

模型陷阱—代码实现

惩罚项和假设检验

小胖

目录

ONE 过度拟合例子

scikit-learn

TWO 机器学习的解决方法

惩罚项、网格搜寻

THREE 统计分析的方法

假设检验

过度拟合例子

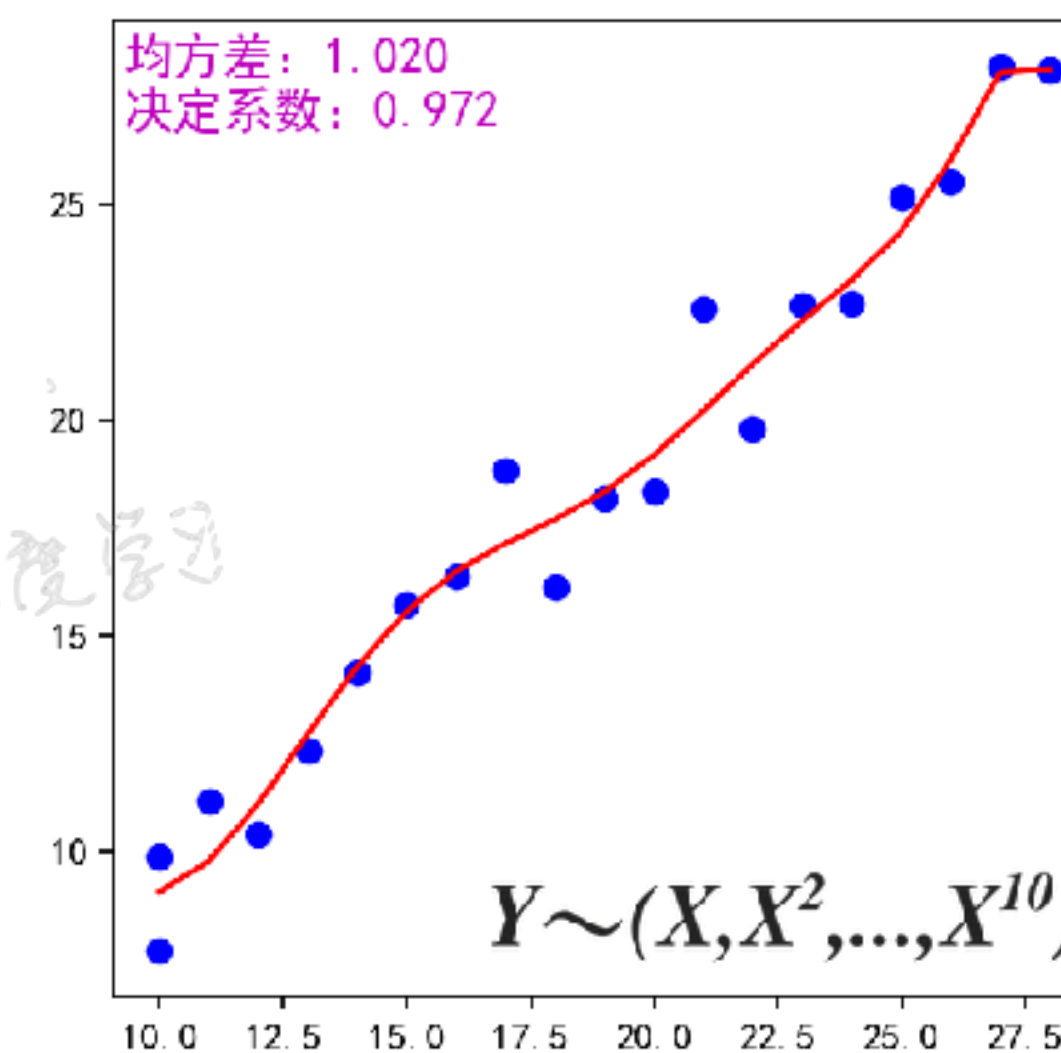
