数据分析和建模中级教程系列讲座

主讲人：元俊

* 本课程的面向群体：

中级数据分析和建模的爱好者

对数据建模和分析有诉求、有兴趣的各行各业的从业人员

* 基础要求：

基本的数学常识

进一步的要求（可选）：

概率论初级 统计学初级 线性代数初级 编程基础

* 解决哪些问题：

基本sense

how to建模和分析

如何解决日常工作和生活的数据问题

课程的风格：

清晰，简洁，明朗

不会展开长篇大论的公式和代码

gitHub...

导论

核心概念 数据 模型 算法

什么是模型？

模型是一个假想的数学公式集合。

但模型不是一次就能建好的。需要大量数据的积累和磨合，去伪存真。素描

模型的作用：

挖掘和预测 风控领域

一个糟糕的模型会导致灾难（赤壁）

算法无处不在，本质上就是把模型训练成更加精确逼真、更加符合预期

模型预测有哪些？

二分|多元|回归

二分分类... Y/N BLACK/WHITE 0/1 TRUE/FALSE

多元分类... LOW/MEDIUM/HIGH 赞成/反对/无所谓

回归 气温，湿度，股价，航行旅程，每天的心跳数

怎么评价模型(二分)？

代价函数

混淆矩阵 P/R/F1 AUC ROC TPR/FPR

艺考的例子

1.下一讲预告 Logistics Regression

2.思考题：年龄是否适合作为建模参数？