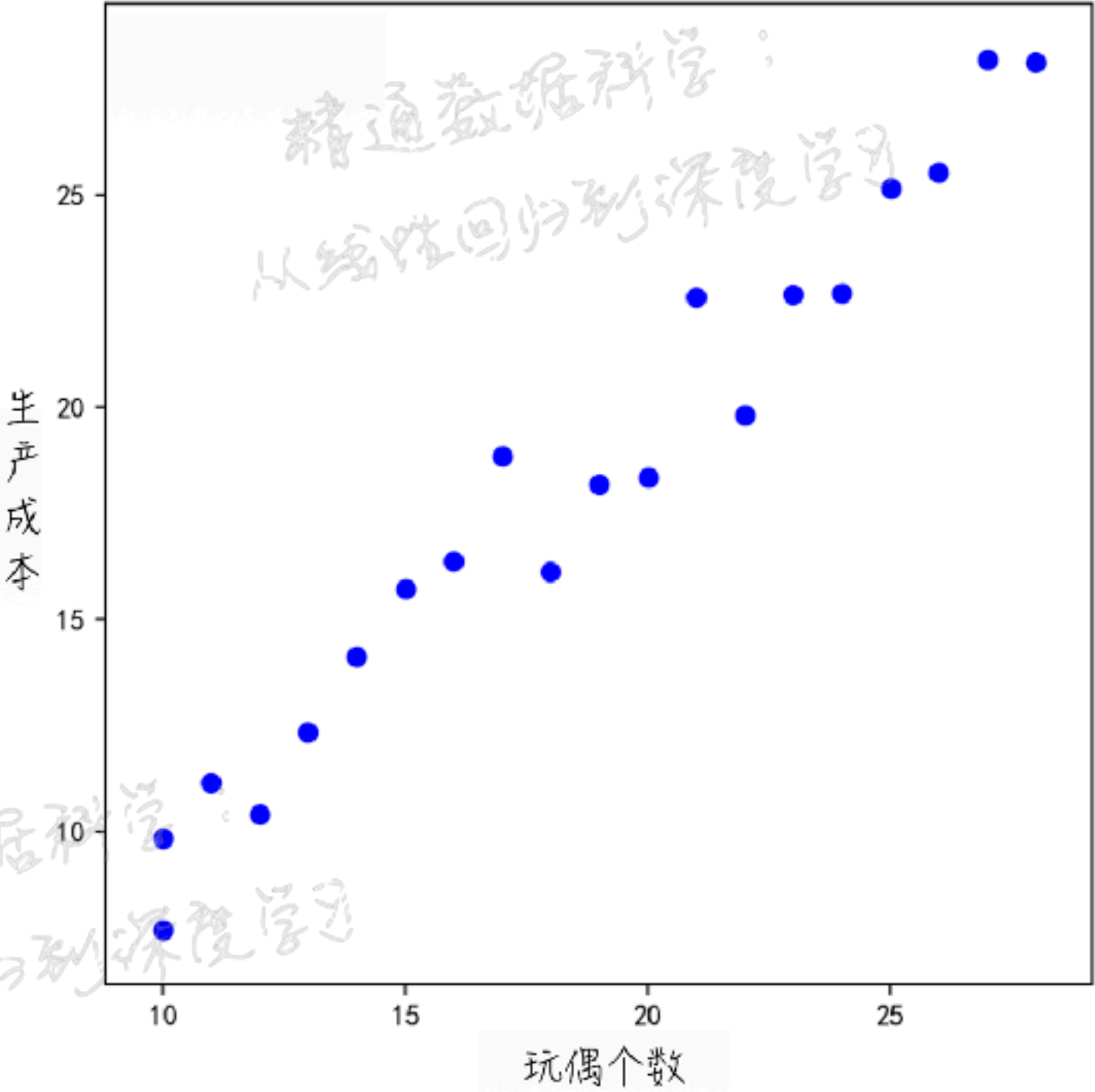
使用Numpy, 生成x和y

使用pandas, 由x和y生成DataFrame

使用matplotlib, 将数据可视化

