

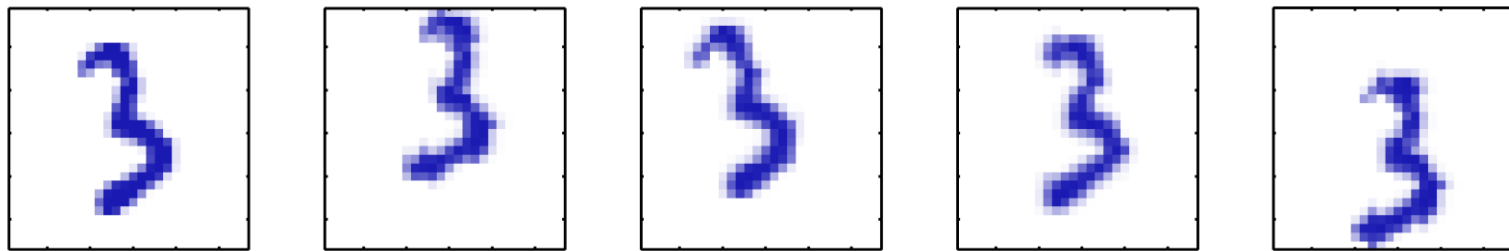


機械学習 Machine Learning

連続潜在変数 Continuous Latent Variables

福島 誠 Makoto Fukushima

情報科学部
School of Informatics and Data Science



データ空間の次元
Dimensionality
of the data space

$100 \times 100 = 10,000$ pixels

実効次元
Intrinsic
dimensionality

3 (垂直方向と水平方向の平行移動, そして回転)
(vertical and horizontal translations and rotations)

主成分分析 Principal component analysis (PCA)

主成分空間と呼ばれる低次元の線形空間上へのデータ点の直行射影を考える。

Consider the orthogonal projection of a data set onto a lower-dimensional linear space, known as the principal subspace.

主成分分析の2種類の異なる定義

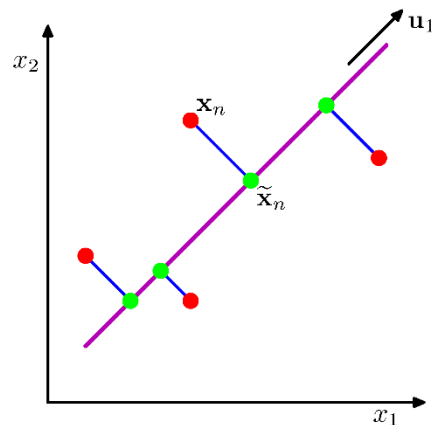
Two different definitions of PCA

- 1) 射影されたデータの分散を最大化する線形射影。

The linear projection that maximizes the variance of the projected data. Hotelling (1933)

- 2) もともとのデータ点と射影した点の間の2乗距離の平均値で定義される射影のコスト関数を最小化する線形射影。 Pearson (1901)

The linear projection that minimizes the average projection cost, defined as the mean squared distance between the data points and their projections.



主成分分析 Principal component analysis (PCA)

分散最大化による定式化 Maximum variance formulation

観測値のデータ集合 A data set of observations

$\{\mathbf{x}_n\}$ $n = 1, \dots, N$ D 次元の変数 D dimensional variable

目的は、射影されたデータ点の分散を最大化しながら、データ点を次元 $M < D$ の空間上に射影すること。

The goal is to project the data onto a space having dimensionality $M < D$ while maximizing the variance of the projected data.

主成分分析 Principal component analysis (PCA)

分散最大化による定式化 Maximum variance formulation

$M = 1$ のとき When $M = 1$

射影されたデータの分散は

The variance of the projected data is

$$\frac{1}{N} \sum_{n=1}^N \{ \mathbf{u}_1^T \mathbf{x}_n - \mathbf{u}_1^T \bar{\mathbf{x}} \}^2 = \mathbf{u}_1^T \mathbf{S} \mathbf{u}_1$$

$$\mathbf{u}_1^T \mathbf{u}_1 = 1$$

$$\bar{\mathbf{x}} = \frac{1}{N} \sum_{n=1}^N \mathbf{x}_n$$

$$\mathbf{S} = \frac{1}{N} \sum_{n=1}^N (\mathbf{x}_n - \bar{\mathbf{x}})(\mathbf{x}_n - \bar{\mathbf{x}})^T$$

$\mathbf{u}_1^T \mathbf{S} \mathbf{u}_1 + \lambda_1 (1 - \mathbf{u}_1^T \mathbf{u}_1)$ を \mathbf{u}_1 について最大化する.

