目次

目次

確率的主成分分析 (PPCA)

- 主成分分析の確率モデルによる表現
  - 主成分分析は、潜在変数を含む確率モデルの、最尤推定として導出できることを示す
  - このような主成分分析の再定式化を確率的主成分分析という

  - そして、主成分分析を行うための EM アルゴリズムを導出してみる

確率的主成分分析 (PPCA)

- 主成分分析の確率モデルによる表現
  - 主成分分析では、D 個の変数 $\{x_1, x_2, \ldots, x_D\}$ を線形に組み合わせて、真にアクティブな M 個の変数を見いだす
  - 主部分空間に存在する潜在変数を、M 次元のベクトル $\boldsymbol{z}$ として明示的に表現する
  - D 次元のデータ $\boldsymbol{x}$ は、主部分空間内の M 次元のベクトル $\boldsymbol{z}$ から、以下のように、ガウス分布から確率的に生成されたと考える

$$p(\boldsymbol{x}|\boldsymbol{z}) = \mathcal{N}(\boldsymbol{x}|\boldsymbol{W}\boldsymbol{z} + \boldsymbol{\mu}, \sigma^2 \boldsymbol{I}) \tag{1}$$

  - $D \times M$ 行列 $\boldsymbol{W}$ の列ベクトルが張る空間 (線形部分空間) は、データ空間における主部分空間に対応する
  - D 次ベクトル $\boldsymbol{\mu}$ は、観測変数 $\boldsymbol{x}$ の平均である

確率的主成分分析 (PPCA)

- 潜在変数 $\boldsymbol{z}$ に関する事前分布 $p(\boldsymbol{z})$ も、平均 0 で共分散行列が単位行列 $\boldsymbol{I}$ のガウス分布とする
$$p(\boldsymbol{z}) = \mathcal{N}(\boldsymbol{z}|\boldsymbol{0}, \boldsymbol{I}) \tag{2}$$
- 事前分布 $p(\boldsymbol{z})$ の平均が 0、共分散行列が $\boldsymbol{I}$ でも、一般性は失われていない
- 事前分布 $p(\boldsymbol{z})$ として、より一般的なガウス分布を仮定しても、得られる確率モデルは結果的に等価になる

- 生成モデルの観点から、確率的主成分分析モデルを考えることができる

確率的主成分分析 (PPCA)

- 生成モデルの観点
  - 生成モデルとは、実際に観測されるデータ $\boldsymbol{x}$ が、どのように生成されるかという過程を、確率分布などを使ってモデル化したものである
  - ここでは、潜在変数 $\boldsymbol{z}$ を含んだ生成モデルを考えている

- データの生成過程
  1. 潜在変数 $\boldsymbol{z}_i$ を、事前分布 $p(\boldsymbol{z})$ からサンプリングする
  2. $\boldsymbol{z}_i$ を使って、条件付き分布 $p(\boldsymbol{x}|\boldsymbol{z})$ からデータ $\boldsymbol{x}_i$ をサンプリングする

- データの生成過程に関する注意点
  - D 次元の観測変数 $\boldsymbol{x}_i$ は、M 次元の潜在変数 $\boldsymbol{z}_i$ に線形変換を施したあと ($\boldsymbol{x}_i = \boldsymbol{W}\boldsymbol{z}_i + \boldsymbol{\mu}$)、ガウス分布によるノイズ ($\boldsymbol{\epsilon}_i \sim \mathcal{N}(\boldsymbol{\epsilon}|0, \sigma^2 \boldsymbol{I})$) が加えられたものとして定義される ($\boldsymbol{x}_i = \boldsymbol{W}\boldsymbol{z}_i + \boldsymbol{\mu}_i + \boldsymbol{\epsilon}_i$)

  - 観測変数 $\boldsymbol{x}_i$ の裏側には、潜在変数 $\boldsymbol{z}_i$ が潜んでいる
  - 潜在変数 $\boldsymbol{z}_i$ は実際に観測できないが、$\boldsymbol{x}_i$ の本質的な情報を表している

確率的主成分分析 (PPCA)

- パラメータ $\boldsymbol{W}, \boldsymbol{\mu}, \sigma^2$ の最尤推定
  - 観測変数 $\boldsymbol{x}_i$ の裏側には、潜在変数 $\boldsymbol{z}_i$ が対応している
  - パラメータ $\boldsymbol{W}, \boldsymbol{\mu}, \sigma^2$ を通じて、$\boldsymbol{z}_i$ から観測可能な $\boldsymbol{x}_i$ が生成される
  - パラメータの最尤推定を行うために、尤度関数 $p(\boldsymbol{x}|\boldsymbol{W}, \boldsymbol{\mu}, \sigma^2)$ を求める
  - $p(\boldsymbol{x}|\boldsymbol{W}, \boldsymbol{\mu}, \sigma^2)$ は、$\boldsymbol{x}$ と $\boldsymbol{z}$ に関する同時分布 $p(\boldsymbol{x}, \boldsymbol{z}|\boldsymbol{W}, \boldsymbol{\mu}, \sigma^2)$ の、$\boldsymbol{z}$ に関する周辺化によって得られる

$$p(\boldsymbol{x}|\boldsymbol{W}, \boldsymbol{\mu}, \sigma^2) = \int p(\boldsymbol{x}|\boldsymbol{z}, \boldsymbol{W}, \boldsymbol{\mu}, \sigma^2)p(\boldsymbol{z})d\boldsymbol{z} \tag{3}$$

