時系列信号の次元圧縮

西井淳

2022 年 2 月 17 日

1　目的

取得した時系列信号 $\boldsymbol{e}(t) = (e_1(t), e_2(t), \ldots, e_N(t))^t$ をできるだけ少ない基底 $\{\boldsymbol{p}_1, \boldsymbol{p}_2, \ldots, \boldsymbol{p}_M\}$, $\boldsymbol{p}_i = (p_{i1}, p_{i2}, \ldots, p_{iN})^t$, $(i = 1, 2, \ldots, M)$ で表す。

$$\begin{pmatrix} e_1(t) \\ e_2(t) \\ \vdots \\ e_N(t) \end{pmatrix} = c_1(t) \begin{pmatrix} p_{11} \\ p_{12} \\ \vdots \\ p_{1N} \end{pmatrix} + c_2(t) \begin{pmatrix} p_{21} \\ p_{22} \\ \vdots \\ p_{2N} \end{pmatrix} + \cdots \tag{1}$$

$$\boldsymbol{e}(t) = c_1(t)\boldsymbol{p}_1 + c_2(t)\boldsymbol{p}_2 + \cdots \tag{2}$$

ここで，$c_i(t)$ は基底 $\{\boldsymbol{p}_i\}$ を座標軸とする空間で $\boldsymbol{e}(t)$ を表した場合の座標値を表す。この基底は，主成分分析では主成分軸，因子分析では因子に対応する。各基底ベクトルは単位ベクトル $(||\boldsymbol{p}_i|| = 1)$ になるようにとる。

主成分分析では，この基底ベクトルが主成分軸を表し，基底ベクトルの要素を主成分負荷量，係数 c_i を第 i 主成分得点とよぶ。因子分析では基底ベクトルの要素 p_{ij} を因子負荷量，係数 c_i を第 i 共通因子とよぶ。

$$\begin{pmatrix} e_1(t_1) & e_1(t_2) & \cdots \\ e_2(t_1) & e_2(t_2) & \cdots \\ \vdots & & \\ e_N(t_1) & e_N(t_2) & \cdots \end{pmatrix} = \begin{pmatrix} p_{11} & p_{21} & \cdots \\ p_{12} & p_{22} & \cdots \\ \vdots & & \\ p_{1N} & p_{2N} & \cdots \end{pmatrix} \begin{pmatrix} c_1(t_1) & c_1(t_2) & \cdots \\ c_2(t_1) & c_2(t_2) & \cdots \\ \vdots & & \\ c_N(t_1) & c_1(t_2) & \cdots \end{pmatrix} \tag{3}$$

2　主成分分析 (Principle Component Analysis, PCA)

変数の分布を考慮した座標軸を作る。

- 原点をデータ点の平均ベクトルにとる
- 元データを射影した点の分散が大きくなる順に座標軸 (主成分) を選択

- 直交基底
- 得られる座標軸は元データの次元と同じ。その中から累積寄与率に基づいて選択

3　因子分析 (Factor Analysis, FA)

変数間の相関関係を因子によって説明する。

- 原点をデータ点の平均ベクトルにとる。
- 基底の数は分析時に指定
- 基底の選び方はいろいろある
  - 主因子法: 因子寄与が大きなものから順
  - 最尤法: 尤度が最大になる順
  - 最小自乗: 残差を小さくする順
  - etc
- 直交基底になるとは限らない

4　非負値因子分解 (Nonnegative Matrix Factorization, NMF)

- 因子分析の一種
- 基底の要素および係数が正になるように座標軸を決定
- 残差を表現する Variance Accounted For (VAF) によって因子数を決定することが多い