カーネル主成分分析

照屋 佑喜仁

May 1, 2025

1. 高次元空間への写像

高次元空間への写像

- 前回の最後で，カーネル法に基づく主成分分析の基本的な考え方が述べられた.
- 高次元空間に散らばるデータの非線形構造を捉える一つのアプローチがカーネル主成分分析であった.
- 中心化したデータの分散共分散行列 S の固有値問題を，データ間の内積で定義される新たな行列 K の固有値問題に置き換えて主成分を求めることができそう
- この項で，固有値問題の置き換えを，高次元空間へと写像したデータに対して適応する.

高次元空間への写像

入力空間上の n 個の p 次元観測データ $\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_n$ を**特徴空間**とよばれる高次元空間へ写像し，特徴空間上で主成分分析を実行する．

一般に，入力空間上の p 次元データ $\boldsymbol{x} = (x_1, x_2, \cdots, x_p)^T$ を，より高次の r 次元空間へと変換する非線形空間を次のようにベクトルで表記する．

$$\boldsymbol{\Phi}(\boldsymbol{x}) \equiv (\phi_1(\boldsymbol{x}), \phi_2(\boldsymbol{x}), \cdots, \phi_r(\boldsymbol{x}))^T$$

このベクトルの各成分 $\phi_j(\boldsymbol{x})$ は p 変数実数値関数とする．

これのベクトルによって，観測された n 個の p 次元データを r 次元特徴空間上へと写像したデータを

$$\boldsymbol{\Phi}(\boldsymbol{x}_1), \boldsymbol{\Phi}(\boldsymbol{x}_2), \cdots, \boldsymbol{\Phi}(\boldsymbol{x}_n)$$

とする．この特徴空間上の n 個の r 次元データに基づいて主成分分析を実行する．

高次元空間への写像

まず，特徴空間上に散らばる r 次元データの重心に原点を移す．（つまり中心化する．）

標本平均ベクトルは

$$\overline{\boldsymbol{\Phi}} = \frac{1}{n} \sum_{i=1}^{n} \boldsymbol{\Phi}(\boldsymbol{x}_i)$$

となり，次のようにデータを中心化を行う．

$$\boldsymbol{\Phi}_c(\boldsymbol{x}_i) = \boldsymbol{\Phi}(\boldsymbol{x}_i) - \overline{\boldsymbol{\Phi}}, \quad i = 1, 2, \cdots, n$$

高次元空間への写像

中心化したデータの標本平均ベクトルは

$$\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{\Phi}_c(\boldsymbol{x}_i) = \frac{1}{n}\sum_{i=1}^{n}\left\{\boldsymbol{\Phi}(\boldsymbol{x}_i) - \overline{\boldsymbol{\Phi}}\right\} = \mathbf{0}$$

（中心化したため当たり前）

次に，中心化したデータを行ベクトルとする $n \times r$ 行列を

$$Z_c = \left(\boldsymbol{\Phi}_c(\boldsymbol{x}_1), \boldsymbol{\Phi}_c(\boldsymbol{x}_2), \cdots, \boldsymbol{\Phi}_c(\boldsymbol{x}_n)\right)^T$$

とすると，標本分散共分散行列は

$$S_c = \frac{1}{n}\sum_{i=1}^{n}\boldsymbol{\Phi}_c(\boldsymbol{x}_i)\boldsymbol{\Phi}_c(\boldsymbol{x}_i)^T = \frac{1}{n}Z_c^T Z_c$$

高次元空間への写像

前回行ったように，分散共分散行列 S_c の固有値問題を，特徴空間上のデータ間の内積に基づく行列の固有値問題へと置き換える．

一般に，行列 S_c の第 α 番目の固有値 λ_α^F と対応する正規化済みの固有ベクトル $\boldsymbol{w}_\alpha^F$ の関係式

$$S_c \boldsymbol{w}_\alpha^F = \lambda_\alpha^F \boldsymbol{w}_\alpha^F \tag{1}$$