  - $p(\boldsymbol{x}|\boldsymbol{z}, \boldsymbol{W}, \boldsymbol{\mu}, \sigma^2)$ と $p(\boldsymbol{z})$ は、いずれもガウス分布であったので、ガウス分布に関する公式から $p(\boldsymbol{x}|\boldsymbol{W}, \boldsymbol{\mu}, \sigma^2)$ は次のようになる

$$p(\boldsymbol{x}|\boldsymbol{W}, \boldsymbol{\mu}, \sigma^2) = \mathcal{N}(\boldsymbol{x}|\boldsymbol{\mu}, \boldsymbol{C}) \tag{4}$$

確率的主成分分析 (PPCA)

- 行列 $\boldsymbol{C}$ は次のように定義される

$$\boldsymbol{C} = \boldsymbol{W}\boldsymbol{W}^T + \sigma^2\boldsymbol{I} \tag{5}$$

- 更に、潜在変数 $\boldsymbol{z}$ に関する事後分布 $p(\boldsymbol{z}|\boldsymbol{x})$ も計算できる

$$p(\boldsymbol{z}|\boldsymbol{x}, \boldsymbol{W}, \boldsymbol{\mu}, \sigma^2) = \mathcal{N}(\boldsymbol{z}|\boldsymbol{M}^{-1}\boldsymbol{W}^T(\boldsymbol{x} - \boldsymbol{\mu}), \sigma^2\boldsymbol{M}^{-1}) \tag{6}$$

- 行列 $\boldsymbol{M}$ は次のように定義される

$$\boldsymbol{M} = \boldsymbol{W}^T\boldsymbol{W} + \sigma^2\boldsymbol{I} \tag{7}$$

- これらの式は、ガウス分布に関する次の公式を用いれば導出できる

$$p(\boldsymbol{y}|\boldsymbol{x}) = \mathcal{N}(\boldsymbol{y}|\boldsymbol{A}\boldsymbol{x} + \boldsymbol{b}, \boldsymbol{D}) \tag{8}$$

$$p(\boldsymbol{x}) = \mathcal{N}(\boldsymbol{x}|\boldsymbol{\mu}, \boldsymbol{\Sigma}) \tag{9}$$

確率的主成分分析 (PPCA)

- $p(\boldsymbol{y}|\boldsymbol{x})$ と $p(\boldsymbol{x})$ が上式のようなガウス分布であるとき、$p(\boldsymbol{x}|\boldsymbol{y})$ と $p(\boldsymbol{y})$ は次のようになる

$$
\begin{aligned}
p(\boldsymbol{x}|\boldsymbol{y}) &= \mathcal{N}\left(\boldsymbol{x}|\boldsymbol{M}\left(\boldsymbol{A}^T\boldsymbol{D}^{-1}(\boldsymbol{y}-\boldsymbol{b})+\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}\right), \boldsymbol{M}\right) \quad (10) \\
p(\boldsymbol{y}) &= \mathcal{N}(\boldsymbol{y}|\boldsymbol{A}\boldsymbol{\mu}+\boldsymbol{b}, \boldsymbol{D}+\boldsymbol{A}\boldsymbol{\Sigma}\boldsymbol{A}^T) \quad (11) \\
\boldsymbol{M} &= \left(\boldsymbol{A}^T\boldsymbol{D}^{-1}\boldsymbol{A}+\boldsymbol{\Sigma}^{-1}\right)^{-1} \quad (12)
\end{aligned}
$$

確率的主成分分析 (PPCA)

- パラメータ $\boldsymbol{\mu}$ に関する最尤推定
    - N 個の観測データ $\{\boldsymbol{x}_1, \dots, \boldsymbol{x}_N\}$ と、それに対応する潜在変数 $\{\boldsymbol{z}_1, \dots, \boldsymbol{z}_N\}$ を考える
    - ここでの目標は、データ $\mathcal{D} = \{\boldsymbol{x}_1, \dots, \boldsymbol{x}_N\}$ から、パラメータ $\boldsymbol{W}, \boldsymbol{\mu}, \sigma^2$ を最尤推定することである

    - N 個のデータをまとめた行列を $\boldsymbol{X}$ とする (第 i 行ベクトルは $\boldsymbol{x}_i^T$)
    - N 個の潜在変数をまとめた行列を $\boldsymbol{Z}$ とする (第 i 行ベクトルは $\boldsymbol{z}_i^T$)

    - 各データと潜在変数 $\boldsymbol{x}_i, \boldsymbol{z}_i$ は、分布 $p(\boldsymbol{x}|\boldsymbol{z}), p(\boldsymbol{z})$ から独立にサンプリングされるとする (データは i.i.d 標本であるとする)

確率的主成分分析 (PPCA)

- 対数尤度関数 $\ln p(\boldsymbol{X}|\boldsymbol{W}, \boldsymbol{\mu}, \sigma^2)$ は次のように書ける