Maximize $\mathbf{u}_1^T \mathbf{S} \mathbf{u}_1 + \lambda_1 (1 - \mathbf{u}_1^T \mathbf{u}_1)$ with respect to \mathbf{u}_1 .

$\Leftrightarrow \mathbf{S} \mathbf{u}_1 = \lambda_1 \mathbf{u}_1 \Leftrightarrow \mathbf{u}_1^T \mathbf{S} \mathbf{u}_1 = \lambda_1 \Leftrightarrow$ \mathbf{u}_1 は \mathbf{S} の最大固有値 λ_1 に属する固有ベクトル
 \mathbf{u}_1 equal to the eigenvector of \mathbf{S} having the largest eigenvalue λ_1

主成分分析 Principal component analysis (PCA)

分散最大化による定式化 Maximum variance formulation

$M > 1$ のとき When $M > 1$

データ共分散行列 S の, 大きい順に M 個の固有値 $\lambda_1, \dots, \lambda_M$ に対応する M 個の固有ベクトル $\mathbf{u}_1, \dots, \mathbf{u}_M$ により, 射影されたデータの分散を最大にする最適な線形射影が定義される.

The optimal linear projection for which the variance of the projected data is maximized is no defined by the M eigenvectors $\mathbf{u}_1, \dots, \mathbf{u}_M$ of the data covariance matrix S corresponding to the M largest eigenvalues $\lambda_1, \dots, \lambda_M$.

主成分分析 Principal component analysis (PCA)

誤差最小化による定式化 Minimum-error formulation

$$\{\mathbf{u}_i\} \quad i = 1, \dots, D \quad \mathbf{u}_i^T \mathbf{u}_j = \delta_{ij}$$

各データ点は厳密に、完全正規直交系をなす基底ベクトル $\{\mathbf{u}_i\}$ の線形結合で表すことができる。

Each data point can be represented exactly by a linear combination of a complete orthonormal set of basis vectors $\{\mathbf{u}_i\}$.

$$\begin{aligned} \mathbf{x}_n &= \sum_{i=1}^D \alpha_{ni} \mathbf{u}_i \\ &= \sum_{i=1}^D (\mathbf{x}_n^T \mathbf{u}_i) \mathbf{u}_i \end{aligned} \quad \longleftarrow \quad \alpha_{nj} = \mathbf{x}_n^T \mathbf{u}_j$$

主成分分析 Principal component analysis (PCA)

誤差最小化による定式化 Minimum-error formulation

各データ点 \mathbf{x}_n を以下のように近似する.

We approximate each data point \mathbf{x}_n by

$$\tilde{\mathbf{x}}_n = \sum_{i=1}^M z_{ni} \mathbf{u}_i + \sum_{i=M+1}^D b_i \mathbf{u}_i \quad M < D$$

低次元化による歪み Error introduced by the reduction in dimensionality

$$J = \frac{1}{N} \sum_{n=1}^N \|\mathbf{x}_n - \tilde{\mathbf{x}}_n\|^2$$

J を最小にする $\{\mathbf{u}_i\}$, $\{z_{ni}\}$, $\{b_i\}$ を求める.

Derive $\{\mathbf{u}_i\}$, $\{z_{ni}\}$, and $\{b_i\}$ that minimize J .

主成分分析 Principal component analysis (PCA)

誤差最小化による定式化 Minimum-error formulation

J を最小にする $\{z_{ni}\}$, $\{b_i\}$ は
 $\{z_{ni}\}$ and $\{b_i\}$ that minimize J are

$$z_{nj} = \mathbf{x}_n^T \mathbf{u}_j \quad j = 1, \dots, M$$

$$b_j = \bar{\mathbf{x}}^T \mathbf{u}_j \quad j = M+1, \dots, D$$

$$J = \frac{1}{N} \sum_{n=1}^N \|\mathbf{x}_n - \tilde{\mathbf{x}}_n\|^2$$
$$\tilde{\mathbf{x}}_n = \sum_{i=1}^M z_{ni} \mathbf{u}_i + \sum_{i=M+1}^D b_i \mathbf{u}_i$$
$$\bar{\mathbf{x}} = \frac{1}{N} \sum_{n=1}^N \mathbf{x}_n \quad \mathbf{u}_i^T \mathbf{u}_j = \delta_{ij}$$

主成分分析 Principal component analysis (PCA)

