L_1因子分析及其在宏观经济中的应用

答辩人：蒯强 导师：孔新兵 教授

南京审计大学统计与数学学院

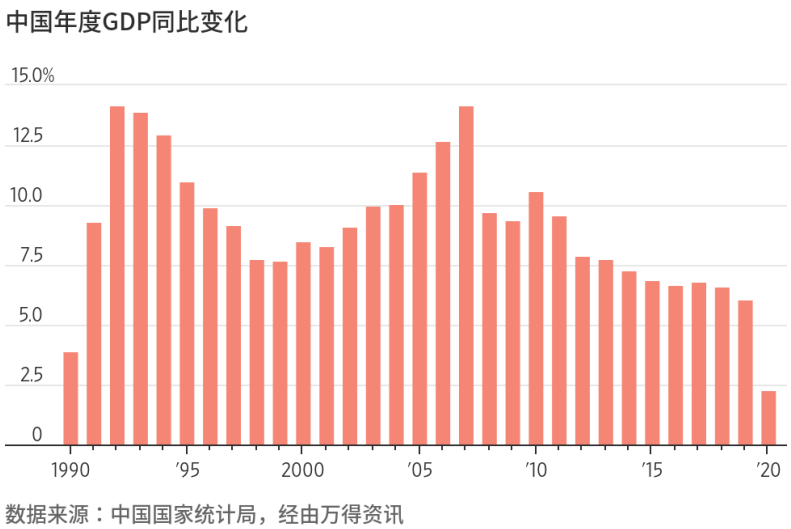
March 17, 2021

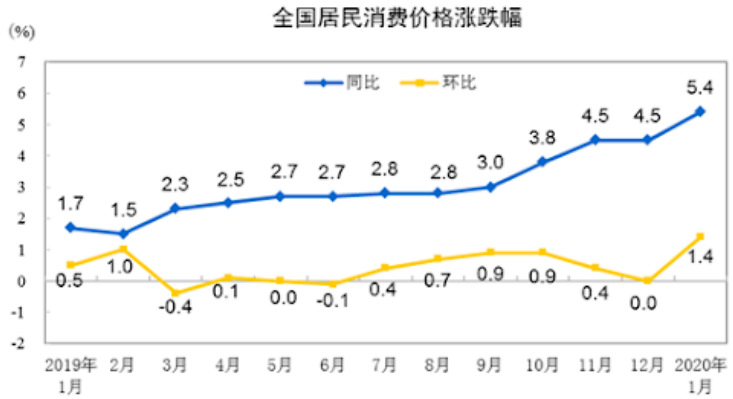
宏观经济和政府调控

- 政府是指导和调控经济运行的主体。国民经济的发展和政府的指导和调控紧密相关。
- 政府部门需要时刻把握国民经济的方方面面，要研究这些经济变量发生变化的原因以及各种经济变量 之间的作用关系。
- 只有这样，政府才能发现经济中存在的问题，并且给出针对性的指导和调控手段。

把控宏观经济数据

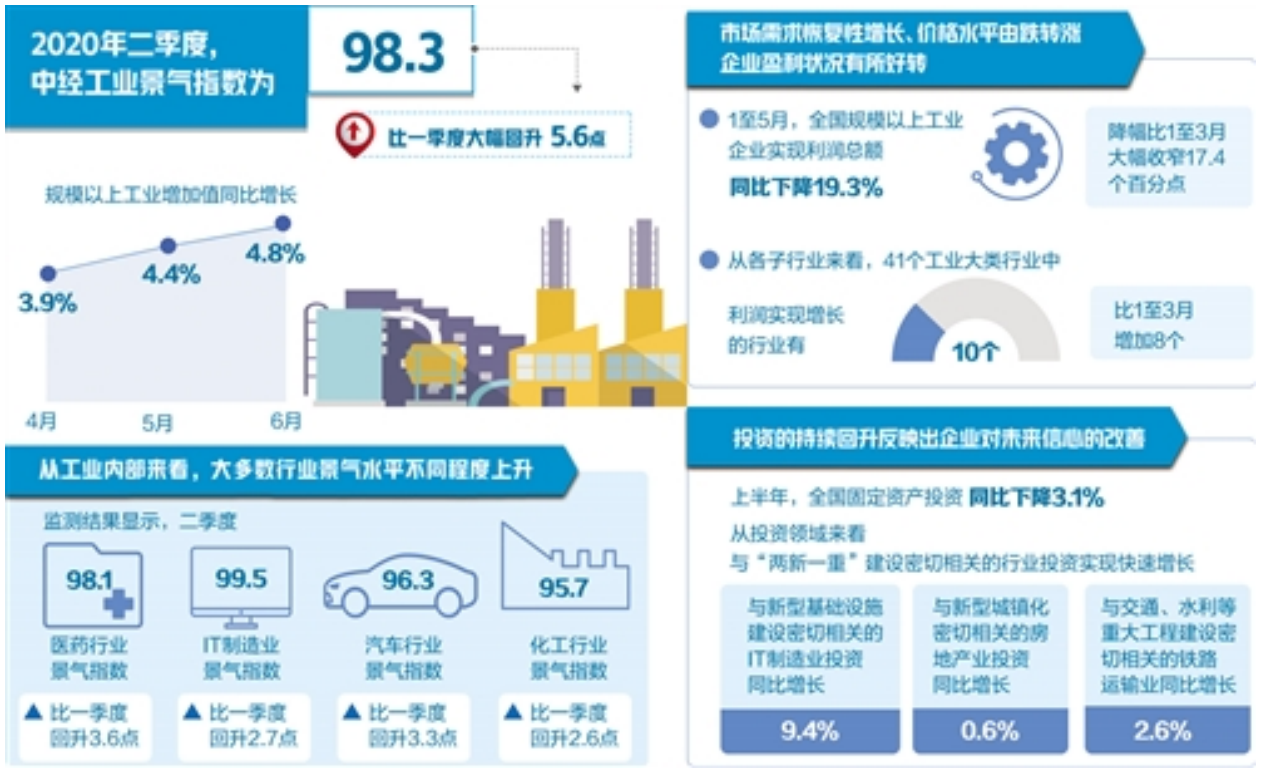
- 政府需要时刻把控宏观经济指标的当前水平。





宏观经济现状分析

- 通过对当前宏观经济指标数据的综合分析，观察当前宏观经济的运行状况，便于行使见机行事的宏观经济政策。



宏观经济预测

- 政府和研究机构常常对重要经济指标做出预测

圖1 IMF對全球經濟預測

國家	2019年	2020年（預測）	2021年（預測）
全　球	2.9%	-3%	5.8%
台　灣	2.7%	-4%	3.5%
新加坡	0.7%	-3.5%	3%
韓　國	2%	-1.2%	3%
香　港	-1.2%	-4.8%	3.9%
日　本	0.7%	-5.2%	3%
中　國	6.1%	1.9%	7.4%
澳　洲	1.8%	-6.7%	6.1%
紐西蘭	2.2%	-7.2%	5.9%

資料來源：IMF, "World Economic Outlook 2020APR"

