2021.12.16改訂版

第10回　主成分分析

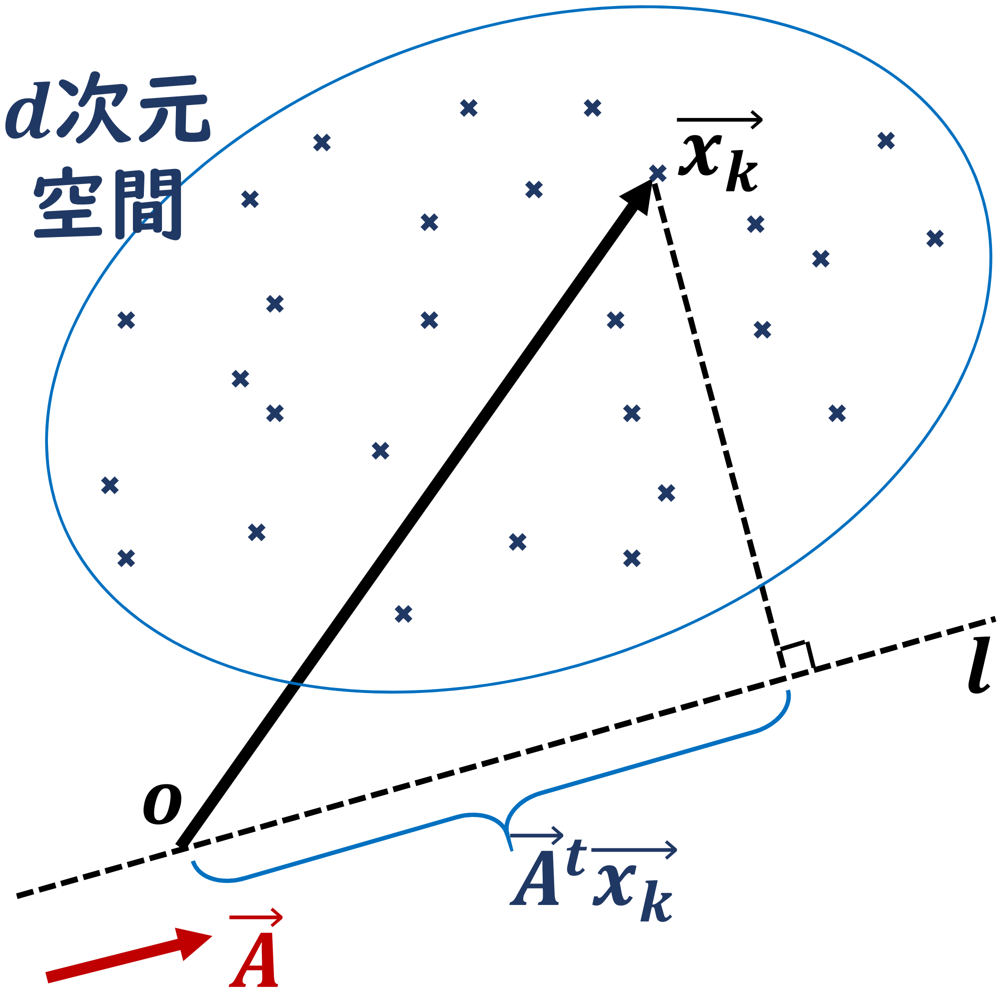
画像認識工学

情報工学科4年

科目担当　鈴木

主成分分析(PCA:Principal Componet Analysis)

d次元空間の多数のデータ$\overrightarrow{x_1}, \overrightarrow{x_2}, \cdots, \overrightarrow{x_n}$の分布において、分散最大の軸$\vec{A}$(主成分)を求める手法を主成分分析という。



$\vec{A}$は直線lの方向ベクトルなので単位ベクトルと仮定しても一般性は失われない。

直線lは原点を通る直線と仮定しても一般性は失われない。

$\vec{x}$を直線lに正射影するとき、原点からその点までの長さは、$\vec{A}^t\overrightarrow{x_k}$である。

正射影後の分散

$\overrightarrow{x_1}, \overrightarrow{x_2}, \cdots, \overrightarrow{x_n}$の平均（平均ベクトル）$\overrightarrow{m}$は次式で表される。

$$\overrightarrow{m} = \frac{1}{n}\sum_{k=1}^{n} \overrightarrow{x_k}$$

$\overrightarrow{x_k}$を直線lに正射影したときの値は$\overrightarrow{A}^t\overrightarrow{x_k}$だから、その平均値$m_A$は次のようになる。

$$m_A = \frac{1}{n}\sum_{k=1}^{n} \overrightarrow{A}^t\overrightarrow{x_k} = \overrightarrow{A}^t\left[\frac{1}{n}\sum_{k=1}^{n} \overrightarrow{x_k}\right] = \overrightarrow{A}^t\overrightarrow{m}$$

直線l上のデータの分散σ^2は次のようになる。

$$\sigma^2 = \frac{1}{n}\sum_{k=1}^{n} \left(\overrightarrow{A}^t\overrightarrow{x_k} - \overrightarrow{A}^t\overrightarrow{m}\right)^2$$