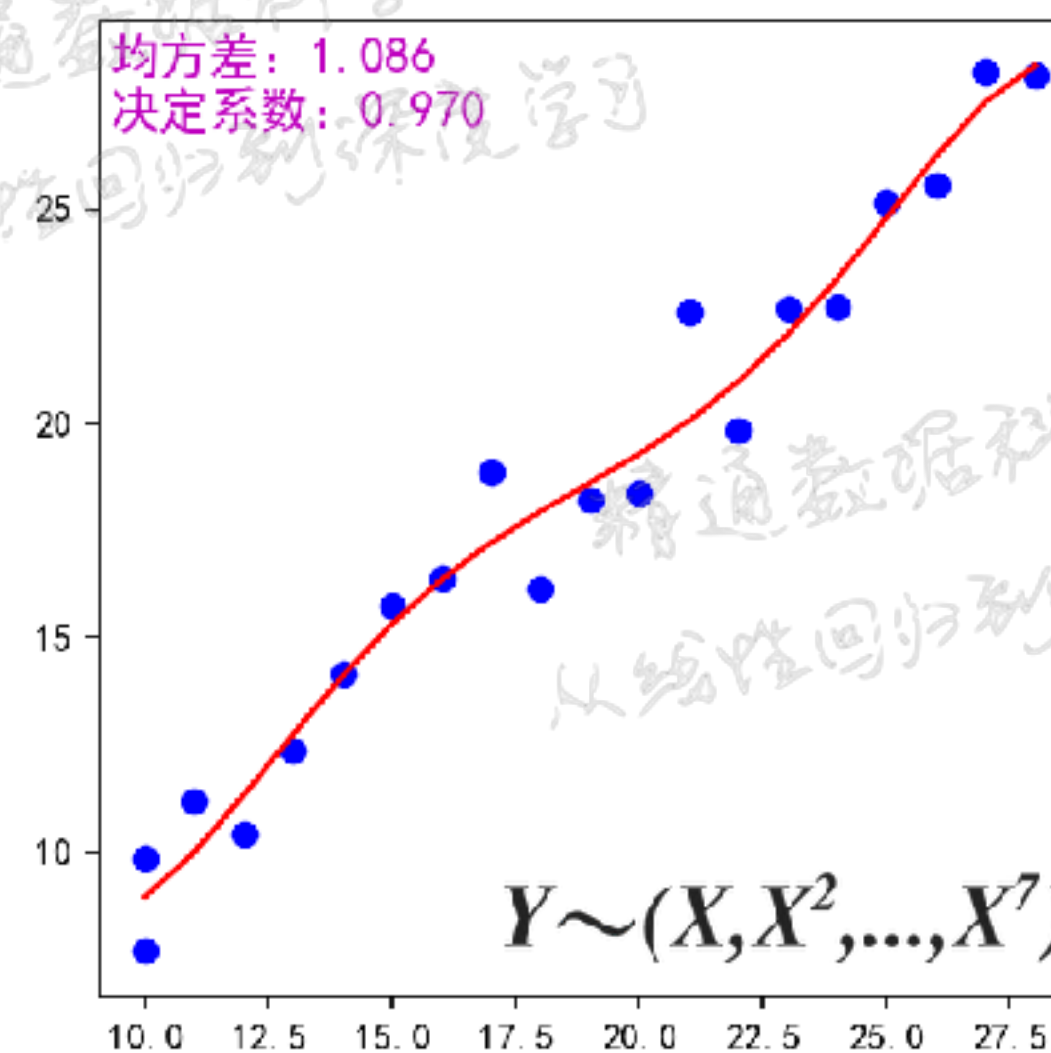
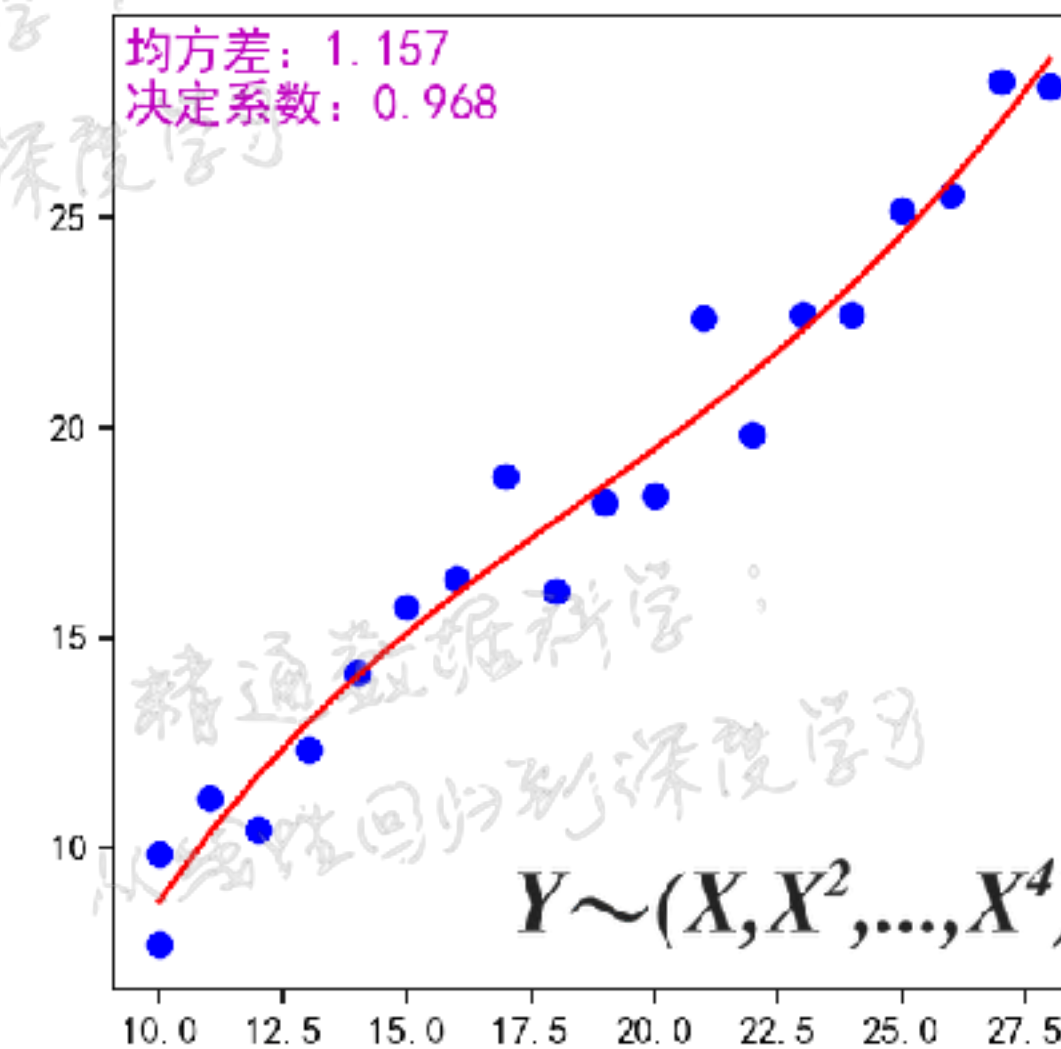
