

2017 年暑期中国人民大学统计学专业知识系列培训班

----大数据挖掘与统计机器学习

主办单位：中国人民大学统计学院

全国应用统计专业学位研究生教育指导委员会秘书处

培训时间：2017 年 7 月 22-23 日

培训地点：中国人民大学明德主楼

授课对象：从事相关专业教学的高校教师；大数据、IT、金融等相关行业从业人员；具有扎实数学、计算机基础的高校学生。

培训目标：

深入理解大数据挖掘与统计机器学习的各种方法、原理；熟练掌握上述方法的 R 语言及 Python 实现；介绍大数据分析案例及其分布式实现。

参考书目：

吕晓玲、宋捷（2016），《大数据挖掘与统计机器学习》，中国人民大学出版社，全国应用统计专业学位研究生教育指导委员会推荐用书。

培训内容：

时间	模块	内容
第一天	线性回归、分类方法，模型评价和选择	<ul style="list-style-type: none"> ● 大数据挖掘与统计机器学习概述：数据智慧 <ul style="list-style-type: none"> ➢ 数据分析的一般流程 ● 线性回归方法 <ul style="list-style-type: none"> ➢ 多元线性回归回顾 ➢ Lasso 与岭回归 ➢ 损失函数加罚的建模框架 ● 线性分类方法 <ul style="list-style-type: none"> ➢ 分类问题综述 ➢ Logistic 回归与线性判别 ● 模型评价与选择 <ul style="list-style-type: none"> ➢ 基本概念 ➢ 理论方法 ➢ 数据重利用方法
	决策树与组合方法、支持向量机	<ul style="list-style-type: none"> ● 决策树与组合方法 <ul style="list-style-type: none"> ➢ 决策树 ➢ Bagging、Boosting、随机森林 ● 支持向量机 <ul style="list-style-type: none"> ➢ 线性可分情况 ➢ 软间隔支持向量机及其拓展

第二天	神经网络与深度学习、聚类分析与推荐系统	<ul style="list-style-type: none"> ● 神经网络与深度学习 <ul style="list-style-type: none"> ➢ 神经网络 ➢ 深度学习 ● 聚类分析与推荐系统 <ul style="list-style-type: none"> ➢ 聚类分析 ➢ 推荐系统
	上机操作与大数据分析案例	<ul style="list-style-type: none"> ● 上述方法的 R 与 Python 的实现 ● 大数据分析案例 <ul style="list-style-type: none"> ➢ 案例一：智能手机用户行为分析 ➢ 案例二：美国航空数据分析 ➢ 案例三：纽约公共自行车数据分析 ➢ 基于数据库、R、Python 的单机实现 ➢ 基于 Hadoop、Hive、Map-Reduce、Spark 的分布式实现

主讲嘉宾：

吕晓玲，中国人民大学统计学院教授，博士生导师；中国人民大学数据挖掘中心主任；五校联合（中国人民大学，北京大学，中国科学院大学，中央财经大学，首都经济贸易大学）大数据分析硕士培养协同创新平台总协调人；北京大数据协会副秘书长。一直从事数据挖掘和统计机器学习领域的理论研究，及其在消费者行为方面的应用研究。主持教育部人文社会科学研究项目以及中国国家自然科学基金项目。数十篇学术论文在 SSCI 以及 SCI 检索的国际学术期刊发表。在数据挖掘以及市场营销方面的项目涉及的领域包括银行、电子商务、交通、教育、广播电视、移动互联网等。

报名方式：

登陆中国人民大学统计学院网站 <http://stat.ruc.edu.cn> 在线咨询报名。

缴费标准及方式：1288 元

1、网上银行缴费（建议）

报名后，我们会将缴费链接和验证码发到报名邮箱中，请点击链接进行网银缴费。

2、银行汇款

收款人全称：中国人民大学 收款人帐号：0200 0076 0902 6400 244

收款人开户银行：中国工商银行紫竹院支行

汇款备注栏中必须注明统计学院和培训项目名称。

3、现场刷卡缴费（仅限工作日）

报到时统一开具等额有效发票。食宿自理。

颁发证书：

圆满完成本项目，符合结业条件，由中国人民大学教育培训管理办公室统一颁发中国人民大学大数据挖掘与机器学习培训班结业证书，加盖中国人民大学教育培训管理办公室和中国人民大学统计学院公章，证书号可登陆中国人民大学教育培训网站查询。

咨询电话：010-62511316，徐老师；010-62511318，张老师

中国人民大学统计学院
2017年6月