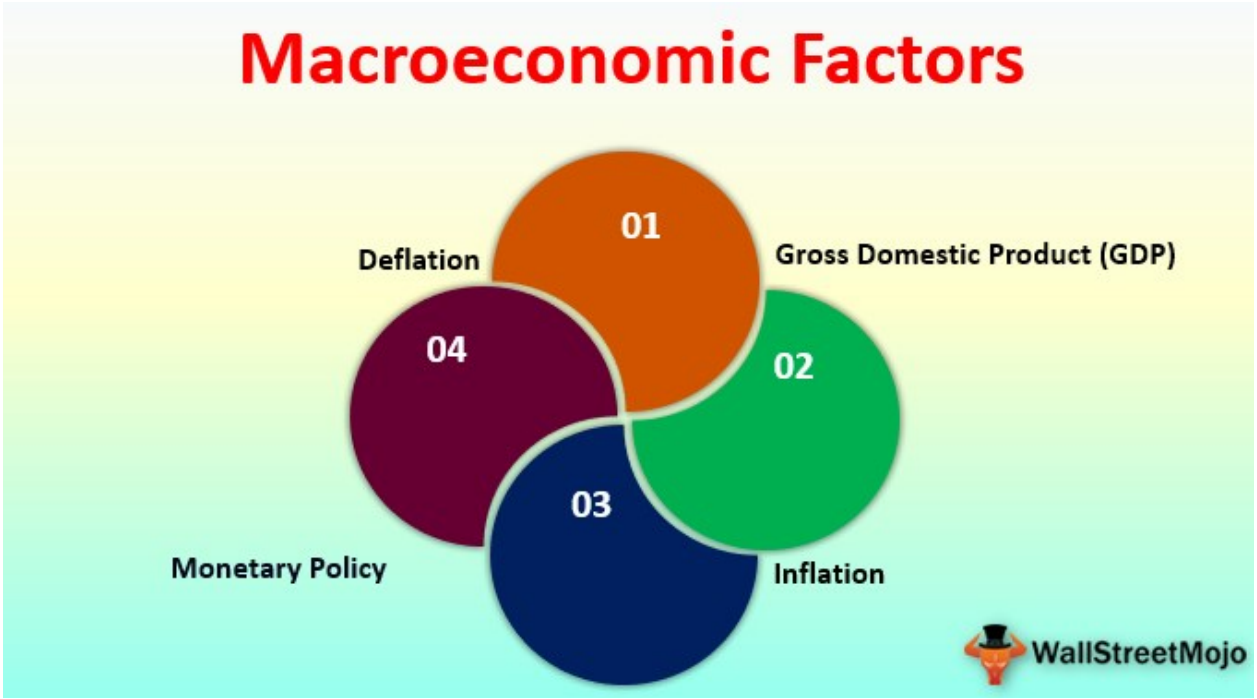
宏观经济预测

	2020年預測值	
	全年	
	預測值	年增率（%）
實質GDP	19,360.24	1.15
民間消費	9,698.94	–0.30
政府消費	2,646.81	2.86
固定資本形成	4,439.10	4.11
民間投資	3,594.48	2.05
公營投資	264.33	21.56
政府投資	580.47	10.90
存貨變動	–11.33	–
貿易順差	2,553.50	–1.49
商品及服務輸出	12,103.47	–3.72
商品及服務輸入	9,567.24	–4.13
物價		
消費者物價指數	102.37	–0.18
躉售物價指數	95.88	–6.19
貨幣供給		
M 1 B	19,851.12	7.43
M 2	47,115.94	4.36

高维宏观经济数据

- 宏观经济指标往往是高维的，难以采用传统模型进行分析。

因子分析是分析高维宏观经济数据的常用手段



扩散指数模型进行宏观经济预测

令y_t为待预测经济变量y在时间t的水平，X_t为p维随机向量，假设(X_t, y_t)服从近似因子模型并且X_t和y_t具有相依性，若X_t, y_t有m维共同因子F_t，即

$$X_t = AF_t + e_t \tag{1}$$

则可以通过式(8)对y_{t+h}进行预测，

$$y_{t+h} = \boldsymbol{\beta}(L)F_t + \boldsymbol{\alpha}(L)y_t + c + e_{t+h} \tag{2}$$

其中滞后算子$\boldsymbol{\beta}(L)$反映了共同因子滞后项的影响，而$\boldsymbol{\alpha}(L)$表示y_t自身的滞后项的影响。

扩散指数模型进行宏观经济预测

Table 1. Simulated Out-of-Sample Forecasting Results: Real Variables, 12-Month Horizon

Forecast method	Industrial production		Personal income		Mfg & trade sales		Nonag. employment	
	Rel. MSE	$\hat{\alpha}$	Rel. MSE	$\hat{\alpha}$	Rel. MSE	$\hat{\alpha}$	Rel. MSE	$\hat{\alpha}$
Benchmark models								
AR	1.00		1.00		1.00		1.00	
LI	.86 (.27)	.57 (.13)	.97 (.21)	.52 (.15)	.82 (.25)	.63 (.17)	.89 (.23)	.56 (.14)
VAR	.97 (.07)	.75 (.68)	.98 (.05)	.68 (.34)	.98 (.04)	.73 (.58)	1.05 (.09)	.22 (.41)
Full dataset ($N = 215$)								
DI-AR, Lag	.57 (.27)	.76 (.13)	.77 (.14)	.76 (.13)	.48 (.25)	.99 (.15)	.91 (.13)	.63 (.18)
DI-AR	.63 (.25)	.71 (.12)	.86 (.16)	.61 (.12)	.57 (.24)	.84 (.18)	.99 (.31)	.51 (.20)
DI	.52 (.26)	.88 (.17)	.86 (.16)	.61 (.12)	.56 (.23)	.94 (.20)	.92 (.26)	.55 (.20)
Balanced panel ($N = 149$)								
DI-AR, Lag	.67 (.25)	.70 (.13)	.82 (.15)	.70 (.13)	.56 (.23)	.91 (.16)	.88 (.14)	.68 (.18)
DI-AR	.67 (.25)	.70 (.12)	.92 (.14)	.57 (.12)	.61 (.23)	.80 (.17)	.88 (.22)	.58 (.17)
DI	.59 (.25)	.81 (.17)	.92 (.14)	.57 (.12)	.57 (.23)	.91 (.18)	.84 (.21)	.62 (.16)
Stacked balance panel								
DI-AR	.65 (.25)	.70 (.12)	.93 (.15)	.56 (.12)	.61 (.22)	.89 (.19)	1.02 (.30)	.49 (.14)
DI	.62 (.25)	.81 (.18)	.93 (.15)	.56 (.12)	.66 (.21)	.85 (.20)	.95 (.24)	.53 (.14)
Full dataset; $m = 1$, $p = BIC$, k fixed								
DI-AR, $k = 1$	1.06 (.11)	.27 (.34)	1.03 (.08)	.34 (.41)	.98 (.06)	.63 (.46)	1.01 (.09)	.49 (.24)
DI-AR, $k = 2$	.63 (.25)	.76 (.14)	.78 (.14)	.77 (.14)	.53 (.24)	.93 (.15)	.77 (.13)	.82 (.15)
DI-AR, $k = 3$	.56 (.26)	.84 (.14)	.77 (.15)	.77 (.13)	.52 (.23)	.99 (.16)	.84 (.14)	.75 (.20)
DI-AR, $k = 4$	.54 (.26)	.85 (.14)	.76 (.15)	.78 (.14)	.51 (.23)	1.01 (.16)	.83 (.15)	.73 (.19)
Full dataset; $m = 1$, $p = 0$, k fixed								
DI, $k = 1$	1.03 (.07)	.30 (.49)	1.01 (.09)	.46 (.34)	.98 (.05)	.67 (.49)	1.01 (.09)	.48 (.24)
DI, $k = 2$	.55 (.25)	.89 (.15)	.78 (.14)	.76 (.13)	.57 (.24)	.95 (.17)	.78 (.13)	.83 (.16)
DI, $k = 3$	.51 (.25)	1.00 (.16)	.77 (.15)	.77 (.13)	.60 (.21)	1.02 (.19)	.84 (.14)	.76 (.19)
DI, $k = 4$	.49 (.25)	1.00 (.16)	.76 (.15)	.78 (.14)	.59 (.22)	1.03 (.20)	.82 (.15)	.75 (.18)
RMSE, AR Model	.049		.027		.045		.017	

扩散指数模型的因子估计

为了得到因子序列，需要对静态因子进行估计，Stock和Watson采用主成分分析作为非参数估计。

2.2 Estimation

In "small-N" dynamic factor models, forecasts are generally constructed using a three-step process (see, e.g., Stock and Watson 1989). First, parametric models are postulated for the joint stochastic process $\{y_{t+h}, X_t, w_t, e_t\}$, and the sample data $\{y_{t+h}, X_t, w_t\}_{t=1}^{T-h}$ are used to estimate the parameters of this process, typically using a Gaussian Maximum likelihood estimator (MLE). Next, these estimated parameters are used in signal extraction algorithms to estimate the unknown value of F_T. Finally, the forecast of y_{T+h} is constructed using this estimated value of the factor and the estimated parameters. When N is large, this process requires estimating many parameters using iterative nonlinear methods, which can be computationally prohibitive. We therefore take a different approach and estimate the dynamic factors nonparametrically using the method of principal components.