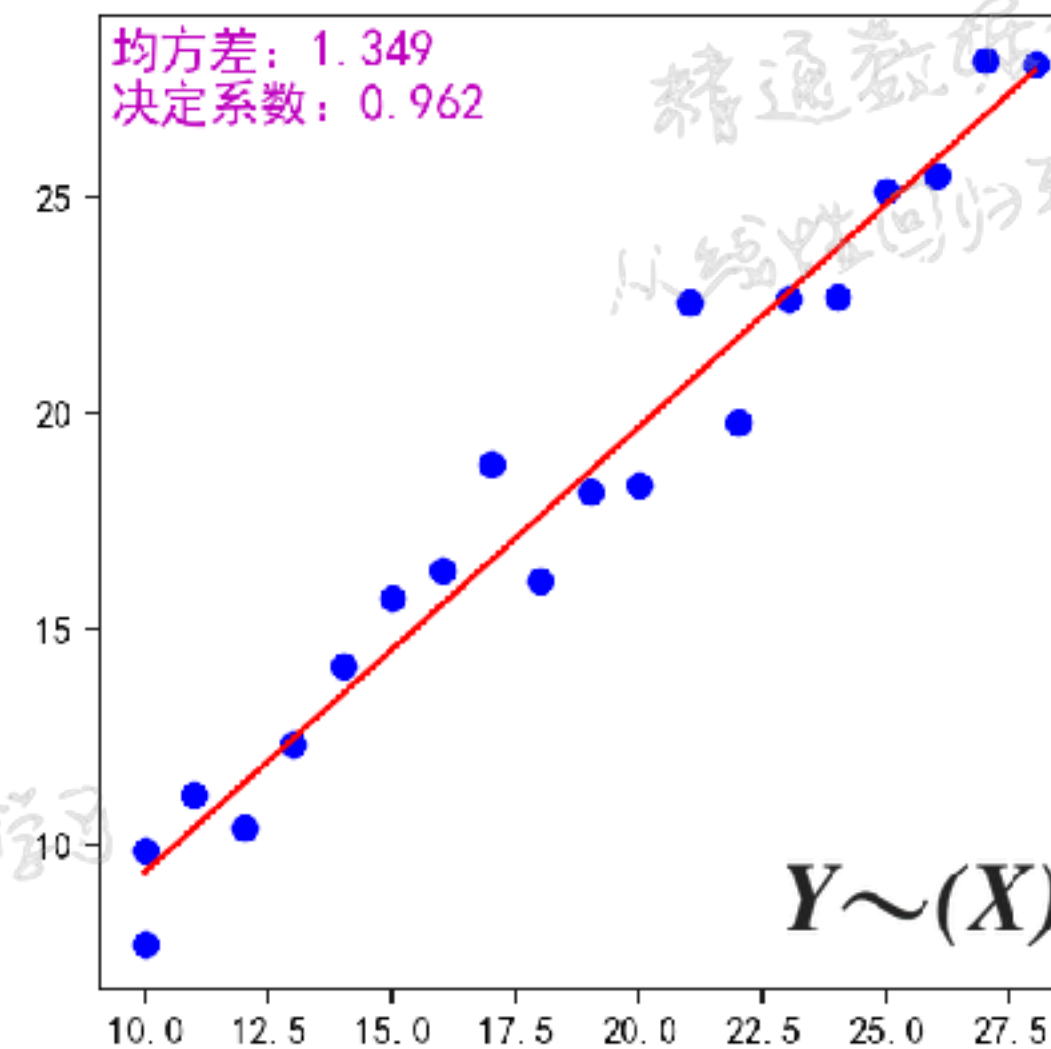
scikit-learn

