

時系列信号の次元圧縮

西井淳

2022 年 2 月 18 日

1 目的

取得した N 次元信号の時系列 $\mathbf{e}(t) = (e_1(t), e_2(t), \dots, e_N(t))^t$ を $M (\leq N)$ 個の基底 $\{\mathbf{p}_1, \mathbf{p}_2, \dots, \mathbf{p}_M\}$, $\mathbf{p}_i = (p_{i1}, p_{i2}, \dots, p_{iN})^t$, $(i = 1, 2, \dots, M)$ で表す。このとき、新しい基底の次元数 M が元データの次元 N より小さければ ($N > M$)、もとの信号をよく表現できるならば、次元圧縮が可能になる。

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

$$\mathbf{e}(t) = c_1(t)\mathbf{p}_1 + c_2(t)\mathbf{p}_2 + \dots \quad (2)$$

$$= (\mathbf{p}_1, \mathbf{p}_2, \dots)\mathbf{c}(t) \quad (3)$$

ここで、 $\mathbf{c}(t) = (c_1(t), c_2(t), \dots)^t$ であり、 $c_i(t)$ は基底 $\{\mathbf{p}_i\}$ を座標軸とする空間で $\mathbf{e}(t)$ を表した場合の座標値を表す。各基底ベクトルは単位ベクトル ($\|\mathbf{p}_i\| = 1$) になるようにとる。

主成分分析では、この基底ベクトル \mathbf{p}_i が主成分軸を表す。また、基底ベクトルの要素 p_{ij} 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 p_{ij} を因子負荷量、 c_i を第 i 共通因子とよぶ。 $E = (\mathbf{e}(t_1), \mathbf{e}(t_2), \dots)$, $P = (\mathbf{p}_1, \mathbf{p}_2, \dots)$, $C = (\mathbf{c}(t_1), \mathbf{c}(t_2), \dots)$ とおくと、以下のように書くことができる。

$$\begin{pmatrix} e_1(t_1) & e_1(t_2) & \dots \\ e_2(t_1) & e_2(t_2) & \dots \\ \vdots & \vdots & \ddots \\ e_N(t_1) & e_N(t_2) & \dots \end{pmatrix} = \begin{pmatrix} p_{11} & p_{21} & \dots \\ p_{12} & p_{22} & \dots \\ \vdots & \vdots & \ddots \\ p_{1N} & p_{2N} & \dots \end{pmatrix} \begin{pmatrix} c_1(t_1) & c_1(t_2) & \dots \\ c_2(t_1) & c_2(t_2) & \dots \\ \vdots & \vdots & \ddots \\ c_N(t_1) & c_N(t_2) & \dots \end{pmatrix} \quad (4)$$

$$E = PC \quad (5)$$

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

主成分分析は変数の分布をよく表現する座標系を取り直すために用いられる手法。

- 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)

因子分析は、変数間の相関関係をその背後に隠れた因子によって説明するための手法。

- 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 ($N = M$ の場合の結果は主成分主成分の結果と一致)
 - 最尤法: 尤度が最大になる順 (元データが正規分布であると)
 - 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 非負値行列因子分解 (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。
おく