σ^2を最大にする
$\overrightarrow{A}$を求める！

分散の解析

$$\sigma^2 = \frac{1}{n}\sum_{k=1}^{n}\left(\vec{A}^t\vec{x_k} - \vec{A}^t\vec{m}\right)^2 = \frac{1}{n}\sum_{k=1}^{n}\left\{\vec{A}^t(\vec{x_k} - \vec{m})\right\}^2$$

$$= \frac{1}{n}\sum_{k=1}^{n}\left\{\underbrace{\vec{A}^t(\vec{x_k} - \vec{m})}_{\text{スカラー}} \cdot \underbrace{\vec{A}^t(\vec{x_k} - \vec{m})}_{\text{スカラー}}\right\}$$

$$\vec{x}^t\vec{y} = \vec{y}^t\vec{x}$$

$$= \frac{1}{n}\sum_{k=1}^{n}\left\{\vec{A}^t(\vec{x_k} - \vec{m}) \cdot (\vec{x_k} - \vec{m})^t\vec{A}\right\}$$

$$= \vec{A}^t\left[\frac{1}{n}\sum_{k=1}^{n}\{(\vec{x_k} - \vec{m})(\vec{x_k} - \vec{m})^t\}\right]\vec{A} = \vec{A}^t\Sigma\vec{A}$$

Work①

d次元空間の多数のデータ$\overrightarrow{x_1}, \overrightarrow{x_2}, \cdots, \overrightarrow{x_n}$を考える。
$\overrightarrow{x_k} = (x_{k1}, x_{k2}, \cdots, x_{kd})^t$とするとき、下記の行列について次の値を求めよ。ただし、$\overrightarrow{m} = (m_1, m_2, \cdots, m_d)^t$は$\overrightarrow{x_1}, \overrightarrow{x_2}, \cdots, \overrightarrow{x_n}$の平均ベクトルとする。

$$\Sigma = \frac{1}{n}\sum_{k=1}^{n}\{(\overrightarrow{x_k} - \overrightarrow{m})(\overrightarrow{x_k} - \overrightarrow{m})^t\}$$

① Σの1行1列成分を求めよ。

② Σの1行2列成分を求めよ。

③ Σのi行j列成分を求めよ。

共分散行列 (Covariance Matrix)

d次元空間の多数のデータ$\overrightarrow{x_1}, \overrightarrow{x_2}, \cdots, \overrightarrow{x_n}$に対し、行列の$i, j$成分が、データの第$i$番目の要素と第$j$番目の要素の共分散になっているような行列$\Sigma$を**分散・共分散行列（Variance-Covariance Matrix）**または**共分散行列（Covariance Matrix）**という。

共分散行列Σはd次の実対称行列であり次式で計算できる。

$$\boldsymbol{\Sigma} = \frac{1}{n}\sum_{k=1}^{n}\{(\overrightarrow{x_k} - \overrightarrow{m})(\overrightarrow{x_k} - \overrightarrow{m})^t\}$$

分散最大性

$$\sigma^2 = \vec{A}^t \Sigma \vec{A} \quad \Sigma = \frac{1}{n}\sum_{k=1}^{n}\{(\vec{x_k} - \vec{m})(\vec{x_k} - \vec{m})^t\}$$

最大化

- Σはデータから求まる定数行列。
- $\|\vec{A}\|$を大きくすればいくらでも大きくなる。
  ($\vec{A}$を単位ベクトルと仮定していた。。。)

$\|\vec{A}\| = 1$という制約のもとで$\sigma^2 = \vec{A}^t \Sigma \vec{A}$を最大にする問題を解けばよい。このような問題を**変分問題**という。

ラグランジュの未定係数法

$g(\vec{x}) = 0$という制約のもとで$f(\vec{x})$が最大（最小）となる$\vec{x}$を求めるためには、λを定数として$F(\vec{x}) = f(\vec{x}) - \lambda g(\vec{x})$とおき、

$$\frac{\partial F(\vec{x})}{\partial x_k} = \frac{\partial F(\vec{x})}{\partial \lambda} = 0$$

$$k = 1, 2, \cdots, d \quad \vec{x}^t = (x_1, x_2, \cdots, x_d)$$

を解けばよい。この手法をラグランジュ（Lagrange）の未定係数（乗数）法という。

Work②

① 二次曲線$x^2+y^2-2xy-\sqrt{2}x-\sqrt{2}y+2=0$上の点のうち、原点に一番近い点を求めよ。

② 三次元曲面 $z=2x^2+3y^2-12y+13$ において，点$(0,3,4)$における接平面の方程式を求めよ.

分散最大性

$\left\|\vec{A}\right\| = 1$、つまり$g(\vec{a}) = \vec{A}^t\vec{A} - 1 = 0$という制約のもとで

$f\left(\vec{A}\right) = \sigma^2$が最大になる$\vec{A} = (A_1, A_2, \cdots, A_d)^t$を求めれば良い。

$$\sigma^2 = \vec{A}^t\Sigma\vec{A} \qquad \Sigma = \frac{1}{n}\sum_{k=1}^{n}\{(\overrightarrow{x_k} - \vec{m})(\overrightarrow{x_k} - \vec{m})^t\}$$

$$F\left(\vec{A}\right) = \vec{A}^t\left[\frac{1}{n}\sum_{k=1}^{n}\{(\overrightarrow{x_k} - \vec{m})(\overrightarrow{x_k} - \vec{m})^t\}\right]\vec{A} - \lambda\left(\vec{A}^t\vec{A} - 1\right)$$

$$\frac{\partial F}{\partial \vec{A}} = \left(\frac{\partial F}{\partial A_1}, \frac{\partial F}{\partial A_2}, \cdots, \frac{\partial F}{\partial A_d}\right)^t = \vec{0} \qquad \frac{\partial F}{\partial \lambda} = 0$$