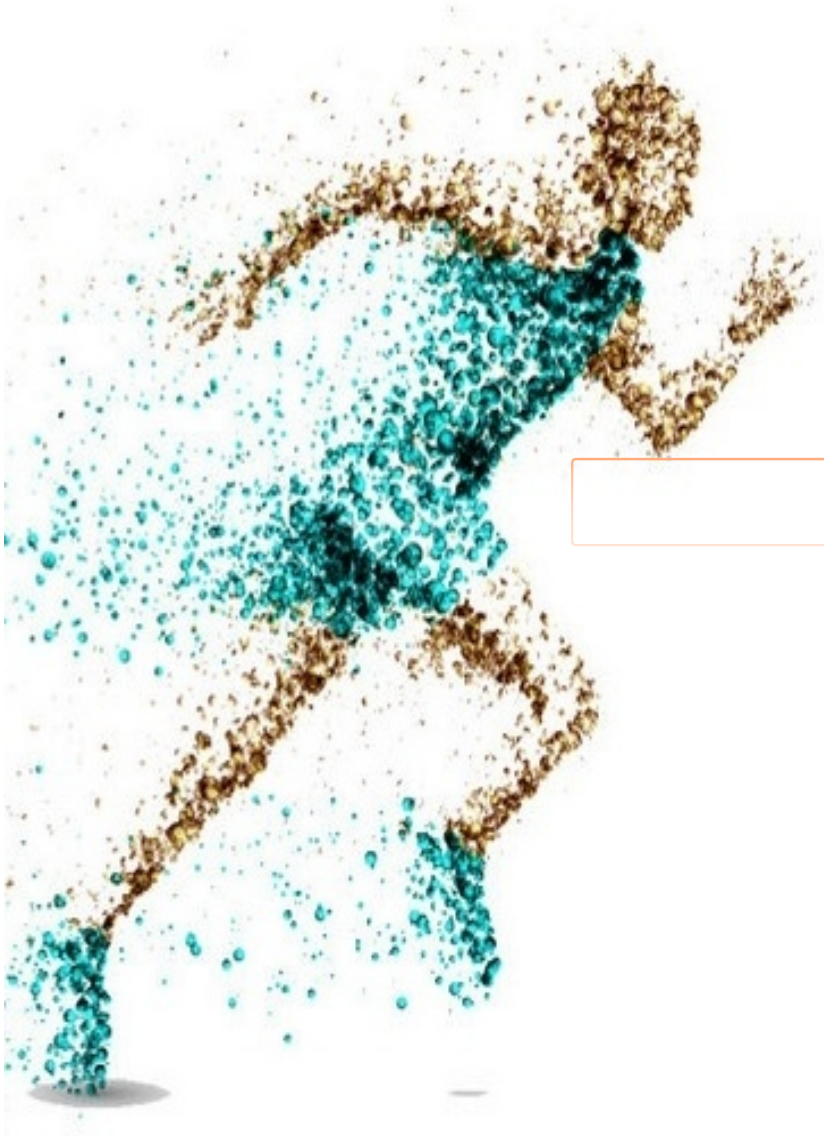


多元统计分析介绍

刘斌
西南财经大学

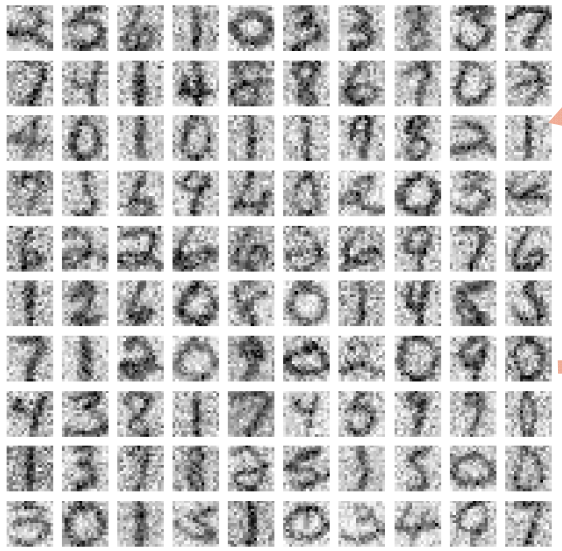
多元统计到底想说一个什么问题

$$\begin{bmatrix} & \end{bmatrix}$$

多个变量或者多个样本之间的关系问题

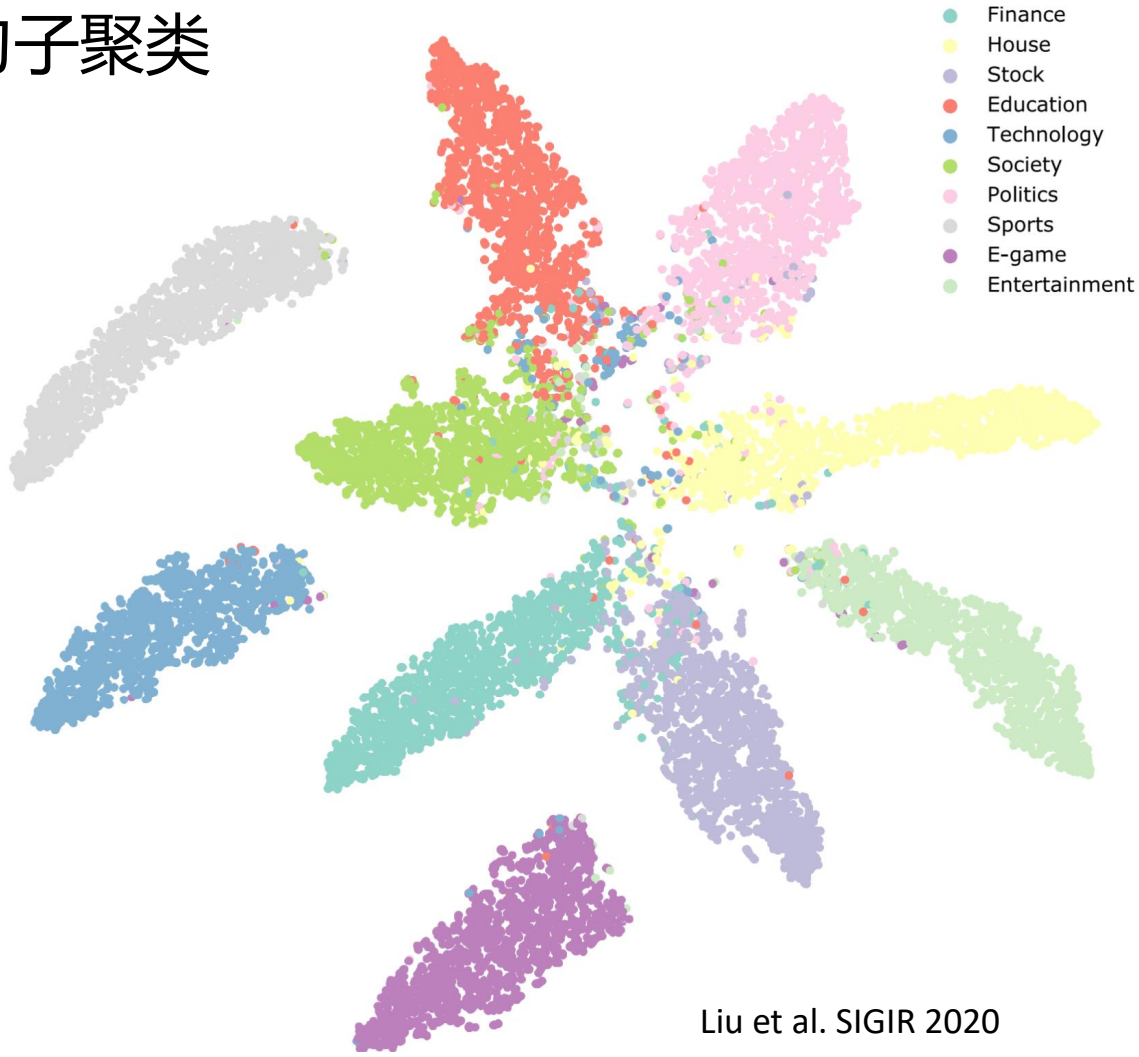
多元统计--降维

- 主成分分析



多元统计--“归”类

- 句子聚类

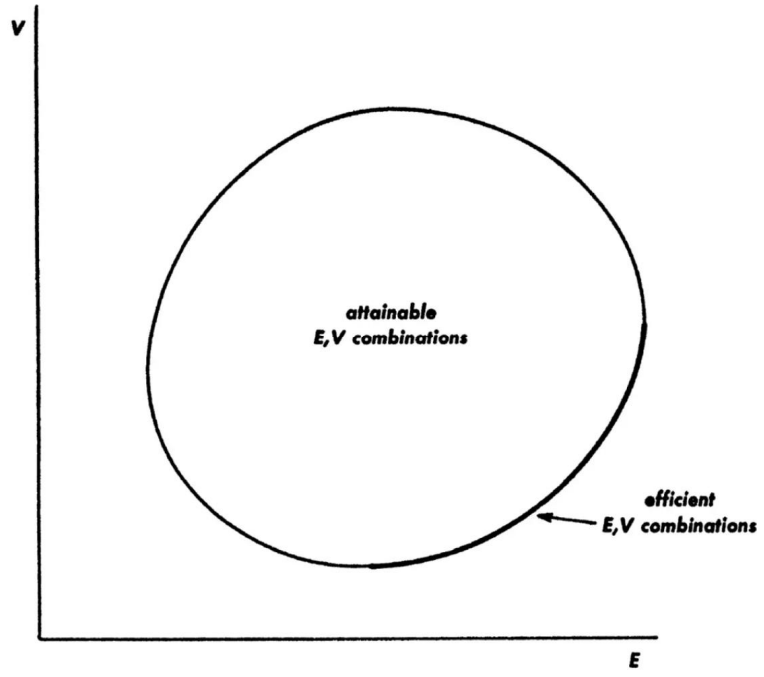


Liu et al. SIGIR 2020

多元统计--因子分析

- 资产定价

$$R_i^e = a_i + \beta_{i,1}\tilde{F}_1 + \cdots + \beta_{i,K}\tilde{F}_K + \varepsilon_i.$$



