Functional data analysis：methods, theories and applications

Time and location: Every Wednesday evening (18.00-19:40) in 3301

Course organizer: 柏杨

Current Instructors: 柏杨、李婷、尤进红，黄涛

Course home: School of Statistics and Data Science, SUFE

Course description:

In many practical problems originating from areas like geophysical sciences, astronomy, medicine, etc., the observations often correspond to measurements of a continuous process, which is a function of time and/or space. With rapid advances in technology, such measurements are becoming more and more frequent, thus pushing the boundaries of high-dimensional data analysis. In data science, high dimensional problems are often approached through dimension free analysis by treating an observation as a function. This is one way to circumvent the technical aspects identified with analyzing high-dimensional data, and thus avoiding the curse of dimensionality. This course is meant to introduce the students to various aspects of dimension free analysis, which obviates the technical and computational hurdles associated with high dimensional data. This way of dealing with high dimensional data is an emerging and rapidly developing field that requires understanding both established methods and newly adopted techniques. The primary objective of this course will be to focus on the application of functional data analysis techniques to real world problems, and thus, mathematical rigor is often traded for adaptability to applications. Beginning with the basics of the analysis of data that may be considered to be “functions”, this course will discuss various visualization and data exploration techniques. Specifically, the course will extensively deal with nonparametric spline smoothing, functional linear models, functional PCA, regularization methods, analysis involving derivatives, registration and nonlinear smoothing. Students are required to work on projects to apply the techniques on real world problems. The preferred software for this course will be ‘R’ and/or Python, however, the students are permitted to use any mathematical software of their liking that have facilities to perform all task in the course (Matlab being one example). Project discussions will enable students to share and compare ideas with each other and to receive specific guidance from the instructors. Efforts will be made to help students to embed real-world problems into mathematical models so that suitable algorithms can be applied with consideration of computational constraints. By surveying special topics, students will be exposed growing range of new methodologies.

Course Prerequisites: Familiarity with multivariate calculus and matrix algebra, basic knowledge of introductory mathematical statistics.

Course objectives:

As a result of successfully completing this course, the student will demonstrate:

- an understanding both technical and conceptual aspects of dealing with data treated as samples of functions,

- knowledge of methods that goes beyond standard multivariate statistical methodology through dimension free approach that accounts for the order in observed variables,
- an understanding how to represent data as functions, when the data are inherently too highly dimensional to be effectively treated as a vector, and
- an understanding the role which derivative functions play in construction solutions to practical problems.
- demonstrate the ability to efficiently handle the data that have an ordering inscribed in them and/or are highly dimensional through functional decomposition,
- demonstrate understanding of the functional data approach which extends from statistical methodology such as the principal component analysis and multivariate regression, and
- master a number of algorithmic approaches, with assistance of software packages, both to analyze and visualize functional data.
- demonstrate available choices for visualization and data exploration in the functional setting,
- identify the type of approach that is most suitable for the problem at hand, and
- critically discuss the results of analysis obtained by a particular method.

函数型数据分析：方法、理论与应用

上课时间/地点：周三下午 6：00-7：45/3301 教室

课程组织： 柏杨（博士、常聘副教授）

授课教师： 柏杨、李婷、尤进红、黄涛

课程单位：上海财经大学统计与数据科学学院

课程描述：

在地球物理科学、天文学、医学等领域的许多实际问题中，观测数据通常对应于对连续过程（作为时间和/或空间函数）的测量。随着技术的快速发展，这类测量变得越来越频繁，从而不断拓展高维数据分析的边界。在数据科学中，处理高维问题常采用无维度分析方法——将观测数据视为函数进行研究。这种思路能规避传统高维数据分析的技术难点，从而避免维度灾难问题。本课程旨在向学生系统介绍无维度分析的各个方面，该方法可有效解决高维数据带来的技术与计算挑战。这种处理高维数据的方式作为一个新兴且快速发展的领域，既需要掌握经典方法，也要求理解最新技术。

本课程的核心目标聚焦于函数型数据分析技术在实际问题中的应用，因此在教学过程中通常会以牺牲数学严谨性为代价来增强对实际应用的适应性。课程将从" 函数型数据" 分析基础展开，系统探讨多种可视化与数据探索技术，重点内容包括：非参数样条平滑、函数型线性模型、函数型主成分分析 (F-PCA)、正则化方法、涉及导数的分析、数据配准与非线性平滑等。学生需通过项目实践应用这些技术解决现实问题。课程推荐使用 R 和/或 Python 软件，但学生也可根据个人偏好选择其他具备课程所需功能的数据分析软件（如 Matlab 等）。项目讨论环节将促进学生之间的观点交流与比较，并得到教师的针对性指导。课程将着力帮助学生将现实问题转化为数学模型，在考虑计算约束的前提下应用合适算法。通过对专题研究的探索，学生将接触到这一领域不断涌现的新方法论体系。

预先知识准备：熟悉微积分和线性代数；数理统计的基本知识。

课程目标：

作为成功完成本课程学习的结果，学生将获得函数型数据分析基本方法及其在数据分析中应用的系统知识。具体来说，学生将掌握：

1. **理解理论与方法：**掌握将数据视为函数样本处理的技术与概念要点，了解通过考虑观测变量内在顺序的无维度分析方法，此类方法超越了传统多元统计方法论的范畴；
2. **数据表征能力：**理解如何将本质上维度极高、无法有效以向量形式处理的数据转化为函数形式表达；
3. **导数应用认知：**认识导数函数在构建实际问题解决方案中的重要作用；
4. **高效处理能力：**能够通过函数型分解技术，有效处理具有内在顺序特征和/或高维度的数据；
5. **方法扩展理解：**理解函数型数据分析方法如何从主成分分析（PCA）、多元回归等经典统计方法中延伸发展而来；
6. **算法工具掌握：**借助软件包掌握多种算法方法，实现函数型数据的分析与可视化；
7. **可视化实践：**展示在函数型数据场景下可用的可视化与数据探索技术选项；

8. **方法适配能力：**针对具体问题识别最适用的分析方法类型；
9. **批判性思维：**对特定方法得出的分析结果开展批判性讨论。