使用pandas, 读取历史数据

将数据分为训练集、测试集
重复整个建模过程

使用PolynomialFeatures, 生成新的特征

使用matplotlib, 将模型结果可视化



惩罚项

生成一系列备选的超参数alphas

使用matplotlib, 将模型参数可视化



机器学习的方法

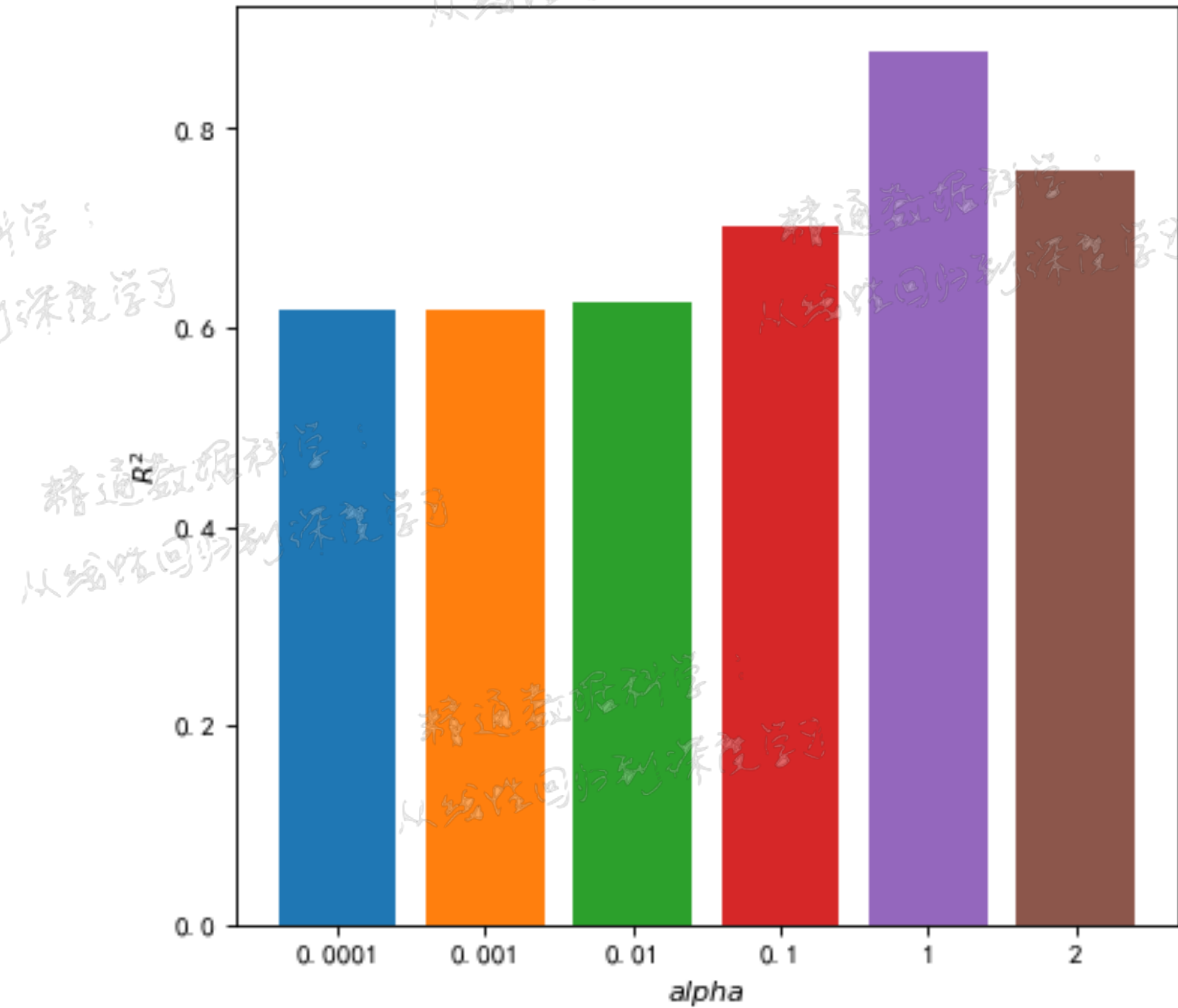
网格搜寻

使用pandas, 读取历史数据

生成一系列备选的超参数alphas

使用GridSearchCV, 构造网络搜寻

使用matplotlib, 将模型结果可视化



统计分析的方法

假设检验

使用pandas, 读取历史数据

在模型中加入不相关的变量z

根据假设检验的结果, 将变量z排除出模型

OLS Regression Results

Dep. Variable:	y	R-squared:	0.963			
Model:	OLS	Adj. R-squared:	0.959			
Method:	Least Squares	F-statistic:	222.8			
Date:	Tue, 25 Apr 2017	Prob (F-statistic):	6.38e-13			
Time:	21:39:40	Log-Likelihood:	-31.141			
No. Observations:	20	AIC:	68.28			
Df Residuals:	17	BIC:	71.27			
Df Model:	2					
Covariance Type:	nonrobust					
=====						
	coef	std err	t	P> t	[0.025	0.975]
-----	-----	-----	-----	-----	-----	-----
const	-0.8983	0.953	-0.942	0.359	-2.910	1.113
x	1.0400	0.050	20.722	0.000	0.934	1.146
z	-0.3619	0.571	-0.634	0.535	-1.566	0.843
=====				=====		
Omnibus:	1.295	Durbin-Watson:	2.209			
Prob(Omnibus):	0.523	Jarque-Bera (JB):	0.852			
Skew:	-0.061	Prob(JB):	0.653			
Kurtosis:	1.996	Cond. No.	66.7			

精通数据科学：
从线性回归到深度学习

THANK YOU