多元统计--因子分析

- 购房、换购预测

$$\mathbf{x}_i = \mathbf{B}\mathbf{f} + \boldsymbol{\epsilon}_i$$

$$\begin{aligned}\boldsymbol{\Sigma} = \mathrm{Cov}(\mathbf{x}) &= \mathbb{E}[(\mathbf{B}\mathbf{f} + \boldsymbol{\epsilon}_i)(\mathbf{B}\mathbf{f} + \boldsymbol{\epsilon}_i)^\top] \\ &= \underbrace{\mathbf{B}\mathrm{Cov}(\mathbf{f})\mathbf{B}^\top}_{\text{common purchase covariance}} + \underbrace{\mathrm{Cov}(\boldsymbol{\epsilon}|\mathbf{f})}_{\text{specific purchase covariance}}\end{aligned}$$

$$\begin{aligned}\mathcal{N}^{cc}(i) &= g(\widehat{\mathbf{B}}\widehat{\mathbf{B}}^\top) = \{j|[\widehat{\mathbf{B}}\widehat{\mathbf{B}}^\top]_{ij} \neq 0\} \\ \mathbf{m}_i^{l-1} &= \mathrm{Aggre}(\{\mathbf{H}_j^{l-1}|j \in \mathcal{N}^{cc}(i)\}) \\ \mathbf{h}_i^l &= \mathrm{Com}(\mathbf{m}_i^{l-1}, \mathbf{h}_i^{l-1}; \mathbf{W}^l), l = 1, \ldots, L,\end{aligned}$$