保存数据为csv格式

生产数据



# 模型实现

scikit-learn

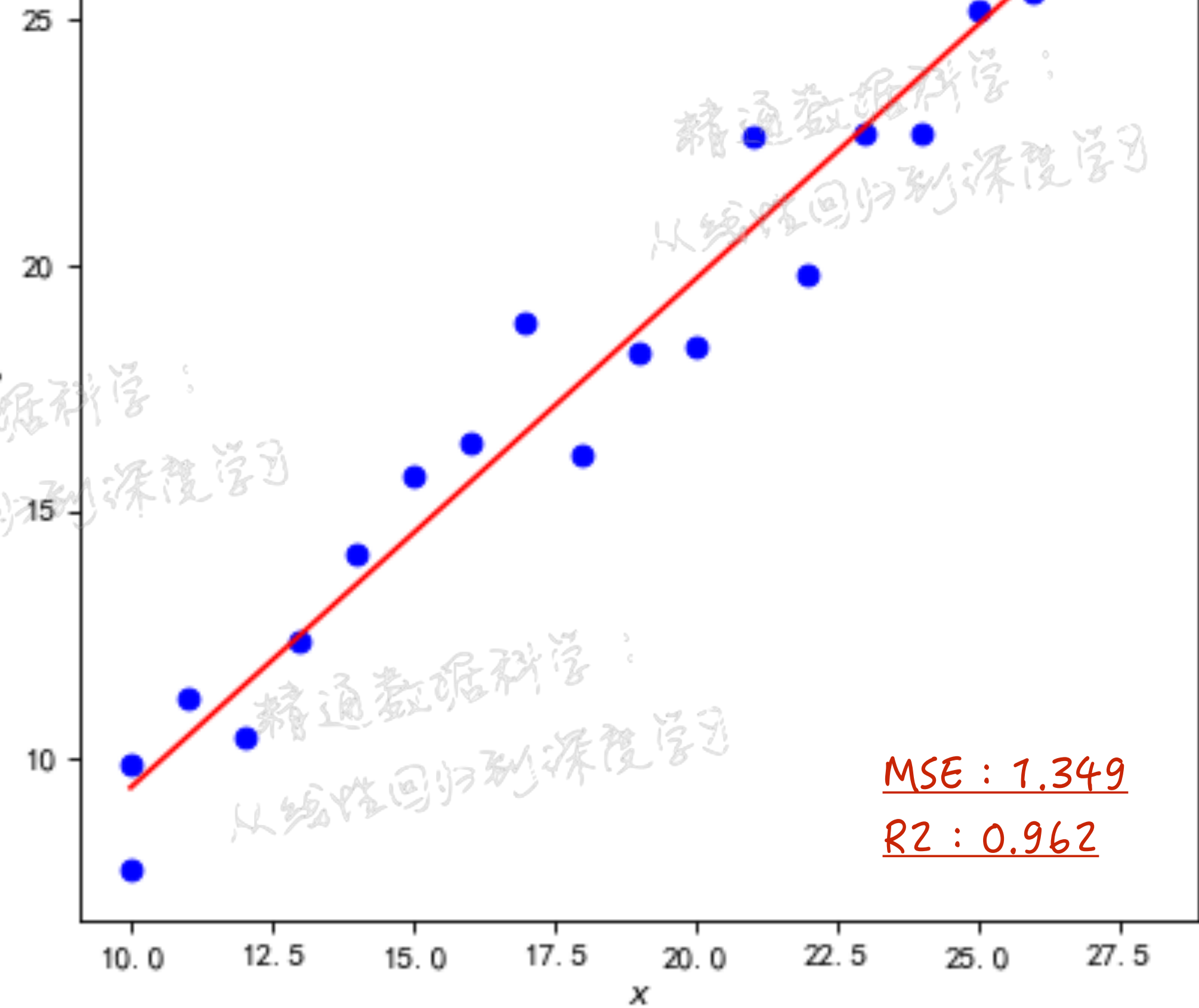
使用pandas, 读取训练数据

使用scikit-learn, 训练模型