Consider the nonlinear least squares objective function,

$$V(\widetilde{F}, \widetilde{\Lambda}) = (NT)^{-1} \sum_i \sum_t (x_{it} - \tilde{\lambda}_i \widetilde{F}_t)^2, \qquad (5)$$

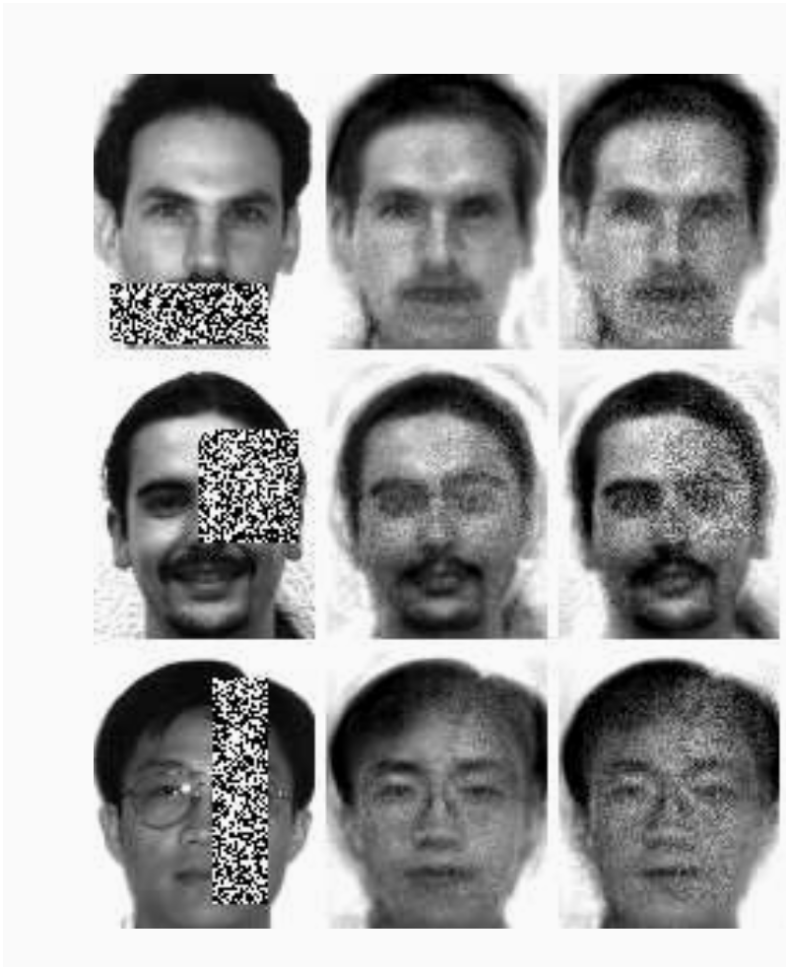
written as a function of hypothetical values of the factors $(\widetilde{F})$ and factor loadings $(\widetilde{\Lambda})$, where $\widetilde{F} = (\widetilde{F}_1 \widetilde{F}_2 \ldots \widetilde{F}_T)'$ and $\tilde{\lambda}_i$ is the ith row of $\widetilde{\Lambda}$. Let $\widehat{F}$ and $\widehat{\Lambda}$ denote the minimizers of $V(\widetilde{F}, \widetilde{\Lambda})$. After concentrating out $\widehat{F}$, minimizing (5) is equivalent to

主成分分析法

- 主成分分析法是常见的降维方法。

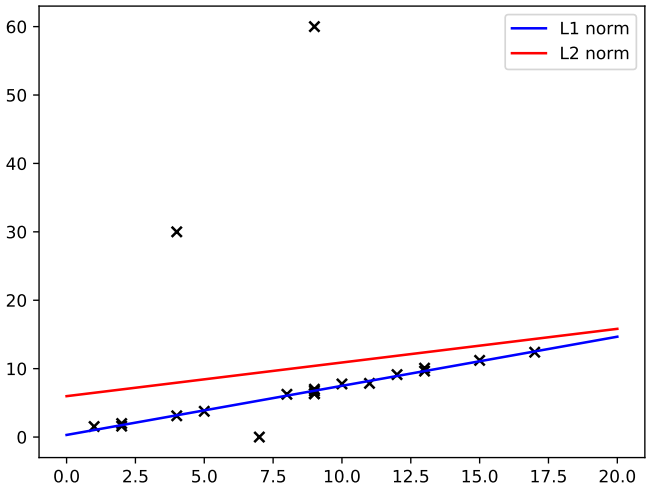
L_2主成分分析不稳健

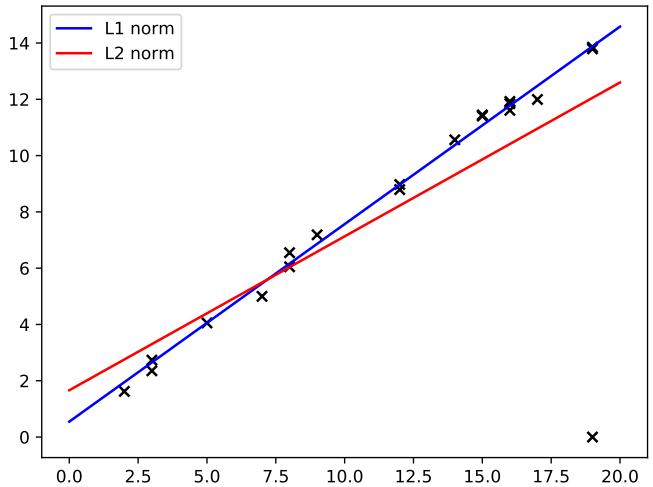
- L_2主成分分析对离群值不稳健。



L_1和L_2的稳健性对比

- L_1范数能够提供更好的稳健性。





L_2主成分分析问题描述

- L_2主成分分析问题描述

$$P_1:\ \hat{A}_{p\times m},\hat{F}_{m\times n}=\arg\min_{A,F}\|X-AF\|_{L_2}=\arg\min_{A,F}\sum_{i=1}^{p}\sum_{j=1}^{m}(x_{ij}-a_i^Tf_i)^2, \tag{3}$$

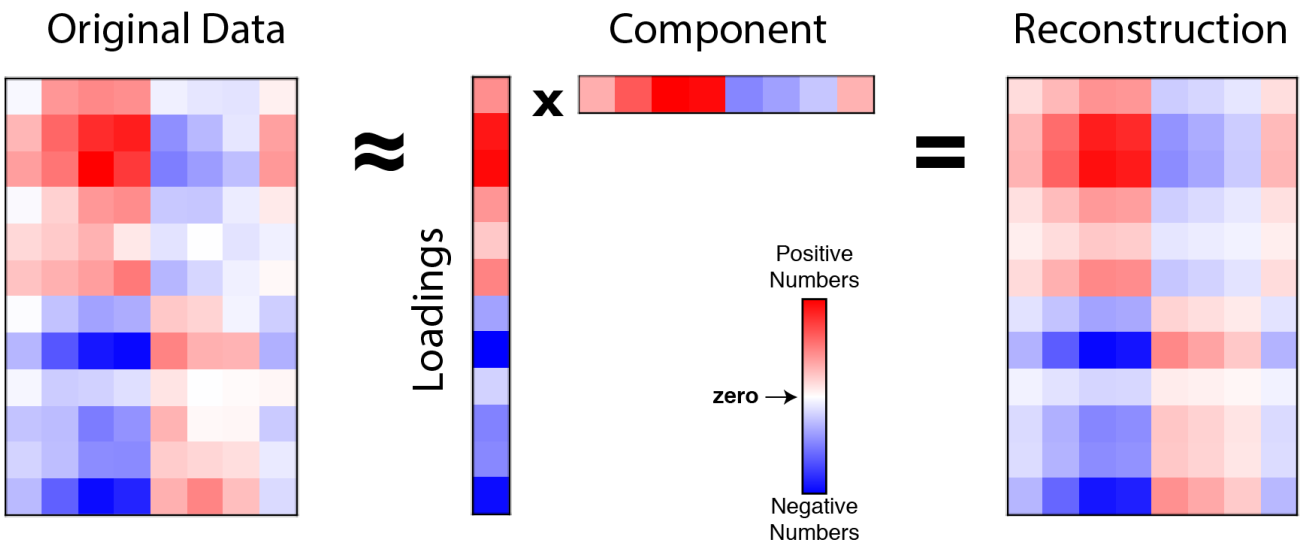
其中X为$p\times n$的高维数据矩阵，A的列构成了X的m维线性子空间的基，这个子空间也称为特征空间。
F为一系数矩阵，给出了X各列元素在特征空间中的坐标，根据矩阵投影理论，在给定A的条件下，$F=A^TX$。
问题P_1可以解释为，需要找到一个合适的投射矩阵，使得数据在低维的投影上升回高维空间后和原矩阵各元素的误差平方和最小。