$$
\begin{aligned}
& \ln p(\boldsymbol{X}|\boldsymbol{W}, \boldsymbol{\mu}, \sigma^2) \\
= & \ln \prod_i p(\boldsymbol{x}_i|\boldsymbol{W}, \boldsymbol{\mu}, \sigma^2) \\
= & \sum_i \ln p(\boldsymbol{x}_i|\boldsymbol{W}, \boldsymbol{\mu}, \sigma^2) \\
= & \sum_i \ln \mathcal{N}(\boldsymbol{x}_i|\boldsymbol{\mu}, \boldsymbol{C}) \\
= & \sum_i \ln \frac{1}{(2\pi)^{\frac{D}{2}}} \frac{1}{|\boldsymbol{C}|^{\frac{1}{2}}} \exp\left(-\frac{1}{2} (\boldsymbol{x}_i - \boldsymbol{\mu})^T \boldsymbol{C}^{-1} (\boldsymbol{x}_i - \boldsymbol{\mu})\right) \\
= & \sum_i \left(-\frac{D}{2} \ln 2\pi - \frac{1}{2} \ln |\boldsymbol{C}| - \frac{1}{2} (\boldsymbol{x}_i - \boldsymbol{\mu})^T \boldsymbol{C}^{-1} (\boldsymbol{x}_i - \boldsymbol{\mu})\right) \\
= & -\frac{ND}{2} \ln 2\pi - \frac{N}{2} \ln |\boldsymbol{C}| - \frac{1}{2} \sum_i (\boldsymbol{x}_i - \boldsymbol{\mu})^T \boldsymbol{C}^{-1} (\boldsymbol{x}_i - \boldsymbol{\mu})
\end{aligned}
$$

確率的主成分分析 (PPCA)

- 対数尤度を $\boldsymbol{\mu}$ で偏微分し、0 と等置することによって、$\boldsymbol{\mu}$ の最尤解が得られる

$$
\begin{aligned}
& \frac{\partial}{\partial \boldsymbol{\mu}} \ln p(\boldsymbol{X}|\boldsymbol{W}, \boldsymbol{\mu}, \sigma^2) \\
= \quad & -\frac{1}{2} \frac{\partial}{\partial \boldsymbol{\mu}} \sum_i (\boldsymbol{x}_i - \boldsymbol{\mu})^T \boldsymbol{C}^{-1} (\boldsymbol{x}_i - \boldsymbol{\mu}) \\
= \quad & -\frac{1}{2} \sum_i \left(-2\boldsymbol{C}^{-1}\right) (\boldsymbol{x}_i - \boldsymbol{\mu}) \\
= \quad & \boldsymbol{C}^{-1} \sum_i (\boldsymbol{x}_i - \boldsymbol{\mu}) \\
= \quad & 0
\end{aligned}
$$

これより $\boldsymbol{\mu}$ の最尤解は以下のようになる

$$
\boldsymbol{\mu} = \frac{1}{N} \sum_i \boldsymbol{x}_i \tag{13}
$$

確率的主成分分析 (PPCA)

- パラメータ $\boldsymbol{W}, \sigma^2$ に関する最尤推定
  - 対数尤度を $\boldsymbol{W}, \sigma^2$ について微分することは形式上は可能である
  - 但し、$\boldsymbol{W}, \sigma^2$ が行列 $\boldsymbol{C}$ の中に入れ子になっている
  - 従って、$\boldsymbol{W}, \sigma^2$ の厳密な閉形式の解を得ることは、複雑かつ困難である

  - そこで、パラメータ $\boldsymbol{W}, \sigma^2$ を求めるための EM アルゴリズムを導出したい

  - EM アルゴリズムは、潜在変数を含む確率モデルについて、パラメータの最尤推定を行うための一般的な枠組みを提供する

- EM アルゴリズムによる主成分分析の導出
  - 完全データ対数尤度関数 $\ln p(\boldsymbol{X}, \boldsymbol{Z} | \boldsymbol{W}, \boldsymbol{\mu}, \sigma^2)$ は次のようになる

$$
\begin{aligned}
& \ln p(\boldsymbol{X}, \boldsymbol{Z} | \boldsymbol{W}, \boldsymbol{\mu}, \sigma^2) \\
= \; & \ln \prod_i p(\boldsymbol{x}_i, \boldsymbol{z}_i | \boldsymbol{W}, \boldsymbol{\mu}, \sigma^2) \\
= \; & \sum_i \ln p(\boldsymbol{x}_i, \boldsymbol{z}_i | \boldsymbol{W}, \boldsymbol{\mu}, \sigma^2) \\
= \; & \sum_i \ln p(\boldsymbol{x}_i | \boldsymbol{z}_i, \boldsymbol{W}, \boldsymbol{\mu}, \sigma^2) p(\boldsymbol{z}_i) \\
= \; & \sum_i \ln \mathcal{N}(\boldsymbol{x}_i | \boldsymbol{W} \boldsymbol{z}_i + \boldsymbol{\mu}, \sigma^2 \boldsymbol{I}) \mathcal{N}(\boldsymbol{z}_i | \boldsymbol{0}, \boldsymbol{I}) \\
= \; & \sum_i \left(\ln \mathcal{N}(\boldsymbol{x}_i | \boldsymbol{W} \boldsymbol{z}_i + \boldsymbol{\mu}, \sigma^2 \boldsymbol{I}) + \ln \mathcal{N}(\boldsymbol{z}_i | \boldsymbol{0}, \boldsymbol{I}) \right)
\end{aligned}
\tag{14}
$$

確率的主成分分析 (PPCA)

