PCAとICA

10月17日

吉田雄介

PCA(1/10)

- PCAとは
  - Principal Component Analysisの略で
    主成分分析と訳される

- 主成分とは
  - 分散行列の最大固有値に対応する固有ベクトル

PCA(2/10)

- R^n上のTこの観測値ベクトルx(1),...,x(T)が与えられたとき単位ベクトル$\|\gamma\|=1$の中で

$$\sum_{t=1}^{T}\left[\left\|x(t)-\bar{x}\right\|^2-\left\{\gamma^T\left(x(t)-\bar{x}\right)\right\}^2\right] \qquad \text{式(1)}$$

を最小にするγを求める
  - 標本平均をx、ベクトルのノルムを||･||で表す

- 標本平均を0とすると式(1)は

$$\sum_{t=1}^{T}\left\{\left\|x(t)\right\|^2-\left(\gamma^T x(t)\right)^2\right\} \qquad \text{式(2)}$$

PCA(3/10)

- 第2項を書き換えると

$$\sum_{t=1}^{T} \left\| \left(\gamma^{T} x(t) \right) \gamma \right\|^{2} \qquad \text{式(3)}$$

と書き換えられ、式(2)は

$$\sum_{t=1}^{T} \left\| x(t) - \left(\gamma^{T} x(t) \right) \gamma \right\|^{2} \qquad \text{式(4)}$$

と書き換えられる

PCA(4/10)

- 観測値$x(t)$をある定ベクトルγの方向に定数倍$s(t)=\gamma^T x(t)$して

  $$x(t) \sim s(t)\gamma \qquad \text{式(5)}$$

  と近似したときにもとの信号を一番良く近似するγを求めるとs(t)は$\gamma^T x(t)$となる
- これより式(4)を書き換えると

  $$\sum_{t=1}^{T} \left\| x(t) - s(t)\gamma \right\|^2 = \sum_{t=1}^{T} \left\{ \left\| x(t) \right\|^2 - \left\| s(t)\gamma \right\|^2 \right\} \qquad \text{式(6)}$$

  となり、これを最小にするγを考える

PCA(5/10)

- 式(5)を最小化するには

$$\sum_{t=1}^{T} \left\| s(t)\gamma \right\|^2 \qquad 式(7)$$

  を最大化する

- s(t)を代入して

$$\sum_{t=1}^{T} \left\| \left(\gamma^T x(t)\right)\gamma \right\|^2 = \sum_{t=1}^{T} \left\| \gamma \left(x(t)^T \gamma\right) \right\|^2$$

$$= \gamma^T \left(\sum_{t=1}^{T} x(t) x(t)^T \right) \gamma \qquad 式(8)$$

PCA(6/10)

- 分散行列を

$$V = \frac{1}{T}\sum_{t=1}^{T} x(t)x(t)^T \qquad 式(9)$$

として、式(8)は

$$\gamma^T V \gamma \qquad 式(10)$$

となり、これを最大にする大きさ1のベクトルを求める問題になる

PCA(7/10)

- 分散行列は正定値なので、
  ある直行行列*S*を使って

$$S^T V S = \begin{pmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{pmatrix} \qquad \text{式(11)}$$

  のように対角化できる。
  - ただし、この対角成分は*V*の固有値で

$$\lambda_1 \geq \cdots \geq \lambda_n \ (\geq 0) \qquad \text{式(12)}$$

    となるように*S*は選ぶものとする

PCA(8/10)

- この対角行列を両側から挟み込んで最大にする大きさ1のベクトルは(1,0,…,0)または(-1,0,…,0)であり、正負は問題ないので(1,0,…,0)だけを考えるとすると

$$\gamma = S\begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} \qquad \text{式(13)}$$

*V*を両側から挟んで最大にするベクトルが求められる

PCA(9/10)

- 式(5)の近似は拡張でき、

  $x(t) \sim s_1(t)\gamma_1 + s_2(t)\gamma_2 + \cdots$ 式(14)

  というように精度を上げることができる

- 主成分が求められた元の観測値から
  主成分方向の成分を取り除いた残差

  $x(t) - s_1(t)\gamma_1$ 式(15)

