课程大纲
COURSE SYLLABUS

1.	**课程代码/名称** **Course Code/Title**	**高维统计分析 High Dimensional Statistics**
2.	**课程性质** **Compulsory/Elective**	**Elective**
3.	**课程学分/学时** **Course Credit/Hours**	3/48
4.	**授课语言** **Teaching Language**	中英双语 English & Chinese
5.	**授课教师** **Instructor(s)**	授课教师：李曾 Zeng Li 所属学系：统计与数据科学系 Department of Statistics and Data Science 联系方式：liz9@sustech.edu.cn
6	**是否面向本科生开放** **Open to undergraduates or not**	否
7.	**先修要求** **Pre-requisites**	（如面向本科生开放，请注明区分内容。 If the course is open to undergraduates, please indicate the difference.） 统计线性模型（MA329），多元统计分析（MA304） Statistical Linear Models (MA329), Multivariate Statistical Analysis (MA 304)

8. 教学目标
Course Objectives

（如面向本科生开放，请注明区分内容。 If the course is open to undergraduates, please indicate the difference.）

本课程旨在引导学生学习学科前沿的高维数据统计分析方法，帮助学生加深对惩罚最小二乘此类方法的理解，达到让学生学会使用适当的统计方法来处理高维数据分析中的问题。

This course aims to guide students learn frontier modeling methods in high dimensional statistical analysis and data science. It helps students deepen their understanding in penalized least square methods and reach the goal of solving practical high dimensional problems using advanced statistical methods and software.

9. 教学方法
Teaching Methods

（如面向本科生开放，请注明区分内容。 If the course is open to undergraduates, please indicate the difference.）

教师授课，课堂讨论

10. 教学内容
Course Contents
（如面向本科生开放，请注明区分内容。 If the course is open to undergraduates, please indicate the difference.）

Section 1	1 介绍（1 hour） 1.1 高维问题的出现 1.2 维数问题的影响 1 Introduction (1 Hour) 1.1 Rise of Dimensionality

	1.2 Impact of Dimensionality
Section 2	2 多元与非参数回归分析 (5 Hours) 2.1 加权最小二乘和岭回归 2.2 Box-Cox 变换 2.3 模型建立与展开 2.4 Reproducing Kernel Hilbert Space 中的回归 2.5 交叉验证 2 Multiple and Nonparametric Regression (5 Hours) 2.1 Weighted Least Squares and Ridge Regression 2.2 Box-Cox transformation 2.3 Model Building and Basis Expansions 2.4 Regression in Reproducing Kernel Hilbert Space 2.5 Leave-one-out and Generalized Cross-validation
Section 3	3 Lasso 线性模型 (8 Hours) 3.1 Lasso 估计量 3.2 交叉验证和推断 3.3 Lasso 估计量的计算 3.4 Lasso 的自由度 3.5 Lasso 解的唯一性 3.6 Lasso 的理论性质 3.7 The Nonnegative Garrote 3.8 Lq 惩罚和贝叶斯估计 3 The Lasso for Linear Models (8 Hours) 3.1 The Lasso estimator 3.2 Cross Validation and Inference 3.3 Computation of Lasso Solution 3.4 Degrees of freedom 3.5 Uniqueness of Lasso Solutions 3.6 A glimpse of the theory 3.7 The Nonnegative Garrote 3.8 Lq penalties and Bayes estimates

Section 4	4 用于变量选择的惩罚最小二乘方法 (14 Hours) 4.1 传统变量选择方法 4.2 凸惩罚最小二乘方法 4.3 Lasso 和 L1 惩罚 4.4 贝叶斯变量选择 4.5 数值算法 4.6 惩罚参数选择 4.7 残差方差和重拟合交叉验证 4 Penalized least square methods for variable selection (12 Hours) 4.1 Classical variable selection Criteria 4.2 Folded concave penalized least squares 4.3 Lasso and L1-regularization 4.4 Bayesian Variable Selection Procedures 4.5 Numerical Algorithms 4.6 Regularization parameter selection 4.7 Residual variance and refitted cross validation
Section 5	5 Lasso 惩罚项的推广 (4 Hours) 5.1 Elastic Net 5.2 Group Lasso 5.3 稀疏可加模型 5.4 Fused lasso 5 Generalizations of the Lasso Penalty (4 Hours) 5.1 Elastic Net 5.2 Group Lasso 5.3 Sparse Additive Models 5.4 Fused Lasso
Section 6	6 优化方法 (4 Hours) 6.1 凸优化条件 6.2 梯度下降 6.3 最小角度回归 6.4 ADMM 6.5 最小最大方法 6 Optimization Methods (4 Hours)

		6.1 Convex Optimality Conditions 6.2 Gradient Descent and Coordinated Descent 6.3 Least Angle Regression 6.4 Alternating Direction Method of Multipliers 6.5 Minorization-Maximization Algorithms
	Section 7	7 矩阵分解、近似和填充 (4 Hours) 7.1 奇异值分解 7.2 缺失数据和矩阵填充 7 Matrix Decomposition, approximation and completion (4 Hours) 7.1 Singular Value Decomposition 7.2 Missing Data and Matrix Completion
	Section 8	8 稀疏多元方法 (8 Hours) 8.1 稀疏主成分分析 8.2 稀疏典型相关分析 8.3 稀疏判别分析 8.4 稀疏聚类分析 8 Sparse Multivariate Methods (8 Hours) 8.1 Sparse PCA 8.2 Sparse CCA 8.3 Sparse LDA 8.4 Sparse Clustering
11.	**课程考核** **Course Assessment**	
	（1.考核形式 Form of examination；2.分数构成 grading policy；3.如面向本科生开放，请注明区分内容。 If the course is open to undergraduates, please indicate the difference.） **平时作业 40% + 期中考试 20% + 期末报告 40% （考查）**	
12.	**教材及其它参考资料** **Textbook and Supplementary Readings**	
	1. Jianqing Fan, Runze Li, Cunhui Zhang and Hui Zou. (2020). Statistical foundations of Data Science. Chapman and Hall/CRC. 2. Hastie, T., Tibshirani, R., Wainwright, M. (2015). Statistical learning with sparsity: the lasso and generalizations. Chapman and Hall/CRC. 3. Bühlmann, P., Van De Geer, S. (2011). Statistics for high-dimensional data: methods, theory, and applications. Springer Science & Business Media. 4. Philippe Rigollet, Jan Christian Hutter. (2019). High dimensional statistics. MIT lecture notes,	