对于问题P_1常用奇异值分解法求解。同样地我们也可考虑其对偶问题P_2，

$$P_2:\ \hat{A}=\arg\max_{A}\|A^TX\|_{L_2},\text{其中 }A^TA=I_m. \tag{4}$$

问题P_2可以理解为，需要找到一个合适的投射矩阵，使得数据在低维空间的投影有最大的方差。在统计学上，数据的方差反映了数据中信息的多少，因此在特征空间中选取方差最大的方向作为主成分是很恰当的。

重构误差最小化


Original Data
≈
Loadings
x
Component
=
Reconstruction
Positive Numbers
zero
Negative Numbers

L_1主成分分析问题描述

- L_1主成分分析问题描述

$$P_3:\ \hat{A}_{p\times m},\hat{F}_{m\times n}=\arg\min_{A,F}\|X-AF\|_{L_1}=\arg\min_{A,F}\sum_{i=1}^{p}\sum_{j=1}^{m}|x_{ij}-a_i^Tf_i|. \tag{5}$$

$$P_4:\ \hat{A}=\arg\max_{A}\|A^TX\|_{L_1}=\arg\max_{A}\sum_{i=1}^{n}\sum_{k=1}^{m}|\sum_{j=1}^{p}a_{jk}x_{ij}|,\ \text{其中}A^TA=I_m \tag{6}$$

一种L_1主成分分析的交替凸优化算法

- Qifake & Kande于2005年提出一种交替凸优化算法。

交替凸优化求解L_1主成分分析 (ACP, Alternate Convex Programming)

1.初始化：给出A，$\mathbf{\Sigma}$的初始值$A^{(0)}$，$\mathbf{\Sigma}^{(0)} = I$，（其中$\mathbf{\Sigma}$为一对角矩阵，I为单位矩阵）；
2.交替凸优化：对于迭代次数$t = 1, \ldots, \tau$：

$$F^{(t)} = \arg\min_{F} \|X - A^{(t-1)}\mathbf{\Sigma}^{(t-1)}F^{T}\|_{L_1}$$

$$A^{(t)} = \arg\min_{A} \|X - A\mathbf{\Sigma}^{(t-1)}(F^{(t)})^{T}\|_{L_1}$$

$$\text{归一化：} \begin{cases} N_a = \mathrm{diag}((A^{(t)})^{T}A^{(t)}) \\ N_f = \mathrm{diag}((F^{(t)})^{T}F^{(t)}) \\ F^{(t)} \leftarrow F^{(t)}N_f^{-1} \\ A^{(t)} \leftarrow A^{(t)}N_a^{-1} \\ \mathbf{\Sigma}^{(t)} \leftarrow N_a\mathbf{\Sigma}^{(t-1)}N_f \end{cases}$$

3.输出结果：$A \leftarrow A^{(\tau)}\mathbf{\Sigma}^{1/2}$，对A进行QR分解取正交矩阵得到$\hat{A}$；$\hat{F} \leftarrow \hat{A}^{T}X$。

模拟实验准备

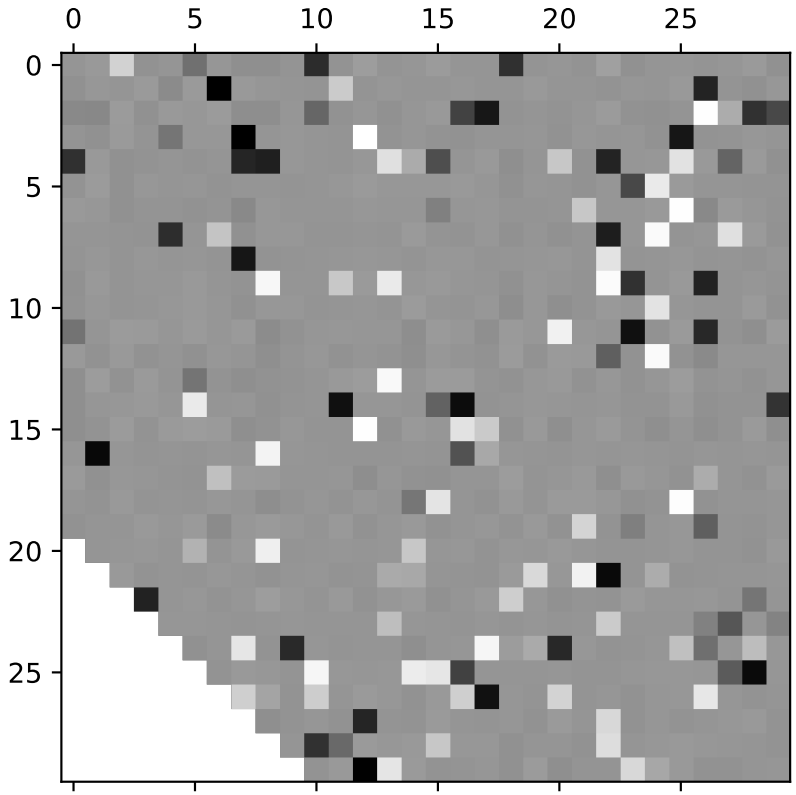
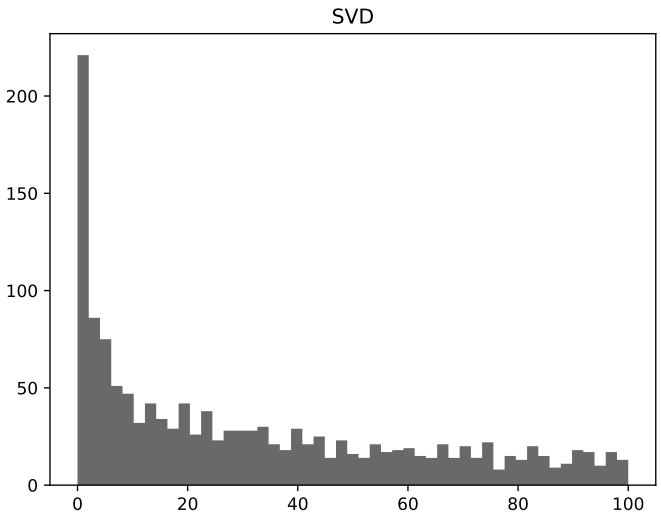
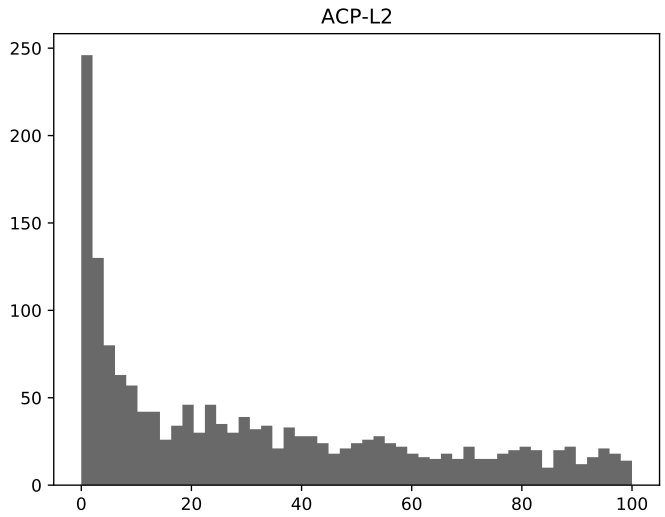
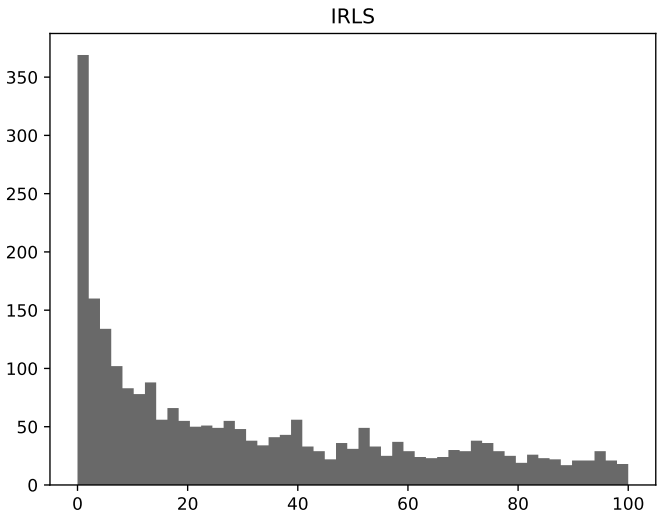