第 1 項 $\ln \mathcal{N}(\boldsymbol{x}_i | \boldsymbol{W}\boldsymbol{z}_i + \boldsymbol{\mu}, \sigma^2 \boldsymbol{I})$ は次のようになる

$$
\begin{aligned}
& \ln \mathcal{N}(\boldsymbol{x}_i | \boldsymbol{W}\boldsymbol{z}_i + \boldsymbol{\mu}, \sigma^2 \boldsymbol{I}) \\
= \; & \ln \Bigg(\frac{1}{(2\pi)^{\frac{D}{2}}} \frac{1}{|\sigma^2 \boldsymbol{I}|^{\frac{1}{2}}} \\
& \quad \exp\left(-\frac{1}{2} (\boldsymbol{x}_i - \boldsymbol{W}\boldsymbol{z}_i - \boldsymbol{\mu})^T \left(\sigma^2 \boldsymbol{I}\right)^{-1} (\boldsymbol{x}_i - \boldsymbol{W}\boldsymbol{z}_i - \boldsymbol{\mu}) \right) \Bigg) \\
= \; & -\frac{D}{2} \ln 2\pi - \frac{1}{2} \ln |\sigma^2 \boldsymbol{I}| - \\
& \quad \frac{1}{2} (\boldsymbol{x}_i - \boldsymbol{W}\boldsymbol{z}_i - \boldsymbol{\mu})^T \left(\sigma^2 \boldsymbol{I}\right)^{-1} (\boldsymbol{x}_i - \boldsymbol{W}\boldsymbol{z}_i - \boldsymbol{\mu}) \\
= \; & -\frac{D}{2} \ln 2\pi - \frac{D}{2} \ln \sigma^2 - \\
& \quad \frac{1}{2\sigma^2} (\boldsymbol{x}_i - \boldsymbol{W}\boldsymbol{z}_i - \boldsymbol{\mu})^T (\boldsymbol{x}_i - \boldsymbol{W}\boldsymbol{z}_i - \boldsymbol{\mu}) \\
= \; & -\frac{D}{2} \ln(2\pi\sigma^2) - \frac{1}{2\sigma^2} (\boldsymbol{x}_i - \boldsymbol{\mu})^T (\boldsymbol{x}_i - \boldsymbol{\mu}) +
\end{aligned}
$$

$$
\begin{aligned}
& \frac{1}{\sigma^2}(\boldsymbol{W}\boldsymbol{z}_i)^T(\boldsymbol{x}_i-\boldsymbol{\mu}) - \frac{1}{2\sigma^2}(\boldsymbol{W}\boldsymbol{z}_i)^T(\boldsymbol{W}\boldsymbol{z}_i) \\
= \; & -\frac{D}{2}\ln(2\pi\sigma^2) - \frac{1}{2\sigma^2}(\boldsymbol{x}_i-\boldsymbol{\mu})^T(\boldsymbol{x}_i-\boldsymbol{\mu}) + \\
& \frac{1}{\sigma^2}\boldsymbol{z}_i^T\boldsymbol{W}^T(\boldsymbol{x}_i-\boldsymbol{\mu}) - \frac{1}{2\sigma^2}\boldsymbol{z}_i^T\boldsymbol{W}^T\boldsymbol{W}\boldsymbol{z}_i \\
= \; & -\frac{D}{2}\ln(2\pi\sigma^2) - \frac{1}{2\sigma^2}||\boldsymbol{x}_i-\boldsymbol{\mu}||^2 + \\
& \frac{1}{\sigma^2}\boldsymbol{z}_i^T\boldsymbol{W}^T(\boldsymbol{x}_i-\boldsymbol{\mu}) - \frac{1}{2\sigma^2}\mathrm{Tr}\left(\boldsymbol{z}_i^T\boldsymbol{W}^T\boldsymbol{W}\boldsymbol{z}_i\right) \\
= \; & -\frac{D}{2}\ln(2\pi\sigma^2) - \frac{1}{2\sigma^2}||\boldsymbol{x}_i-\boldsymbol{\mu}||^2 + \\
& \frac{1}{\sigma^2}\boldsymbol{z}_i^T\boldsymbol{W}^T(\boldsymbol{x}_i-\boldsymbol{\mu}) - \frac{1}{2\sigma^2}\mathrm{Tr}\left(\boldsymbol{z}_i\boldsymbol{z}_i^T\boldsymbol{W}^T\boldsymbol{W}\right)
\end{aligned} \tag{15}
$$

第 2 項 $\ln\mathcal{N}(\boldsymbol{z}_i|\boldsymbol{0},\boldsymbol{I})$ は次のようになる

$$
\ln\mathcal{N}(\boldsymbol{z}_i|\boldsymbol{0},\boldsymbol{I})
$$

確率的主成分分析 (PPCA)

$$
\begin{aligned}
&= \ln \left(\frac{1}{(2\pi)^{\frac{M}{2}}} \exp \left(-\frac{1}{2} \boldsymbol{z}_i^T \boldsymbol{z}_i \right) \right) \\
&= -\frac{M}{2} \ln 2\pi - \frac{1}{2} \boldsymbol{z}_i^T \boldsymbol{z}_i \\
&= -\frac{M}{2} \ln 2\pi - \frac{1}{2} \mathrm{Tr} \left(\boldsymbol{z}_i^T \boldsymbol{z}_i \right) \\
&= -\frac{M}{2} \ln 2\pi - \frac{1}{2} \mathrm{Tr} \left(\boldsymbol{z}_i \boldsymbol{z}_i^T \right) \qquad (16)
\end{aligned}
$$