计算MSE和R2, 评估模型

使用matplotlib, 将模型结果可视化

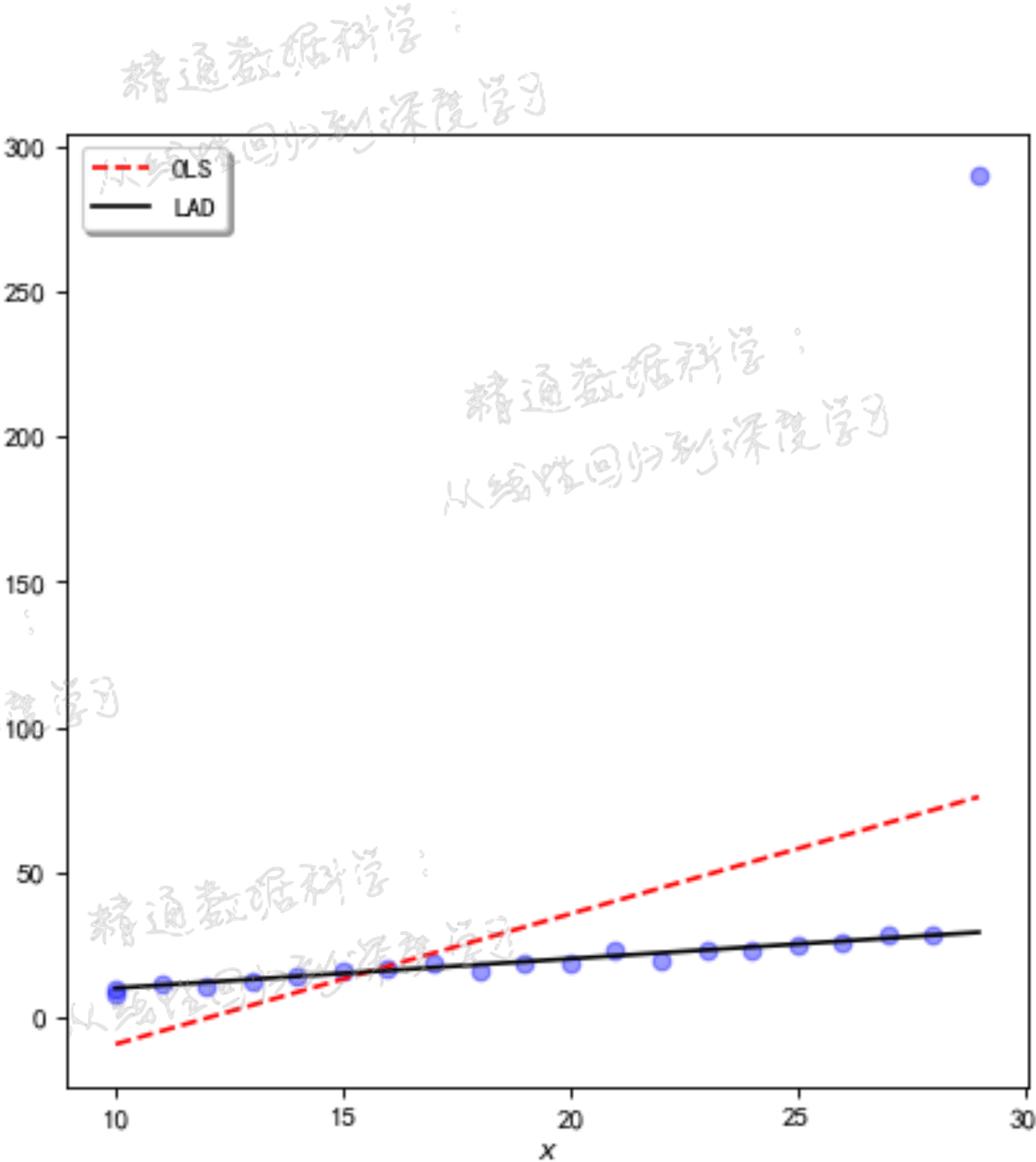