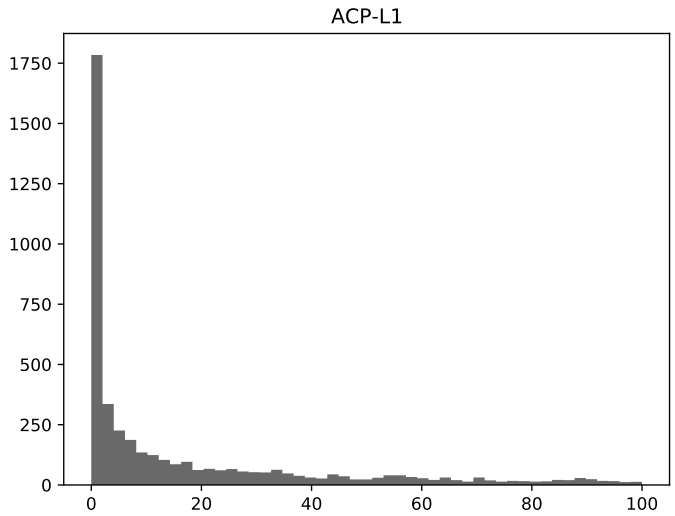
Figure: 30 × 30的模拟矩阵

模拟实验结果


SVD
0
50
100
150
200
0
20
40
60
80
100

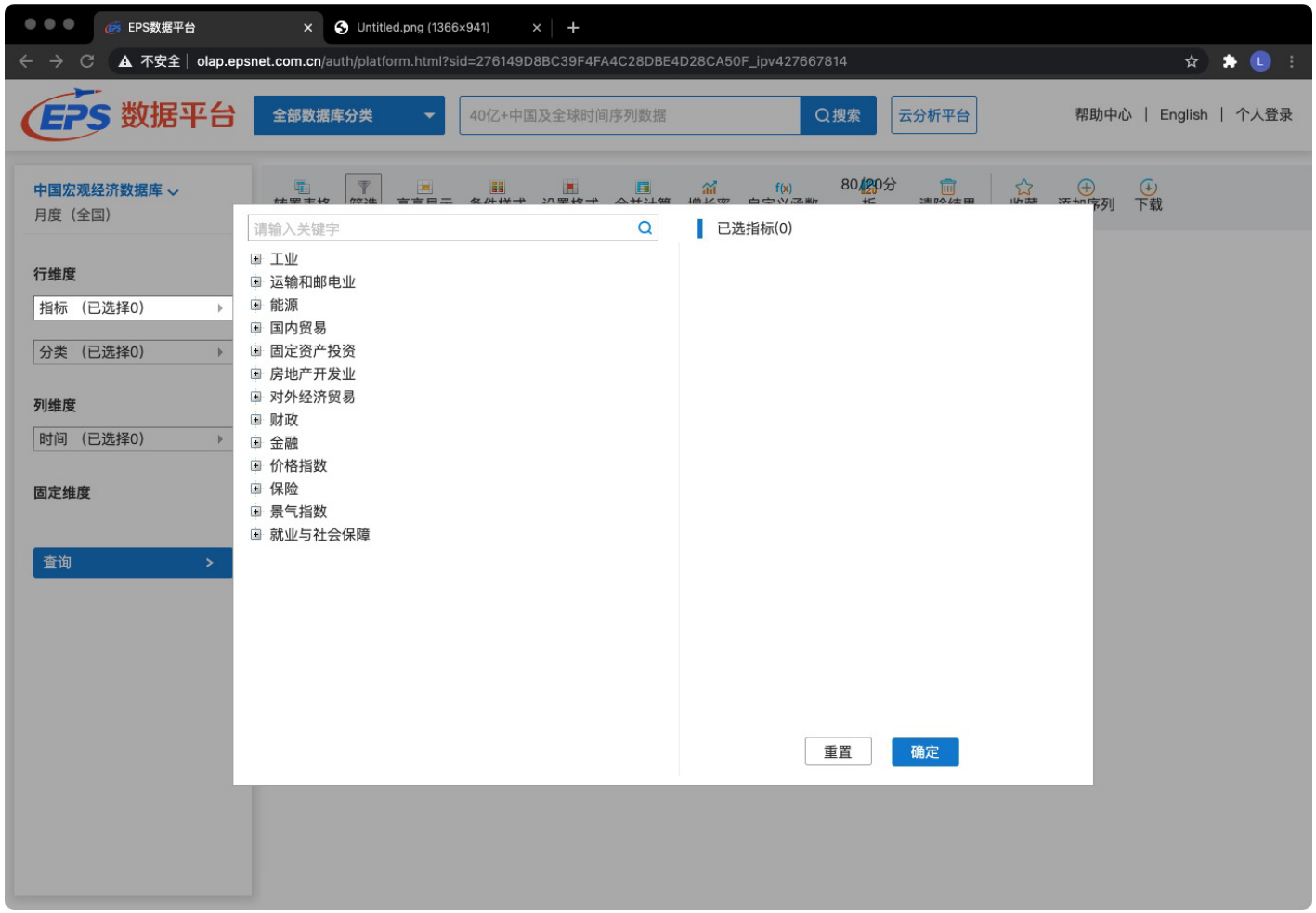

ACP-L2
0
50
100
150
200
250
0
20
40
60
80
100


IRLS
0
50
100
150
200
250
300
350
0
20
40
60
80
100


ACP-L1
0
250
500
750
1000
1250
1500
1750
0
20
40
60
80
100

基于国内主要月度宏观经济数据的实证研究

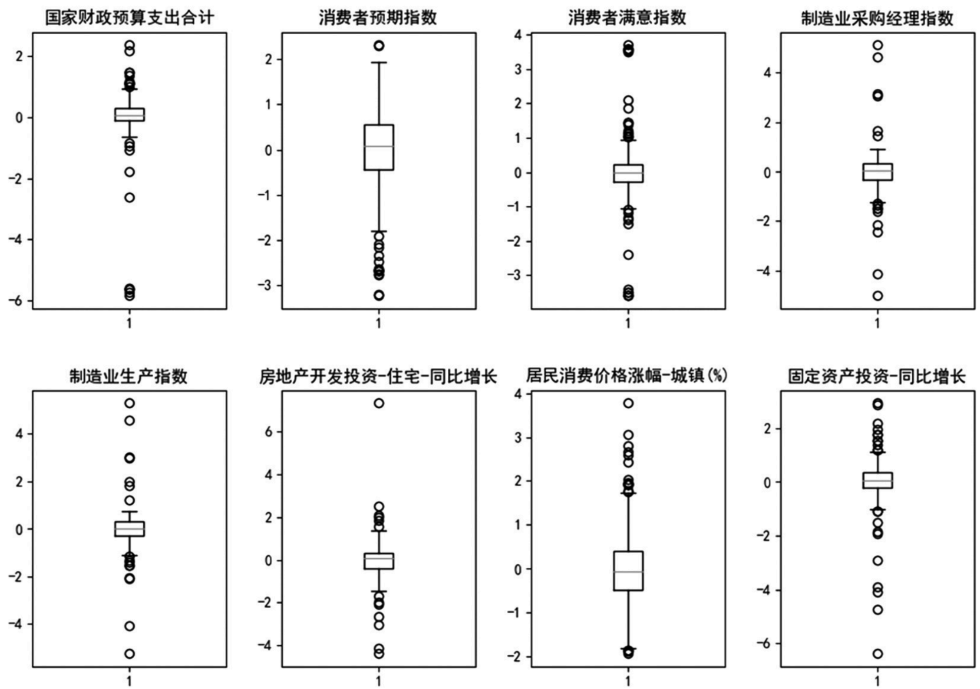
- 数据来源：EPS数据平台-全国月度宏观经济数据



基于国内主要月度宏观经济数据的实证研究

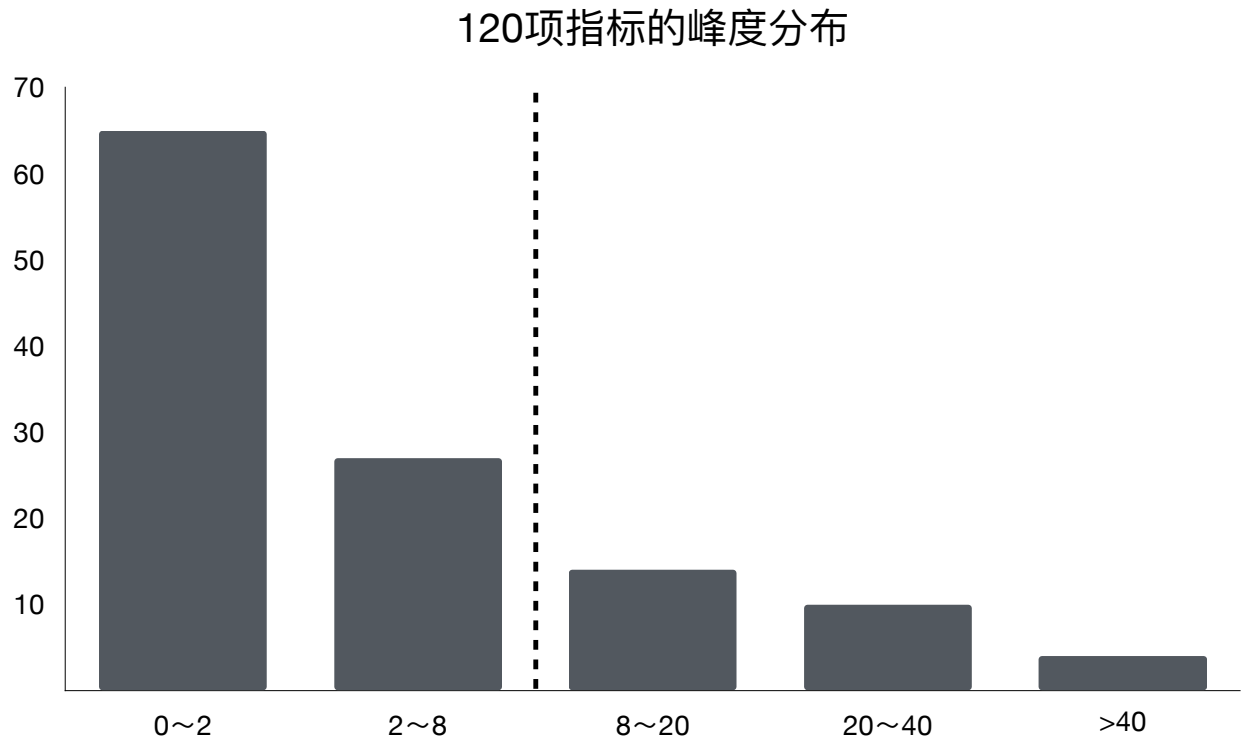
- 宏观经济数据的重尾性

部分指标箱形图



基于国内主要月度宏观经济数据的实证研究

- 宏观经济数据的重尾性



利用扩散指数模型进行预测

- 扩散指数模型预测

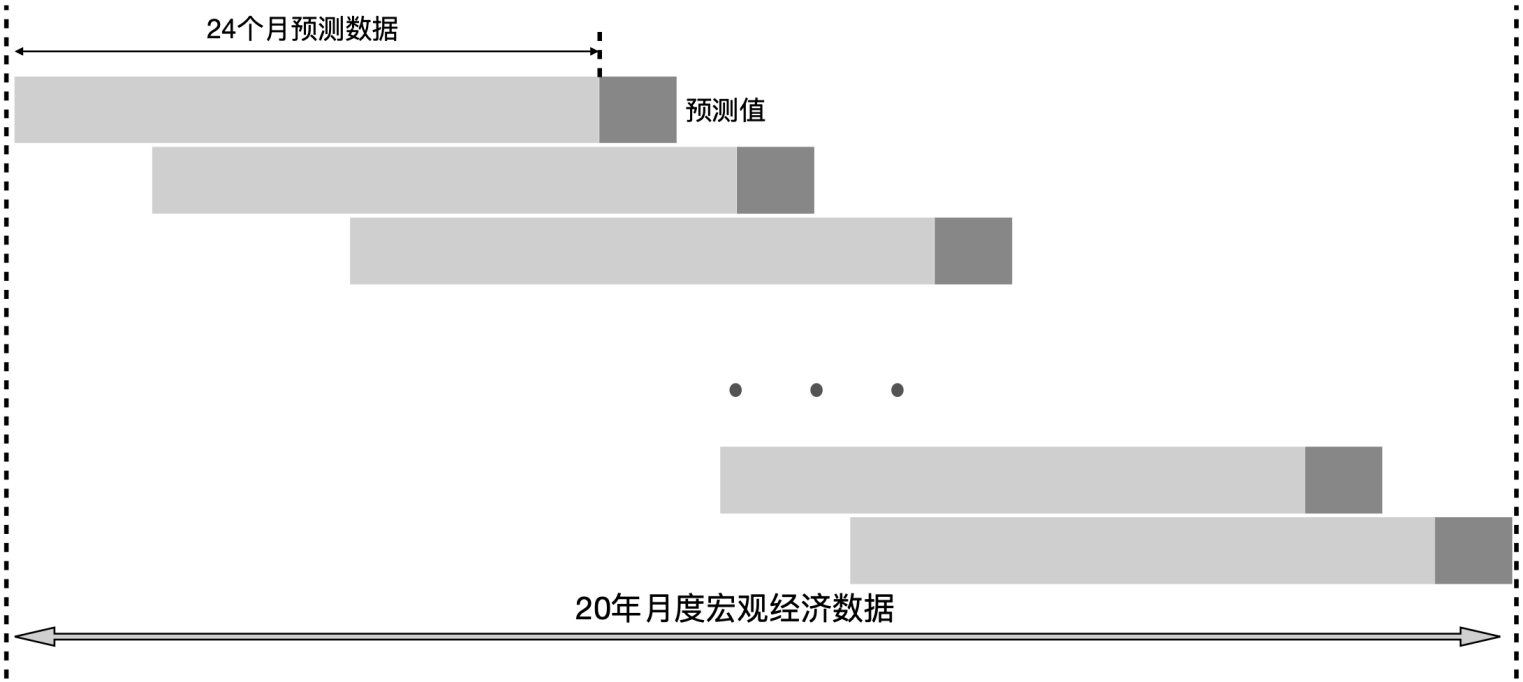
$$X_t = AF_t + e_t \tag{7}$$