これより $\ln p(\boldsymbol{X}, \boldsymbol{Z} | \boldsymbol{W}, \boldsymbol{\mu}, \sigma^2)$ は次のようになる

$$
\begin{aligned}
& \ln p(\boldsymbol{X}, \boldsymbol{Z} | \boldsymbol{W}, \boldsymbol{\mu}, \sigma^2) \\
= & \sum_i \left(\ln \mathcal{N}(\boldsymbol{x}_i | \boldsymbol{W} \boldsymbol{z}_i + \boldsymbol{\mu}, \sigma^2 \boldsymbol{I}) + \ln \mathcal{N}(\boldsymbol{z}_i | \boldsymbol{0}, \boldsymbol{I}) \right) \\
= & \sum_i \left\{ -\frac{D}{2} \ln(2\pi\sigma^2) - \frac{1}{2\sigma^2} ||\boldsymbol{x}_i - \boldsymbol{\mu}||^2 + \right.
\end{aligned}
$$

$$\frac{1}{\sigma^2}\boldsymbol{z}_i^T\boldsymbol{W}^T(\boldsymbol{x}_i-\boldsymbol{\mu})-\frac{1}{2\sigma^2}\mathrm{Tr}\left(\boldsymbol{z}_i\boldsymbol{z}_i^T\boldsymbol{W}^T\boldsymbol{W}\right)-\frac{M}{2}\ln 2\pi-\frac{1}{2}\mathrm{Tr}\left(\boldsymbol{z}_i\boldsymbol{z}_i^T\right)\Bigg\} \tag{17}$$

- $\boldsymbol{\mu}$ は、全データの平均として得られることが分かっている
- 従って、$\boldsymbol{\mu}$ は既知であるとして、$\bar{\boldsymbol{x}}$ と書くことにする

- 潜在変数の事後分布 $p(\boldsymbol{Z}|\boldsymbol{X},\boldsymbol{W},\boldsymbol{\mu},\sigma^2)$ に関する期待値を取ると、次のようになる ($\mathbb{E}[\cdot]$ は、事後分布による期待値を表す)

$$\begin{aligned}&\mathbb{E}\left[\ln p(\boldsymbol{X},\boldsymbol{Z}|\boldsymbol{W},\boldsymbol{\mu},\sigma^2)\right]\\=\;&\sum_i\Bigg\{-\frac{D}{2}\ln(2\pi\sigma^2)-\frac{1}{2\sigma^2}||\boldsymbol{x}_i-\boldsymbol{\mu}||^2+\\&\frac{1}{\sigma^2}\mathbb{E}[\boldsymbol{z}_i]^T\boldsymbol{W}^T(\boldsymbol{x}_i-\boldsymbol{\mu})-\frac{1}{2\sigma^2}\mathrm{Tr}\left(\mathbb{E}\left[\boldsymbol{z}_i\boldsymbol{z}_i^T\right]\boldsymbol{W}^T\boldsymbol{W}\right)-\\&\frac{M}{2}\ln 2\pi-\frac{1}{2}\mathrm{Tr}\left(\mathbb{E}\left[\boldsymbol{z}_i\boldsymbol{z}_i^T\right]\right)\Bigg\}\end{aligned} \tag{18}$$

確率的主成分分析 (PPCA)

- 事後分布は以下のようであった ($\boldsymbol{M} = \boldsymbol{W}^T\boldsymbol{W} + \sigma^2\boldsymbol{I}$)

$$p(\boldsymbol{z}|\boldsymbol{x}, \boldsymbol{W}, \boldsymbol{\mu}, \sigma^2) = \mathcal{N}(\boldsymbol{z}|\boldsymbol{M}^{-1}\boldsymbol{W}^T(\boldsymbol{x} - \boldsymbol{\mu}), \sigma^2\boldsymbol{M}^{-1}) \qquad (19)$$

- これより $\mathbb{E}[\boldsymbol{z}_i]$ と $\mathbb{E}\left[\boldsymbol{z}_i\boldsymbol{z}_i^T\right]$ は次のようになる

$$\mathbb{E}[\boldsymbol{z}_i] = \boldsymbol{M}^{-1}\boldsymbol{W}^T(\boldsymbol{x}_i - \boldsymbol{\mu}) \qquad (20)$$

$$\mathbb{E}\left[\boldsymbol{z}_i\boldsymbol{z}_i^T\right] = \sigma^2\boldsymbol{M}^{-1} + \mathbb{E}[\boldsymbol{z}_i]\,\mathbb{E}[\boldsymbol{z}_i]^T \qquad (21)$$

- $\mathbb{E}\left[\boldsymbol{z}_i\boldsymbol{z}_i^T\right]$ は、$\mathbb{E}\left[\boldsymbol{z}_i\boldsymbol{z}_i^T\right] = \mathrm{cov}[\boldsymbol{z}_i] + \mathbb{E}[\boldsymbol{z}_i]\,\mathbb{E}[\boldsymbol{z}_i]^T$ から得られる
- $\mathrm{cov}[\boldsymbol{z}]$ は、確率変数 $\boldsymbol{z}$ の共分散行列を意味する

- E ステップで計算するのは、$\mathbb{E}[\boldsymbol{z}_i]$ と $\mathbb{E}\left[\boldsymbol{z}_i\boldsymbol{z}_i^T\right]$ の 2 つである
- これらの値は、古いパラメータ $\boldsymbol{M}, \boldsymbol{W}, \boldsymbol{\mu}, \sigma^2$ を使って計算される

- M ステップでは、上記の期待値 $\mathbb{E}\left[\ln p(\boldsymbol{X}, \boldsymbol{Z}|\boldsymbol{W}, \boldsymbol{\mu}, \sigma^2)\right]$ を、各パラメータ $\boldsymbol{W}, \sigma^2$ について最大化する

確率的主成分分析 (PPCA)