から，固有ベクトルは，中心化した特徴空間上のデータを基底として次のように表すことができるのであった．（前回）

高次元空間への写像

Review

$$S_c \boldsymbol{w}_\alpha^F = \lambda_\alpha^F \boldsymbol{w}_\alpha^F$$

の左辺は次のように書き表せる

$$\begin{aligned} S_c \boldsymbol{w}_\alpha^F &= \frac{1}{n} \sum_{i=1}^{n} \boldsymbol{\Phi}_c(\boldsymbol{x}_i) \boldsymbol{\Phi}_c(\boldsymbol{x}_i)^T \boldsymbol{w}_\alpha^F \\ &= \frac{1}{n} \sum_{i=1}^{n} \left\{ \boldsymbol{\Phi}_c(\boldsymbol{x}_i)^T \boldsymbol{w}_\alpha^F \right\} \boldsymbol{\Phi}_c(\boldsymbol{x}_i) \quad \because \boldsymbol{\Phi}_c(\boldsymbol{x}_i)^T \boldsymbol{w}_\alpha^F \text{ はスカラー} \end{aligned}$$

高次元空間への写像

改めて固有ベクトルは次で表せる.

$$\boldsymbol{w}_\alpha^F = Z_c^T \boldsymbol{c}_\alpha^F, \quad Z_c^T = (\boldsymbol{\Phi}_c(\boldsymbol{x}_1), \boldsymbol{\Phi}_c(\boldsymbol{x}_2), \cdots, \boldsymbol{\Phi}_c(\boldsymbol{x}_n),) \tag{3}$$

ただし, $\boldsymbol{c}_\alpha^T$ は n 次元係数ベクトルとする. これを (1) に代入すると

$$\frac{1}{n} Z_c^T Z_c Z_c^T \boldsymbol{c}_\alpha^F = \lambda_\alpha^F Z_c^T \boldsymbol{c}_\alpha^F$$

左から Z_c を掛けて分母を払うと

$$Z_c Z_c^T Z_c Z_c^T \boldsymbol{c}_\alpha^F = n\lambda_\alpha^F Z_c Z_c^T \boldsymbol{c}_\alpha^F \tag{2}$$

を得る. $Z_c Z_c^T$ は次のように特徴空間上の中心化したデータ間の内積を成分とする n 次元対称行列である.

高次元空間への写像

$$
\begin{aligned}
K_c &\equiv Z_c Z_c^T \\
&= \begin{pmatrix}
\boldsymbol{\Phi}_c(\boldsymbol{x}_1)^T \boldsymbol{\Phi}_c(\boldsymbol{x}_1) & \boldsymbol{\Phi}_c(\boldsymbol{x}_1)^T \boldsymbol{\Phi}_c(\boldsymbol{x}_2) & \cdots & \boldsymbol{\Phi}_c(\boldsymbol{x}_1)^T \boldsymbol{\Phi}_c(\boldsymbol{x}_n) \\
\boldsymbol{\Phi}_c(\boldsymbol{x}_2)^T \boldsymbol{\Phi}_c(\boldsymbol{x}_1) & \boldsymbol{\Phi}_c(\boldsymbol{x}_2)^T \boldsymbol{\Phi}_c(\boldsymbol{x}_2) & \cdots & \boldsymbol{\Phi}_c(\boldsymbol{x}_2)^T \boldsymbol{\Phi}_c(\boldsymbol{x}_n) \\
\vdots & \vdots & \ddots & \vdots \\
\boldsymbol{\Phi}_c(\boldsymbol{x}_n)^T \boldsymbol{\Phi}_c(\boldsymbol{x}_1) & \boldsymbol{\Phi}_c(\boldsymbol{x}_n)^T \boldsymbol{\Phi}_c(\boldsymbol{x}_2) & \dots & \boldsymbol{\Phi}_c(\boldsymbol{x}_n)^T \boldsymbol{\Phi}_c(\boldsymbol{x}_n)
\end{pmatrix}
\end{aligned}
$$

よって (2) は次のように表される.

$$K_c^2 \boldsymbol{w}_\alpha^F = n\lambda_\alpha^F \boldsymbol{w}_\alpha^F$$

高次元空間への写像

したがって，第 α 番目の大きさの固有値 $\lambda_\alpha^F (\neq 0)$ と対応する固有ベクトル $\boldsymbol{c}_\alpha^F$ を求める問題は，次の n 次対称行列 K_c の固有値問題へと帰着される．

$$K_c \boldsymbol{c}_\alpha^F = n \lambda_\alpha^F \boldsymbol{c}_\alpha^F$$

求めた固有ベクトル $\boldsymbol{c}_\alpha^F$ を (3) へ代入すると