$$y_{t+h} = \boldsymbol{\beta}(L)F_t + \boldsymbol{\alpha}(L)y_t + c + e_{t+h} \tag{8}$$

- 因子个数选择，Bai&Ng信息准则。

利用扩散指数模型进行预测

- 滑动窗口预测



实证结果

- 预测结果表明：L_1因子在短期预测中比L_2因子更准确。长期预测，表现也良好。

Table: 向前一个月预测结果

	MSE（L1 PCA）	MSE（IM）	MAE（L1 PCA）	MAE（IM）	MPAE（L1 PCA）	MPAE（IM）
消费者满意指数*	0.81	1.49	0.87	1.67	0.43	1.55
工业生产者出厂价格指数*	0.70	1.98	0.75	1.54	0.96	1.60
货币供应量M2*	0.76	1.25	0.90	1.19	0.45	0.97
固定资产投资总额*	0.89	1.03	0.97	0.99	0.81	1.32
房地产开发投资总额*	0.79	0.98	0.95	1.00	0.91	1.20
社会消费品零售总额*	0.84	1.25	0.87	1.12	0.45	1.17
制造业采购经理指数	1.24	1.69	1.06	1.43	0.90	1.50
住宅新开工面积总数*	0.89	2.40	0.85	1.98	0.45	2.22
股票流通市值*	0.99	3.99	0.98	2.84	0.81	4.51
消费者信心指数*	0.93	1.01	0.98	0.99	0.98	1.56