- 期待値を最大化するような、$\boldsymbol{W}, \sigma^2$ の計算式は容易に導出できる
- $\boldsymbol{W}$ について偏微分し、0 と等置すると次のようになる

$$
\begin{aligned}
&\frac{\partial}{\partial \boldsymbol{W}} \mathbb{E}\left[\ln p(\boldsymbol{X}, \boldsymbol{Z} | \boldsymbol{W}, \boldsymbol{\mu}, \sigma^2)\right] \\
= \quad & \frac{\partial}{\partial \boldsymbol{W}} \sum_i \left\{ \frac{1}{\sigma^2} \mathbb{E}\left[\boldsymbol{z}_i\right]^T \boldsymbol{W}^T \left(\boldsymbol{x}_i - \boldsymbol{\mu}\right) - \frac{1}{2\sigma^2} \operatorname{Tr}\left(\mathbb{E}\left[\boldsymbol{z}_i \boldsymbol{z}_i^T\right] \boldsymbol{W}^T \boldsymbol{W}\right) \right\} \\
= \quad & \frac{1}{2\sigma^2} \sum_i \frac{\partial}{\partial \boldsymbol{W}} \left\{ 2\mathbb{E}\left[\boldsymbol{z}_i\right]^T \boldsymbol{W}^T \left(\boldsymbol{x}_i - \boldsymbol{\mu}\right) - \operatorname{Tr}\left(\mathbb{E}\left[\boldsymbol{z}_i \boldsymbol{z}_i^T\right] \boldsymbol{W}^T \boldsymbol{W}\right) \right\} \\
= \quad & -\frac{1}{2\sigma^2} \sum_i \frac{\partial}{\partial \boldsymbol{W}} \left\{ 2 \operatorname{Tr}\left(\mathbb{E}\left[\boldsymbol{z}_i\right]^T \boldsymbol{W}^T \left(\boldsymbol{x}_i - \boldsymbol{\mu}\right)\right) - \right. \\
& \qquad \left. \operatorname{Tr}\left(\boldsymbol{W} \mathbb{E}\left[\boldsymbol{z}_i \boldsymbol{z}_i^T\right] \boldsymbol{W}^T\right) \right\} \\
= \quad & -\frac{1}{2\sigma^2} \sum_i \frac{\partial}{\partial \boldsymbol{W}} \left\{ 2 \operatorname{Tr}\left(\boldsymbol{W}^T \left(\boldsymbol{x}_i - \boldsymbol{\mu}\right) \mathbb{E}\left[\boldsymbol{z}_i\right]^T\right) - \right.
\end{aligned}
$$

$$
\begin{aligned}
& \mathrm{Tr}\left(\boldsymbol{W}\mathbb{E}\left[\boldsymbol{z}_i\boldsymbol{z}_i^T\right]\boldsymbol{W}^T\right)\Big\} \\
= & -\frac{1}{2\sigma^2}\sum_i\left(2\left(\boldsymbol{x}_i-\boldsymbol{\mu}\right)\mathbb{E}\left[\boldsymbol{z}_i\right]^T-\boldsymbol{W}\left(\mathbb{E}\left[\boldsymbol{z}_i\boldsymbol{z}_i^T\right]+\mathbb{E}\left[\boldsymbol{z}_i\boldsymbol{z}_i^T\right]^T\right)\right) \\
= & -\frac{1}{2\sigma^2}\sum_i\left(2\left(\boldsymbol{x}_i-\boldsymbol{\mu}\right)\mathbb{E}\left[\boldsymbol{z}_i\right]^T-2\boldsymbol{W}\mathbb{E}\left[\boldsymbol{z}_i\boldsymbol{z}_i^T\right]\right) \\
= & \; 0
\end{aligned}
$$

- ここで、行列に関するトレースの微分の公式を用いた

$$
\frac{\partial}{\partial \boldsymbol{A}}\mathrm{Tr}\left(\boldsymbol{A}\boldsymbol{B}\right) = \boldsymbol{B}^T \tag{22}
$$

$$
\frac{\partial}{\partial \boldsymbol{A}}\mathrm{Tr}\left(\boldsymbol{A}^T\boldsymbol{B}\right) = \boldsymbol{B} \tag{23}
$$

$$
\frac{\partial}{\partial \boldsymbol{A}}\mathrm{Tr}\left(\boldsymbol{A}^T\boldsymbol{B}\boldsymbol{A}\right) = \left(\boldsymbol{B}+\boldsymbol{B}^T\right)\boldsymbol{A} \tag{24}
$$

$$
\frac{\partial}{\partial \boldsymbol{A}}\mathrm{Tr}\left(\boldsymbol{A}\boldsymbol{B}\boldsymbol{A}^T\right) = \boldsymbol{A}\left(\boldsymbol{B}+\boldsymbol{B}^T\right) \tag{25}
$$

確率的主成分分析 (PPCA)

- 上式をパラメータ $\boldsymbol{W}$ について解く

$$
\begin{aligned}
& -\frac{1}{2\sigma^2}\sum_i \left(2\left(\boldsymbol{x}_i-\boldsymbol{\mu}\right)\mathbb{E}\left[\boldsymbol{z}_i\right]^T - 2\boldsymbol{W}\mathbb{E}\left[\boldsymbol{z}_i\boldsymbol{z}_i^T\right]\right) = 0 \\
\Rightarrow \quad & \boldsymbol{W}\left(\sum_i \mathbb{E}\left[\boldsymbol{z}_i\boldsymbol{z}_i^T\right]\right) = \sum_i \left(\boldsymbol{x}_i-\boldsymbol{\mu}\right)\mathbb{E}\left[\boldsymbol{z}_i\right]^T \\
\Rightarrow \quad & \boldsymbol{W} = \left(\sum_i \left(\boldsymbol{x}_i-\boldsymbol{\mu}\right)\mathbb{E}\left[\boldsymbol{z}_i\right]^T\right)\left(\sum_i \mathbb{E}\left[\boldsymbol{z}_i\boldsymbol{z}_i^T\right]\right)^{-1} \qquad (26)
\end{aligned}
$$