誤差最小化による定式化 Minimum-error formulation

J を最小にする $\{\mathbf{u}_i\}$ は . . .
 $\{\mathbf{u}_i\}$ that minimize J are ...

$$\mathbf{x}_n - \tilde{\mathbf{x}}_n = \sum_{i=M+1}^D \{(\mathbf{x}_n - \bar{\mathbf{x}})^T \mathbf{u}_i\} \mathbf{u}_i$$

$$\Rightarrow J = \frac{1}{N} \sum_{n=1}^N \sum_{i=M+1}^D (\mathbf{x}_n^T \mathbf{u}_i - \bar{\mathbf{x}}^T \mathbf{u}_i)^2$$
$$= \sum_{i=M+1}^D \mathbf{u}_i^T \mathbf{S} \mathbf{u}_i$$

$$J = \frac{1}{N} \sum_{n=1}^N \|\mathbf{x}_n - \tilde{\mathbf{x}}_n\|^2$$

$$\mathbf{x}_n = \sum_{i=1}^D (\mathbf{x}_n^T \mathbf{u}_i) \mathbf{u}_i$$

$$\mathbf{S} = \frac{1}{N} \sum_{n=1}^N (\mathbf{x}_n - \bar{\mathbf{x}})(\mathbf{x}_n - \bar{\mathbf{x}})^T$$

$$\mathbf{u}_i^T \mathbf{u}_j = \delta_{ij}$$

主成分分析 Principal component analysis (PCA)

誤差最小化による定式化 Minimum-error formulation

$$D = 2 \quad M = 1$$

J を最小にする \mathbf{u}_2 は, 小さい方の固有値に対応した固有ベクトルとなる.

\mathbf{u}_2 that minimize J is the eigenvector corresponding to the smaller eigenvalue.

$$J = \mathbf{u}_2^T \mathbf{S} \mathbf{u}_2$$

$$\tilde{J} = \mathbf{u}_2^T \mathbf{S} \mathbf{u}_2 + \lambda_2 (1 - \mathbf{u}_2^T \mathbf{u}_2)$$

\tilde{J} の \mathbf{u}_2 に関する導関数を0とおくと,

By setting the derivative of \tilde{J} w.r.t. \mathbf{u}_2 to 0,

$$\mathbf{S} \mathbf{u}_2 = \lambda_2 \mathbf{u}_2$$

$$J = \lambda_2$$

$$J = \frac{1}{N} \sum_{n=1}^N \|\mathbf{x}_n - \tilde{\mathbf{x}}_n\|^2$$

$$\mathbf{x}_n = \sum_{i=1}^D (\mathbf{x}_n^T \mathbf{u}_i) \mathbf{u}_i$$

$$\mathbf{S} = \frac{1}{N} \sum_{n=1}^N (\mathbf{x}_n - \bar{\mathbf{x}})(\mathbf{x}_n - \bar{\mathbf{x}})^T$$

$$\mathbf{u}_i^T \mathbf{u}_j = \delta_{ij}$$

主成分分析 Principal component analysis (PCA)

誤差最小化による定式化 Minimum-error formulation

J を最小にする $\{\mathbf{u}_i\}$ は, 以下の式で与えられる共分散行列の固有ベクトルとなる.
 $\{\mathbf{u}_i\}$ that minimize J are eigenvectors of the covariance matrix given by