交替凸优化算法的计算性能问题

- 交替凸优化算法需要求解多个最小绝对值回归问题，在变量较多情况下，线性规划算法计算性能较差。

$$\boldsymbol{F}^{(t)} = \arg\min_{\boldsymbol{F}} \|\boldsymbol{X} - \boldsymbol{A}^{(t-1)}\boldsymbol{F}\|_{L_1} \tag{2.12}$$

$$\boldsymbol{A}^{(t)} = \arg\min_{\boldsymbol{A}} \|\boldsymbol{X} - \boldsymbol{A}\boldsymbol{F}^{(t)}\|_{L_1} \tag{2.13}$$

我们改写式(2.12)中的目标函数，

$$\xi(\boldsymbol{F}) = \|\boldsymbol{X} - \boldsymbol{A}^{(t-1)}\boldsymbol{F}\|_{L_1} = \sum_{j=1}^{n} |\boldsymbol{x}_j - \boldsymbol{A}^{(t-1)}\boldsymbol{f}_j| \tag{2.14}$$

其中 $\boldsymbol{x}_j$ 是矩阵 $\boldsymbol{X}$ 的第 j 列，$\boldsymbol{f}_j$ 是 $\boldsymbol{F}$ 的第 j 列。于是式(2.12)问题可以分解为 n 个独立的子优化问题，求解相应的 $\boldsymbol{f}_j$：

$$\boldsymbol{f}_j^{(t)} = \arg\min_{\boldsymbol{\theta}} |\boldsymbol{A}^{(t-1)}\boldsymbol{\theta} - \boldsymbol{x}_j| \tag{2.15}$$

同样地，(2.13)可以转化为下面的 p 个独立的子优化问题，

$$\boldsymbol{a}_i^{(t)} = \arg\min_{\boldsymbol{\theta}} |\boldsymbol{x}_i - \boldsymbol{\theta}^T\boldsymbol{F}^{(t-1)}| \tag{2.16}$$

其中 $\boldsymbol{a}_i$ 为 $\boldsymbol{A}$ 的第 i 行，而 $\boldsymbol{x}_i$ 为 $\boldsymbol{X}$ 的第 i 行。

最小绝对值回归的优化1: 聚类——迭代拆解算法

- Park等人于2016年提出一种聚类——迭代拆解算法，通过减小问题规模来提升计算性能。

$$\boldsymbol{F}^{(t)} = \arg\min_{\boldsymbol{F}} \|\boldsymbol{X} - \boldsymbol{A}^{(t-1)}\boldsymbol{F}\|_{L_1} \tag{2.12}$$

$$\boldsymbol{A}^{(t)} = \arg\min_{\boldsymbol{A}} \|\boldsymbol{X} - \boldsymbol{A}\boldsymbol{F}^{(t)}\|_{L_1} \tag{2.13}$$

我们改写式(2.12)中的目标函数，

$$\xi(\boldsymbol{F}) = \|\boldsymbol{X} - \boldsymbol{A}^{(t-1)}\boldsymbol{F}\|_{L_1} = \sum_{j=1}^{n} |\boldsymbol{x}_j - \boldsymbol{A}^{(t-1)}\boldsymbol{f}_j| \tag{2.14}$$

其中 $\boldsymbol{x}_j$ 是矩阵 $\boldsymbol{X}$ 的第 j 列，$\boldsymbol{f}_j$ 是 $\boldsymbol{F}$ 的第 j 列。于是式(2.12)问题可以分解为 n 个独立的子优化问题，求解相应的 $\boldsymbol{f}_j$:

$$\boldsymbol{f}_j^{(t)} = \arg\min_{\boldsymbol{\theta}} |\boldsymbol{A}^{(t-1)}\boldsymbol{\theta} - \boldsymbol{x}_j| \tag{2.15}$$

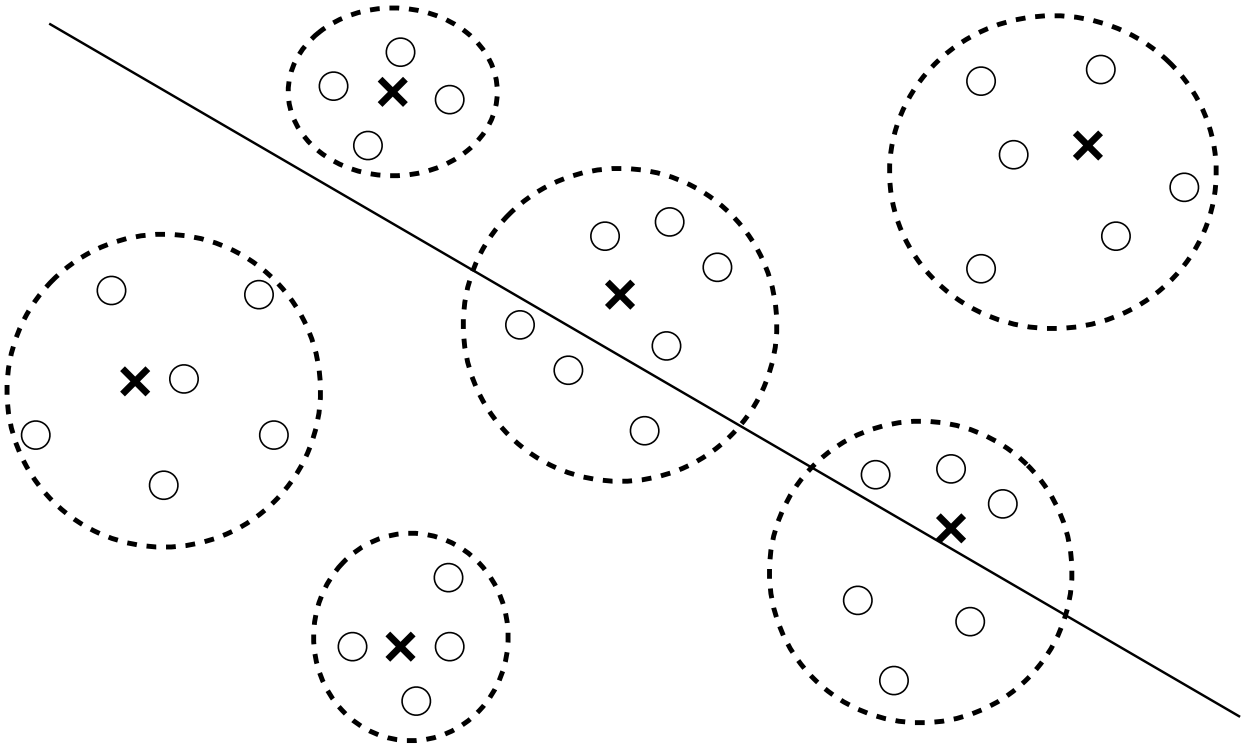
同样地，(2.13)可以转化为下面的 p 个独立的子优化问题，

$$\boldsymbol{a}_i^{(t)} = \arg\min_{\boldsymbol{\theta}} |\boldsymbol{x}_i - \boldsymbol{\theta}^T\boldsymbol{F}^{(t-1)}| \tag{2.16}$$

其中 $\boldsymbol{a}_i$ 为 $\boldsymbol{A}$ 的第 i 行，而 $\boldsymbol{x}_i$ 为 $\boldsymbol{X}$ 的第 i 行。