- これより、パラメータ $\boldsymbol{W}$ の更新式は次のようになる

$$
\boldsymbol{W}_{\mathrm{new}} \leftarrow \left(\sum_i \left(\boldsymbol{x}_i-\bar{\boldsymbol{x}}\right)\mathbb{E}\left[\boldsymbol{z}_i\right]^T\right)\left(\sum_i \mathbb{E}\left[\boldsymbol{z}_i\boldsymbol{z}_i^T\right]\right)^{-1} \qquad (27)
$$

- σ^2 について偏微分し、0 と等置すると次のようになる

$$
\begin{aligned}
& \frac{\partial}{\partial \sigma^2}\mathbb{E}\left[\ln p(\boldsymbol{X}, \boldsymbol{Z}|\boldsymbol{W}, \boldsymbol{\mu}, \sigma^2)\right] \\
= \quad & \frac{\partial}{\partial \sigma^2}\sum_i \Bigg\{ -\frac{D}{2}\ln(2\pi\sigma^2) - \frac{1}{2\sigma^2}||\boldsymbol{x}_i - \boldsymbol{\mu}||^2 + \\
& \quad \frac{1}{\sigma^2}\mathbb{E}\left[\boldsymbol{z}_i\right]^T \boldsymbol{W}^T\left(\boldsymbol{x}_i - \boldsymbol{\mu}\right) - \frac{1}{2\sigma^2}\operatorname{Tr}\left(\mathbb{E}\left[\boldsymbol{z}_i\boldsymbol{z}_i^T\right]\boldsymbol{W}^T\boldsymbol{W}\right)\Bigg\} \\
= \quad & \sum_i \Bigg\{ -\frac{D}{2}\frac{2\pi}{2\pi\sigma^2} + \frac{1}{2\sigma^4}||\boldsymbol{x}_i - \boldsymbol{\mu}||^2 - \\
& \quad \frac{1}{\sigma^4}\mathbb{E}\left[\boldsymbol{z}_i\right]^T \boldsymbol{W}^T\left(\boldsymbol{x}_i - \boldsymbol{\mu}\right) + \frac{1}{2\sigma^4}\operatorname{Tr}\left(\mathbb{E}\left[\boldsymbol{z}_i\boldsymbol{z}_i^T\right]\boldsymbol{W}^T\boldsymbol{W}\right)\Bigg\} \\
= \quad & -\frac{ND}{2}\frac{1}{\sigma^2} + \sum_i \Bigg\{ \frac{1}{2\sigma^4}||\boldsymbol{x}_i - \boldsymbol{\mu}||^2 - \\
& \quad \frac{1}{\sigma^4}\mathbb{E}\left[\boldsymbol{z}_i\right]^T \boldsymbol{W}^T\left(\boldsymbol{x}_i - \boldsymbol{\mu}\right) + \frac{1}{2\sigma^4}\operatorname{Tr}\left(\mathbb{E}\left[\boldsymbol{z}_i\boldsymbol{z}_i^T\right]\boldsymbol{W}^T\boldsymbol{W}\right)\Bigg\} = 0
\end{aligned}
$$

確率的主成分分析 (PPCA)

- 上式をパラメータ σ^2 について解く

$$-ND\sigma^2 + \sum_i \left\{ ||\boldsymbol{x}_i - \boldsymbol{\mu}||^2 - 2\mathbb{E}\left[\boldsymbol{z}_i\right]^T \boldsymbol{W}^T \left(\boldsymbol{x}_i - \boldsymbol{\mu}\right) + \mathrm{Tr}\left(\mathbb{E}\left[\boldsymbol{z}_i \boldsymbol{z}_i^T\right] \boldsymbol{W}^T \boldsymbol{W}\right) \right\} = 0$$

$$\Rightarrow \quad \sigma^2 = \frac{1}{ND} \sum_i \left\{ ||\boldsymbol{x}_i - \boldsymbol{\mu}||^2 - 2\mathbb{E}\left[\boldsymbol{z}_i\right]^T \boldsymbol{W}^T \left(\boldsymbol{x}_i - \boldsymbol{\mu}\right) + \mathrm{Tr}\left(\mathbb{E}\left[\boldsymbol{z}_i \boldsymbol{z}_i^T\right] \boldsymbol{W}^T \boldsymbol{W}\right) \right\} \tag{28}$$

- これより、パラメータ σ^2 の更新式は次のようになる

$$\sigma^2_{\mathrm{new}} \leftarrow \frac{1}{ND} \sum_i \left\{ ||\boldsymbol{x}_i - \bar{\boldsymbol{x}}||^2 - 2\mathbb{E}\left[\boldsymbol{z}_i\right]^T \boldsymbol{W}^T_{\mathrm{new}} \left(\boldsymbol{x}_i - \bar{\boldsymbol{x}}\right) + \mathrm{Tr}\left(\mathbb{E}\left[\boldsymbol{z}_i \boldsymbol{z}_i^T\right] \boldsymbol{W}^T_{\mathrm{new}} \boldsymbol{W}_{\mathrm{new}}\right) \right\} \tag{29}$$

確率的主成分分析 (PPCA)