$$\mathbf{S}\mathbf{u}_i = \lambda_i \mathbf{u}_i$$

対応する歪み尺度の値

The corresponding value of the error measures

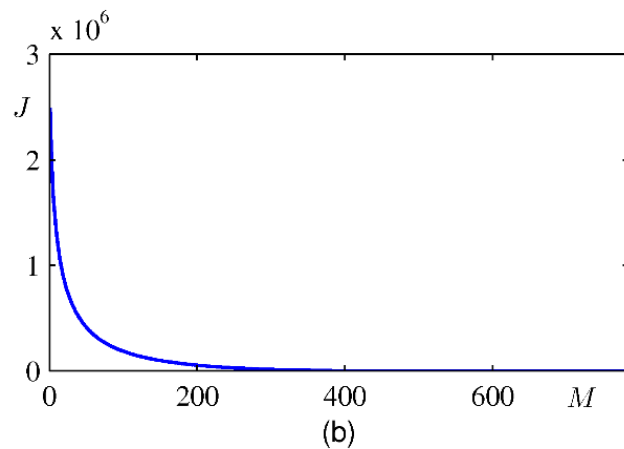
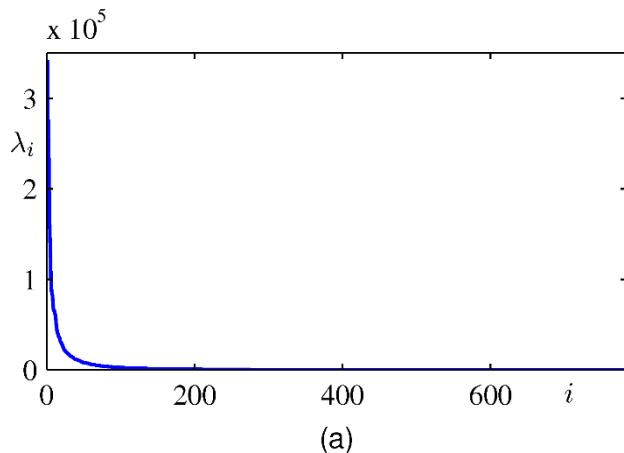
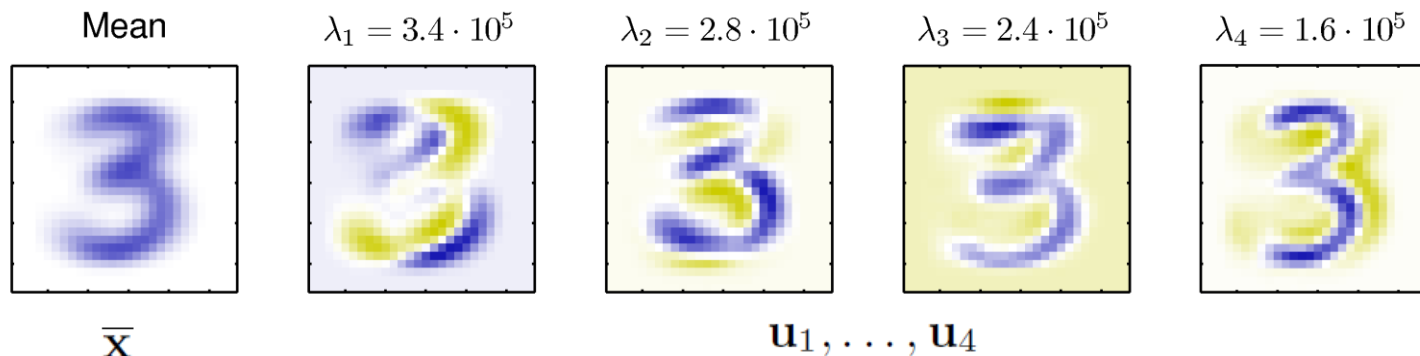
$$J = \sum_{i=M+1}^D \lambda_i$$

は, 小さい順から $D - M$ 個の固有値に対応する固有ベクトルを $\{\mathbf{u}_i\}$
($i = M + 1, \dots, D$) として選ぶことで最小になる.

is minimized by selecting the eigenvectors $\{\mathbf{u}_i\}$ for $i = M + 1, \dots, D$
to be those having the $D - M$ smallest eigenvalues.

主成分分析 Principal component analysis (PCA)

データ圧縮 Data compression

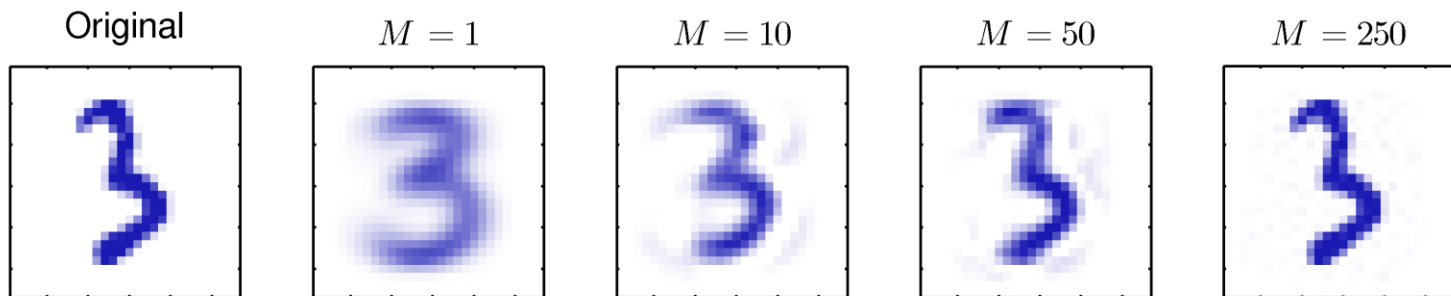


主成分分析 Principal component analysis (PCA)