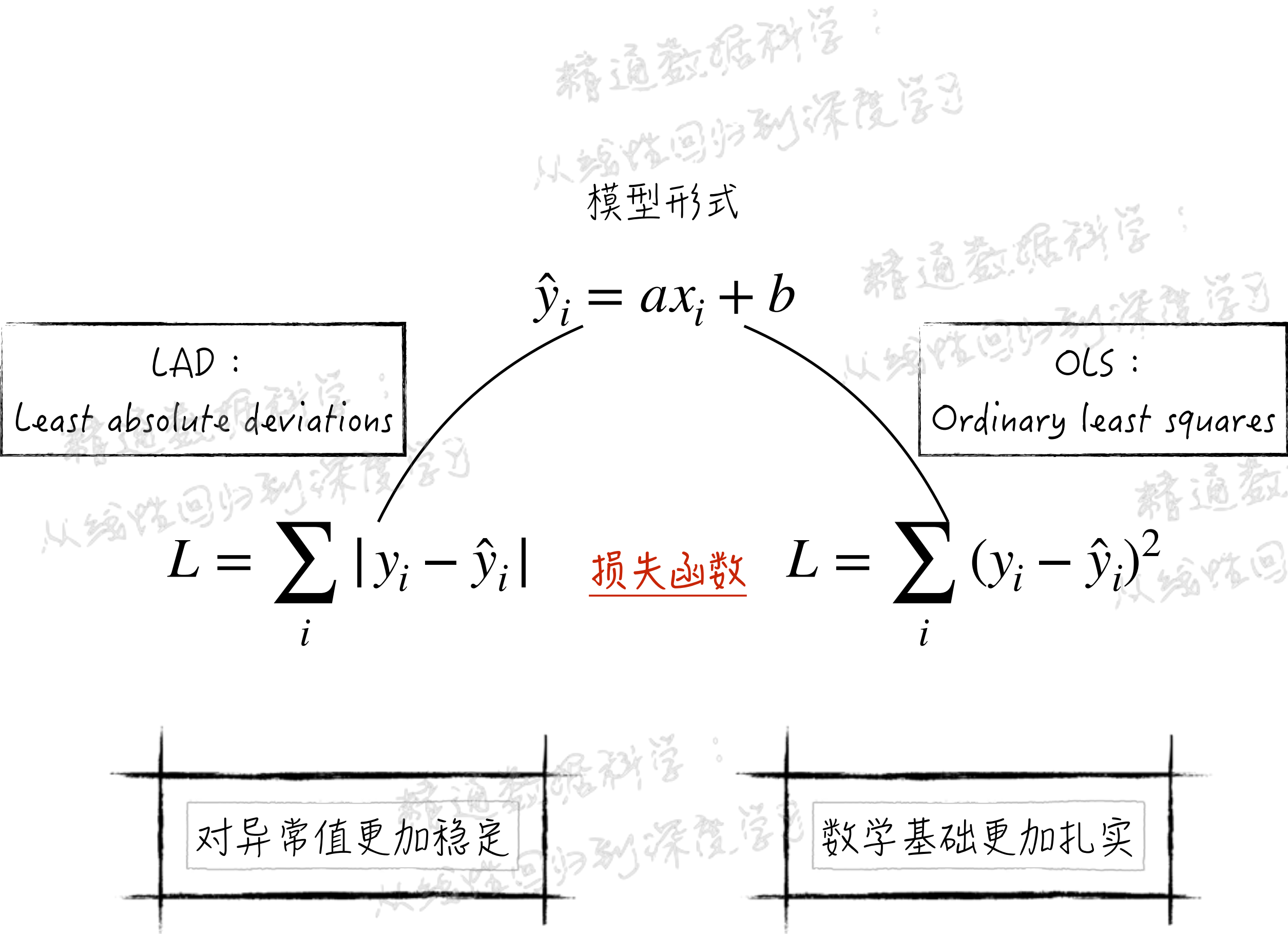
$$y = 1.033x - 0.950$$





# 两种线性回归对比

OLS v.s. LAD



精通数据科学：  
从线性回归到深度学习

# THANK YOU