  に対して同じように主成分分析を繰り返し行うことで

  $x(t) = s_1(t)\gamma_1 + s_2(t)\gamma_2 + \cdots + s(t)_n(t)\gamma_n$ 式(16)

  という分解が得られる

PCA(10/10)

- $\frac{1}{T}\sum_{t=1}^{T} s_i(t)s_j(t)$ （式(17)）について
  - i≠jのときはγ_iとγ_jが直交しているので式(17)は0となり、係数は無相関である
  - i＝jのときは式(17)はλ_iとなる
- これより

$$\frac{1}{T}\sum_{t=1}^{T}\left\|x(t)\right\|^2 = \sum_{i=1}^{n}\lambda_i \qquad \text{式(18)}$$

とパワーの分解を与えていると解釈できる

ICA(1/4)

- $y(t) = Bx(t)$ 式(19)
  という変換で$y(t)$の各成分が独立なので
  Bを使って$x(t)$を分解すると
  $x(t) = B^{-1}y(t)$ 式(20)
  となり、B^{-1}の各列ベクトルをγ_i、
  $y(t)$の各成分を$s_i(t)$として考える

ICA(2/4)

- $x(t) = B^{-1}Bx(t)$

$$= B^{-1}\left\{\begin{pmatrix} 1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & 0 \end{pmatrix} + \cdots + \begin{pmatrix} 0 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & 1 \end{pmatrix}\right\}Bx(t)$$

$$= B^{-1}\left\{\begin{pmatrix} 1 \\ \vdots \\ 0 \end{pmatrix}\begin{pmatrix} 1 & \cdots & 0 \end{pmatrix} + \cdots + \begin{pmatrix} 0 \\ \vdots \\ 1 \end{pmatrix}\begin{pmatrix} 0 & \cdots & 1 \end{pmatrix}\right\}Bx(t)$$

$$= \gamma_1\left(\gamma_1^* x(t)\right) + \cdots + \gamma_n\left(\gamma_n^* x(t)\right)$$

$$= s_1(t)\gamma_1 + \cdots + s_n(t)\gamma_n \qquad \text{式(21)}$$

ここでγ_iは行列B^{-1}の第i番目の列ベクトル

γ_i^*は行列Bの第i番目の行ベクトル

ICA(3/4)

- $\|\gamma_i\|=1$としてもγ_iが互いに直交していないので

$$\frac{1}{T}\sum_{t=1}^{T}\|x(t)\|^2 \neq \frac{1}{T}\sum_{i=1}^{n}\sum_{t=1}^{T}s_i(t)^2 \qquad \text{式(22)}$$

- $s_i(t)$と$s_j(t)$はy_i, y_jの独立性から相関が0になるので、係数が無相関であるため

$$\frac{1}{T}\sum_{t=1}^{T}s_i(t)s_j(t) = 0 \qquad \text{式(23)}$$

ICA(4/4)

- $x(t)^*$は$x(t)$の双対座標であり、この場合はユークリッド空間なので$x(t)$の転置を考え、

$$s_i(t)^* = x(t)^* \gamma_i = x(t)^T \gamma_i$$ 式(24)

を定義し

$$x(t)^T = \gamma_1^* s_1(t)^* + \cdots + \gamma_n^* s_n(t)^*$$ 式(25)

という分解を考えると、基底の双対性から

$$\frac{1}{T}\sum_{t=1}^{T}\left\| x(t) \right\|^2 = \frac{1}{T}\sum_{i=1}^{n}\sum_{t=1}^{T} s_i(t) s_i(t)^*$$ 式(26)

という分解が成り立つ

PCAとICAの共通点と違い

- 共通点

  $$x(t) = s_1(t)\gamma_1 + s_2(t)\gamma_2 + \cdots + s(t)_n(t)\gamma_n$$

  という形で確率変数ベクトルの分解を考える

- 違い
  - 分解のために採用する基準
    $\gamma_1,\ldots,\gamma_n$が線形独立ならどんな分解も可能
  - ICAは基底が直交してなくてもよいが
    分解の係数が独立になるように基底を選ぶ
  - PCAはγとして互いに直交する方向への分解を考え、
    分解の係数が無相関になるように基底を選ぶ