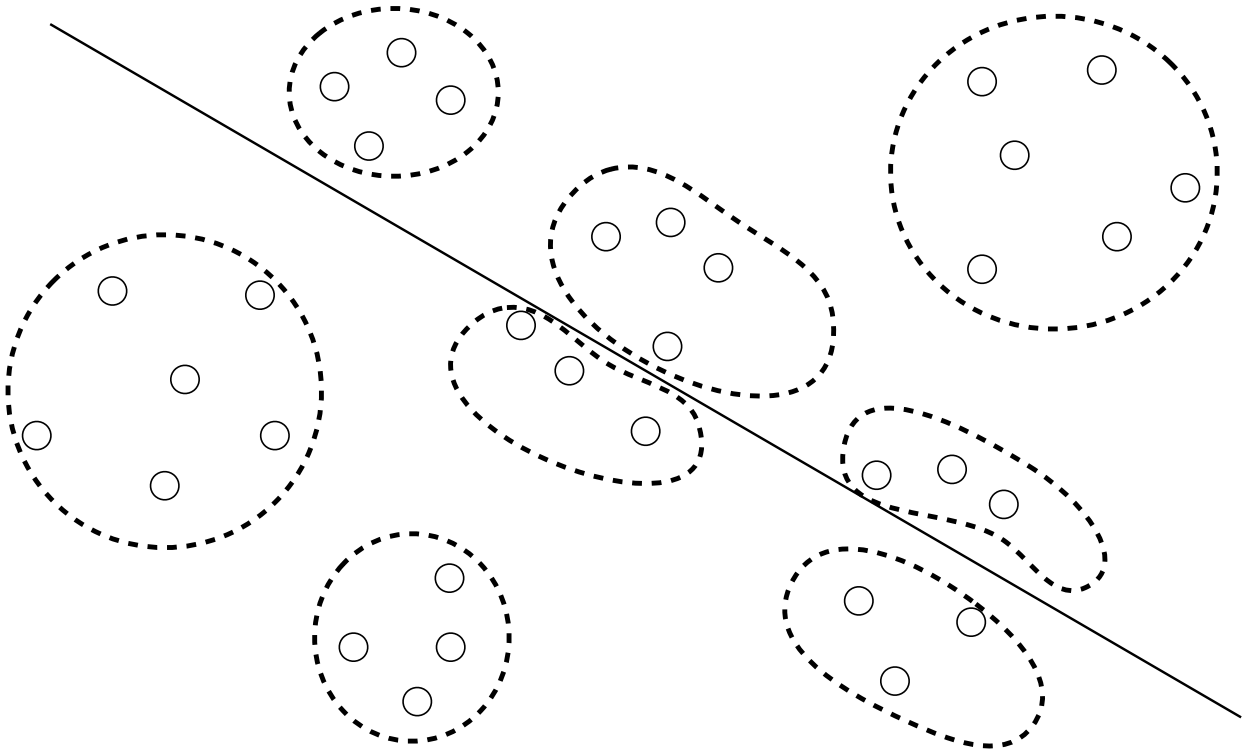
最小绝对值回归的优化1: 聚类——迭代拆解算法

- 初始聚类，使用任意聚类方法，我们这里使用K-means。



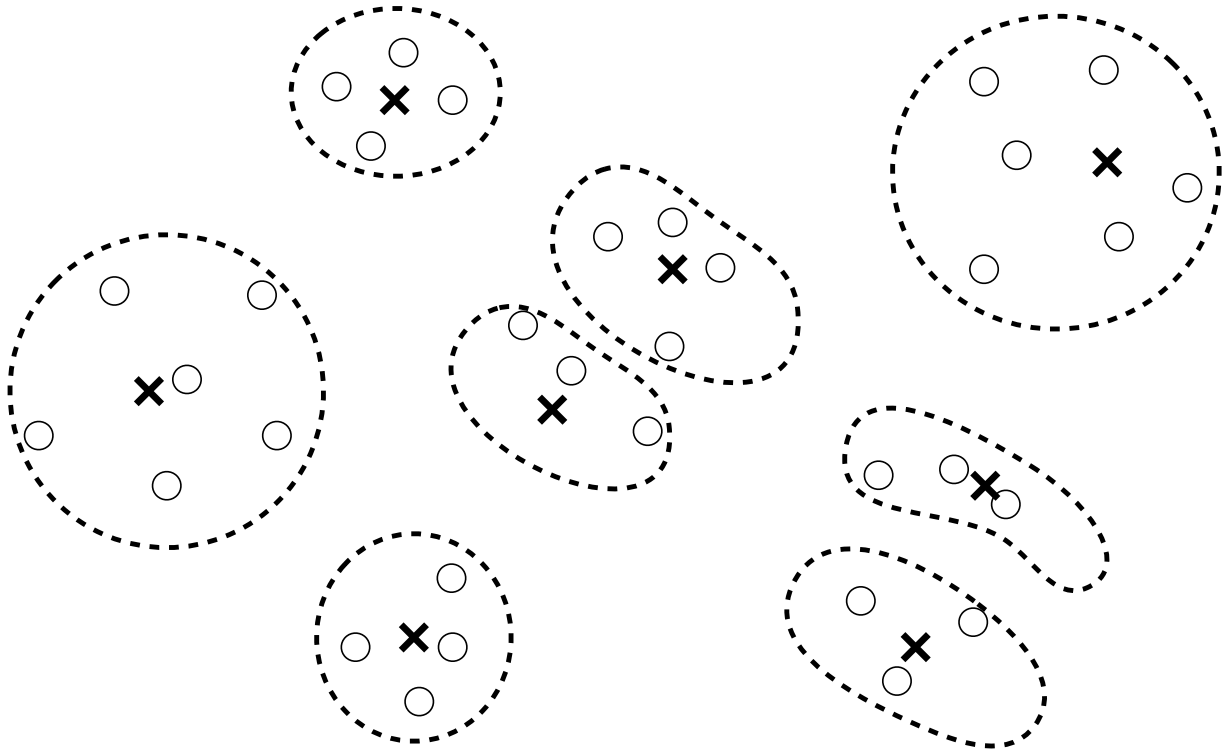
最小绝对值回归的优化1: 聚类——迭代拆解算法

- 计算β，按规则对聚类进行拆解。



最小绝对值回归的优化1: 聚类——迭代拆解算法

- 在新的聚类上重新计算β。



最小绝对值回归的优化1: 聚类——迭代拆解算法

- 可以证明，该方法最终获得原问题的最优解（和在全部数据上解原优化问题有相同的解）。

下面证明最后一次迭代的解 $\hat{\boldsymbol{\beta}}^{(\tau)}$ 就是(3.4)的解 $\boldsymbol{\beta}^*$，

$$\begin{aligned}
\xi^* &= \sum_{i\in I}|y_i - \sum_{j\in J} x_{ij}\boldsymbol{\beta}_j^*| = \sum_{k\in K^{(t)}}\sum_{i\in C_k^{(t)}}|y_i - \sum_{j\in J} x_{ij}\boldsymbol{\beta}_j^*| \\
&\geq \sum_{k\in K^{(t)}}|\sum_{i\in C_k^{(t)}}(y_i - \sum_{j\in J} x_{ij}\boldsymbol{\beta}_j^*)| = \sum_{k\in K^{(t)}}|C_k^{(t)}||y_k^{(t)} - \sum_{j\in J} x_{kj}^{(t)}\boldsymbol{\beta}_j^*| \\
&\geq \sum_{k\in K^{(t)}}|C_k^{(t)}||y_K^{(t)} - \sum_{j\in J} x_{kj}^{(t)}\hat{\boldsymbol{\beta}}_j^{(t)}| = \sum_{k\in K^{(t)}}|\sum_{i\in C_k^{(t)}} y_i - \sum_{i\in C_k^{(t)}}\sum_{j\in J} x_{ij}\hat{\boldsymbol{\beta}}_j^{(t)}| \\
&= \sum_{k\in K^{(t)}}\sum_{i\in C_k^{(t)}}|y_i - \sum_{j\in J} x_{ij}\hat{\boldsymbol{\beta}}_j^{(t)}| = \sum_{i\in I}|y_i - \sum_{j\in J} x_{ij}\hat{\boldsymbol{\beta}}_j^{(t)}| = \xi^{(t)}
\end{aligned}$$

因为 $\hat{\boldsymbol{\beta}}_j^{(t)}$ 是(3.4)的可行解，又显然 $\xi^* \leq \xi^{(t)}$，这就证明了 $\xi^* = \xi^{(t)}$，注意到 $\sum_{k\in K^{(t)}}|C_k^{(t)}||y_K^{(t)} - \sum_{j\in J} x_{kj}\hat{\boldsymbol{\beta}}_j^{(t)}|$ 就是 $F^{(t)}$，因此 $\xi^{(t)} = F^{(t)}$。因此最后一次迭代 $F^{(\tau)}$ 的最优解 $\hat{\boldsymbol{\beta}}^{(\tau)}$ 就是原问题的最优解。