- 以上で、観測データ $\mathcal{D} = \{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N\}$ から、主成分分析のパラメータ $\boldsymbol{W}, \sigma^2$ を推定するための EM アルゴリズムが得られた
- M ステップのパラメータの更新では、E ステップで求めた $\mathbb{E}[\boldsymbol{z}_i]$ と $\mathbb{E}\left[\boldsymbol{z}_i \boldsymbol{z}_i^T\right]$ を使用している
- これらの更新式を改良することで、オンライン型の EM アルゴリズムを導出することも可能である

目次

確率的主成分分析 (PPCA) のまとめ

EM アルゴリズムによる確率的主成分分析モデルの推定

- 目的は、確率的主成分分析モデルが与えられているとき、そのパラメータ $(\boldsymbol{W}, \boldsymbol{\mu}, \sigma^2)$ について、尤度関数 $\ln p(\boldsymbol{X}|\boldsymbol{W}, \boldsymbol{\mu}, \sigma^2)$ を最大化することである

1. $\boldsymbol{\mu}$ の決定: $\boldsymbol{\mu}$ を、観測データ $\boldsymbol{x}$ の標本平均として求める (これ以降、$\boldsymbol{\mu}$ は既知として扱い、$\bar{\boldsymbol{x}}$ と書く)

$$\boldsymbol{\mu} = \bar{\boldsymbol{x}} = \frac{1}{N} \sum_i \boldsymbol{x}_i \tag{30}$$

2. 初期化: パラメータ $\boldsymbol{W}_{\mathrm{old}}, \sigma^2_{\mathrm{old}}$ を適当に初期化し、対数尤度 $\ln p(\boldsymbol{X}|\boldsymbol{W}, \sigma^2)$ の初期値を計算する

確率的主成分分析 (PPCA) のまとめ

3. E ステップ: 現在のパラメータを用いて、以下の量を計算する

$$\begin{aligned}\mathbb{E}\left[\boldsymbol{z}_i\right] &= \boldsymbol{M}^{-1}\boldsymbol{W}_{\mathrm{old}}^T\left(\boldsymbol{x}_i - \bar{\boldsymbol{x}}\right) & (31)\\ \mathbb{E}\left[\boldsymbol{z}_i\boldsymbol{z}_i^T\right] &= \sigma_{\mathrm{old}}^2\boldsymbol{M}^{-1} + \mathbb{E}\left[\boldsymbol{z}_i\right]\mathbb{E}\left[\boldsymbol{z}_i\right]^T & (32)\end{aligned}$$

但し

$$\boldsymbol{M} = \boldsymbol{W}_{\mathrm{old}}^T\boldsymbol{W}_{\mathrm{old}} + \sigma_{\mathrm{old}}^2\boldsymbol{I} \quad (33)$$

確率的主成分分析 (PPCA) のまとめ

4. M ステップ: 現在の $\mathbb{E}[\boldsymbol{z}_i]$ と $\mathbb{E}\left[\boldsymbol{z}_i\boldsymbol{z}_i^T\right]$ を用いて、パラメータを更新する

$$\boldsymbol{W}_{\text{new}} \leftarrow \left(\sum_i (\boldsymbol{x}_i - \bar{\boldsymbol{x}})\,\mathbb{E}[\boldsymbol{z}_i]^T\right)\left(\sum_i \mathbb{E}\left[\boldsymbol{z}_i\boldsymbol{z}_i^T\right]\right)^{-1} \tag{34}$$

$$\begin{aligned}\sigma^2_{\text{new}} \leftarrow \frac{1}{ND}\sum_i \Big\{ &||\boldsymbol{x}_i - \bar{\boldsymbol{x}}||^2 - 2\mathbb{E}[\boldsymbol{z}_i]^T \boldsymbol{W}_{\text{new}}^T(\boldsymbol{x}_i - \bar{\boldsymbol{x}}) + \\ &\mathrm{Tr}\left(\mathbb{E}\left[\boldsymbol{z}_i\boldsymbol{z}_i^T\right]\boldsymbol{W}_{\text{new}}^T\boldsymbol{W}_{\text{new}}\right)\Big\}\end{aligned} \tag{35}$$

確率的主成分分析 (PPCA) のまとめ

5. 対数尤度関数 $\ln p(\boldsymbol{X}|\boldsymbol{W},\boldsymbol{\mu},\sigma^2)$ を計算する

$$\begin{aligned} & \ln p(\boldsymbol{X}|\boldsymbol{W},\boldsymbol{\mu},\sigma^2) \\ = & -\frac{ND}{2}\ln 2\pi - \frac{N}{2}\ln|\boldsymbol{C}| - \frac{1}{2}\sum_i (\boldsymbol{x}_i - \bar{\boldsymbol{x}})^T \boldsymbol{C}^{-1} (\boldsymbol{x}_i - \bar{\boldsymbol{x}}) \end{aligned}$$

但し

$$\boldsymbol{C} = \boldsymbol{W}_{\mathrm{new}}\boldsymbol{W}_{\mathrm{new}}^T + \sigma_{\mathrm{new}}^2 \boldsymbol{I} \tag{36}$$

パラメータの変化量、あるいは対数尤度の変化量をみて、収束性を判定する

6. 収束基準を満たしていなければ、(3) に戻る

$$\boldsymbol{W}_{\mathrm{old}} \leftarrow \boldsymbol{W}_{\mathrm{new}}, \quad \sigma_{\mathrm{old}}^2 \leftarrow \sigma_{\mathrm{new}}^2 \tag{37}$$