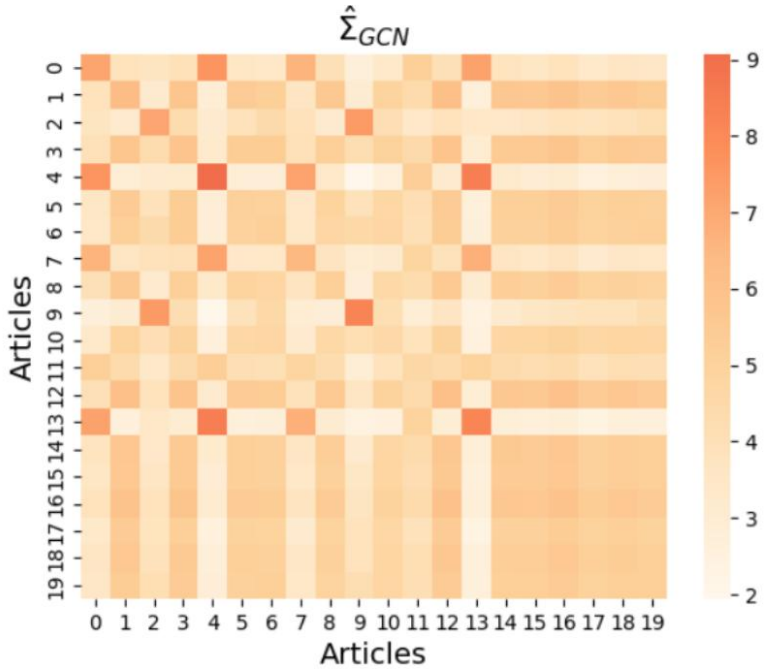
	Accuracy	F1	TPR	TNR	AUC
LR	0.585	0.616	0.671	0.501	0.586
Decision Tree	0.546	0.547	0.555	**0.537**	0.546
GCN	0.571	0.623	0.719	0.423	0.571
GAT	0.570	0.624	0.720	0.424	0.572
ChebyNet	**0.616**	**0.677**	**0.812**	0.423	**0.618**
TAGCN	0.571	0.627	0.729	0.414	0.572
GraphSage	0.573	0.629	0.732	0.417	0.574

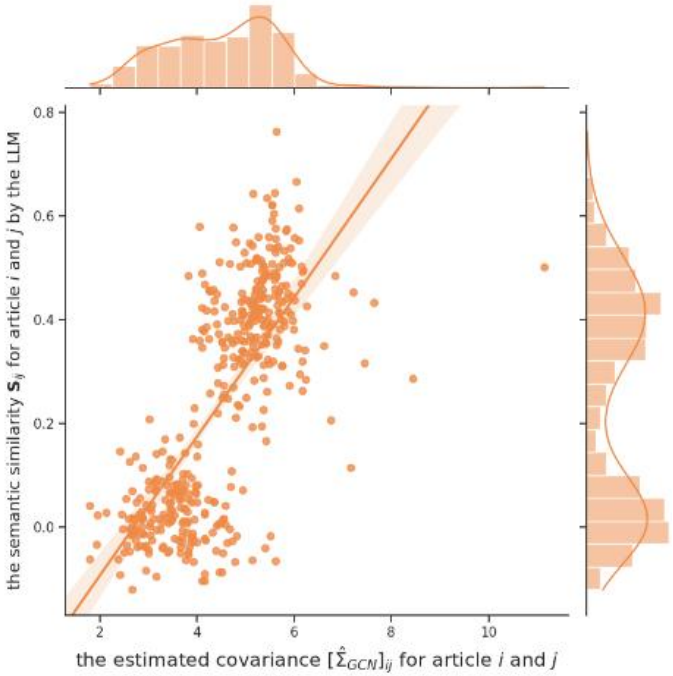
Table 2: The overall performance of predicting future homebuyers for their first purchases. LR is the acronym for Logistic Regression.

Li and Liu， 2023

多元统计--协方差矩阵

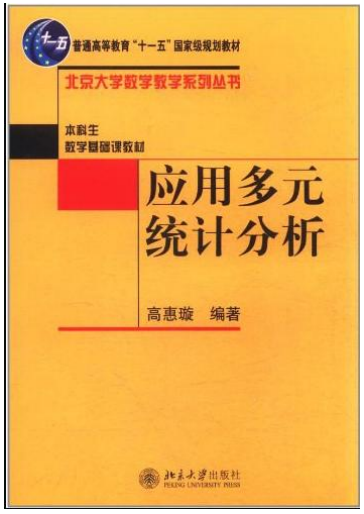
- 图学习





Wang et al. 2023

推荐的教材



https://book.douban.com/subject/1239695/

多元统计研究的问题最接近现代机器学习

它比机器学习讲述地更加系统

如何有效地学习

算法理解 + 编程实现

足量的复习以及练习

每次课至少1/3的时间用于习题练习，每章会留一次课后作业

课程主页：代码、PPT、资料

多元统计分析

优先使用校园网访问
https://github.com/binspage/multiVariateAnaCourse