

本课程学习要求：

1. 知识准备：

线性代数

多元分析

基本概率推理（概率不等式，马氏过程等）

基本数理统计（多元回归，多元正态分布，中心极限定理等）

优化(凸优化)

编程能力：R      多数统计软件包在该环境下运行

Matlab 优化和稀疏矩阵处理能力优异

\*C/Python等，需要时候非结构化数据预处理

2. 本课程与“统计学习”一课为姊妹课，统计学习主要覆盖有监督学习(Supervised Learning)部分，本课程则试图覆盖一些非监督学习的方面并给出一个统一的数学系统。

### 3. 上课内容包括：

#### I. Geometric aspects of data

##### A. Geometric embedding of data:

- Euclidean Embedding (MDS + extensions)

- Hilbertian Embedding (Kernel methods)

##### B. Adaptive embedding and matrix factorization

- Dictionary learning (basis/SVD, frames, nonnegative, topic)

- Hierarchical Embedding (DeepNN)

##### C. Dimensionality Reduction

- Linear adaptive DR and PCA

- Random projections

- Nonlinear DR and manifold learning

##### D. Curse of Dimensionality and High dimensional statistics

- Failure of mean and covariance estimators when  $p > n$

- Sparse representation

- Random matrix theory

#### II. Topological aspects of data

- Clustering on data graphs

- High order homology on data complexes

- Connecting geometry and topology: Hodge Theory

  - Spectral clustering and graph Laplacian

  - Statistical Ranking and graph Helmholtzian

  - Game Theory

  - Localizing the holes: Sensor Network coverage

这些课程的顺序可以调整。而且每一个内容可能不止一次课就可以讲完。