データ圧縮 Data compression

\mathbf{x}_n に対する主成分近似 PCA approximation to \mathbf{x}_n

$$\begin{aligned}\tilde{\mathbf{x}}_n &= \sum_{i=1}^M (\mathbf{x}_n^T \mathbf{u}_i) \mathbf{u}_i + \sum_{i=M+1}^D (\bar{\mathbf{x}}^T \mathbf{u}_i) \mathbf{u}_i \\ &= \bar{\mathbf{x}} + \sum_{i=1}^M (\mathbf{x}_n^T \mathbf{u}_i - \bar{\mathbf{x}}^T \mathbf{u}_i) \mathbf{u}_i \quad \bar{\mathbf{x}} = \sum_{i=1}^D (\bar{\mathbf{x}}^T \mathbf{u}_i) \mathbf{u}_i\end{aligned}$$



主成分分析 Principal component analysis (PCA)

データ白色化 Data whitening

データ標準化 Standardizing the data

平均 Mean = 0

分散 Variance = 1

共分散行列の要素 Elements of the covariance matrix

$$\rho_{ij} = \frac{1}{N} \sum_{n=1}^N \frac{(x_{ni} - \bar{x}_i)}{\sigma_i} \frac{(x_{nj} - \bar{x}_j)}{\sigma_j}$$

データ白色化 Whitening the data

\mathbf{x}_n の変換 Transformation of \mathbf{x}_n : $\mathbf{y}_n = \mathbf{L}^{-1/2} \mathbf{U}^T (\mathbf{x}_n - \bar{\mathbf{x}})$

⇒ 平均 Mean = 0; 共分散行列 Covariance matrix = \mathbf{I}

$$\begin{aligned} \frac{1}{N} \sum_{n=1}^N \mathbf{y}_n \mathbf{y}_n^T &= \frac{1}{N} \sum_{n=1}^N \mathbf{L}^{-1/2} \mathbf{U}^T (\mathbf{x}_n - \bar{\mathbf{x}}) (\mathbf{x}_n - \bar{\mathbf{x}})^T \mathbf{U} \mathbf{L}^{-1/2} \\ &= \mathbf{L}^{-1/2} \mathbf{U}^T \mathbf{S} \mathbf{U} \mathbf{L}^{-1/2} = \mathbf{L}^{-1/2} \mathbf{L} \mathbf{L}^{-1/2} = \mathbf{I} \end{aligned}$$

$$\mathbf{S} \mathbf{u}_i = \lambda_i \mathbf{u}_i$$

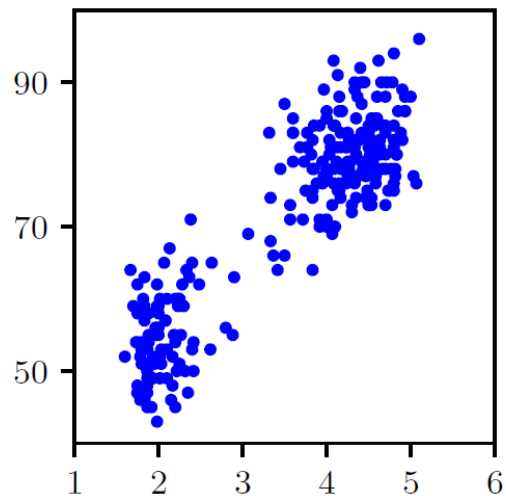
$$\mathbf{S} \mathbf{U} = \mathbf{U} \mathbf{L}$$

$$\bar{\mathbf{x}} = \frac{1}{N} \sum_{n=1}^N \mathbf{x}_n$$

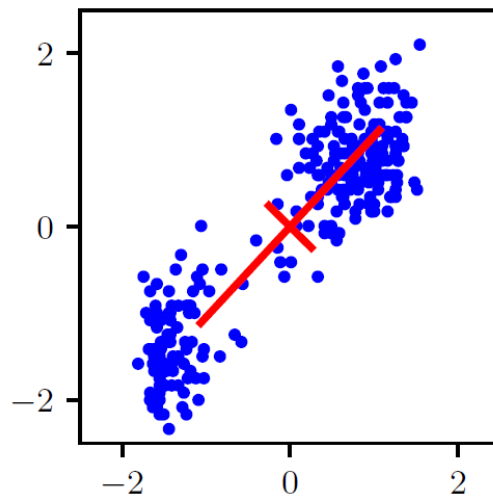
主成分分析 Principal component analysis (PCA)

