



deepshare.net

深度之眼

二分类比赛分享会

导师：hphacker战队

本次分享内容

Course content



1、赛题背景

2、特征工程

3、小trick

4、深度思考



deepshare.net

深度之眼

赛题背景

介绍本次比赛基本框架

赛题背景



结构化数据

类别特征

	ID	age	job	marital	education	default	balance	housing	loan	contact	day	month	duration	campaign	pdays	previous	poutcome	y
0	1	43	management	married	tertiary	no	291	yes	no	unknown	9	may	150	2	-1	0	unknown	0
1	2	42	technician	divorced	primary	no	5076	yes	no	cellular	7	apr	99	1	251	2	other	0
2	3	47	admin.	married	secondary	no	104	yes	yes	cellular	14	jul	77	2	-1	0	unknown	0
3	4	28	management	single	secondary	no	-994	yes	yes	cellular	18	jul	174	2	-1	0	unknown	0
4	5	42	technician	divorced	secondary	no	2974	yes	no	unknown	21	may	187	5	-1	0	unknown	0

结构化数据

数值特征

特征工程

数据决定了机器学习的上限，而算法
只是尽可能逼近这个上限

特征工程

树模型：LGB



deepshare.net

深度之眼

保持单调性的特征变换没有必要

类别特征和数值特征在LGB中的区别

NaN有单独的处理方式(类似于XGB) (-1)

特征工程

编码



deepshare.net

深度之眼

ID特征：去除

'job', 'marital', 'education', 'contact', 'poutcome'

Default特征：0, 1 编码

数值编码

Housing特征：0, 1 编码

unknown没有当作NaN

Loan特征：0, 1 编码

Month特征：1-12 编码

特征工程



特征粗粒化

季度特征

是否有存款

特征细粒化

总天数特征

特征周期化（非单调）

本月第几周特征

年龄的模特征

答疑问题

特征组合统计

关于月日

'campaign', 'job', 'marital', 'education',
'default', 'housing', 'loan', 'contact',
'poutcome'

特征相互统计

'job', 'marital', 'education'

'balance', 'duration'

小trick

小技巧

小trick

1、Bagging策略和类别特征选择

小trick

2、Pdays的问题

小trick

3、样本迁移

小\trick



2019-07-06 12:06:52

0.94040014

深度思考

飞机和飞鸟形态、结构和原理都不相同，但都能飞翔，人工智能的未来也许如此——卡卡罗特

深度思考



1、缺失值问题

2、类别特征编码和组合

3、时间序列



deepshare.net

深度之眼

互动时间

结语

—— 结 语 ——

谢谢大家！





deepshare.net

深度之眼

联系我们：

电话：18001992849

邮箱：service@deepshare.net

QQ：2677693114



公众号



客服微信