データ白色化 Data whitening

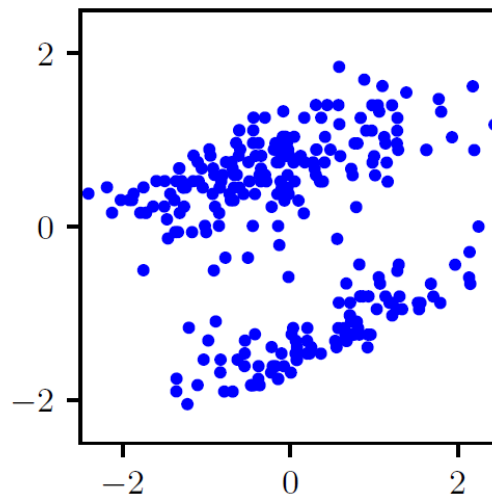
もともとのデータ
Original data



標準化されたデータ
Standardized data



白色化されたデータ
Whitened data



主成分分析 Principal component analysis (PCA)

高次元データ High-dimensional data

$$N < D$$

\mathbf{X} の n 行目 n -th row of \mathbf{X} : $(\mathbf{x}_n - \bar{\mathbf{x}})^T$

共分散行列 Covariance matrix: $\mathbf{S} = N^{-1}\mathbf{X}^T\mathbf{X}$

\mathbf{X} のサイズ
The size of \mathbf{X}
 $N \times D$

$$\begin{array}{lcl}
 \begin{array}{c}
 \frac{1}{N}\mathbf{X}^T\mathbf{X}\mathbf{u}_i = \lambda_i\mathbf{u}_i \\
 \downarrow \\
 \frac{1}{N}\mathbf{X}\mathbf{X}^T(\mathbf{X}\mathbf{u}_i) = \lambda_i(\mathbf{X}\mathbf{u}_i) \\
 \downarrow \\
 \frac{1}{N}\mathbf{X}\mathbf{X}^T\mathbf{v}_i = \lambda_i\mathbf{v}_i
 \end{array}
 & \xrightarrow{\hspace{1cm}} &
 \begin{array}{l}
 \left(\frac{1}{N}\mathbf{X}^T\mathbf{X}\right)(\mathbf{X}^T\mathbf{v}_i) = \lambda_i(\mathbf{X}^T\mathbf{v}_i) \rightarrow \mathbf{u}_i \propto \mathbf{X}^T\mathbf{v}_i \\
 \downarrow \\
 \mathbf{u}_i = \frac{1}{(N\lambda_i)^{1/2}}\mathbf{X}^T\mathbf{v}_i
 \end{array}
 \end{array}$$

$D \times D$ ----- $N \times N$

$\|\mathbf{u}_i\| = 1$

確率的潜在変数 Probabilistic latent variables

確率的主成分分析 Probabilistic PCA Tipping and Bishop (1997, 1999); Roweis (1998)

主成分分析は、ある確率的潜在変数モデルの最尤解としても表現される。
PCA can also be expressed as the maximum likelihood solution of a probabilistic latent variable model.

この主成分分析の再定式化は確率的主成分分析として知られている。
This reformulation of PCA is known as probabilistic PCA.

後で見るように、これは因子分析と密接な関係をもっている。
As we will see later, it is closely related to factor analysis. Basilevsky (1994)

確率的潜在変数 Probabilistic latent variables

確率的主成分分析の利点 Advantages of probabilistic PCA

確率的主成分分析は制約付きのガウス分布に基づいており、モデルがデータ集合の主要な相関の構造を捉えることを可能にしつつ、自由パラメータの数を制限できる。 Probabilistic PCA represents a constrained form of a Gaussian distribution in which the number of free parameters can be restricted while still allowing the model to capture the dominant correlations in a data set.

主成分分析を行うためのEMアルゴリズムを導くことができ、これは上位の固有ベクトルのみが必要な状況では計算効率が良く、途中でデータ共分散行列を計算する必要もない。

We can derive an EM algorithm for PCA that is computationally efficient in situations where only a few leading eigenvectors are required and that avoid having to evaluate the data covariance matrix as an intermediate step.

確率的潜在変数 Probabilistic latent variables

確率的主成分分析の利点 Advantages of probabilistic PCA

確率モデルとEMアルゴリズムの組み合わせによりデータ集合内の欠損値を扱える。
The combination of a probabilistic model and EM allows us to deal with missing values in the data set.

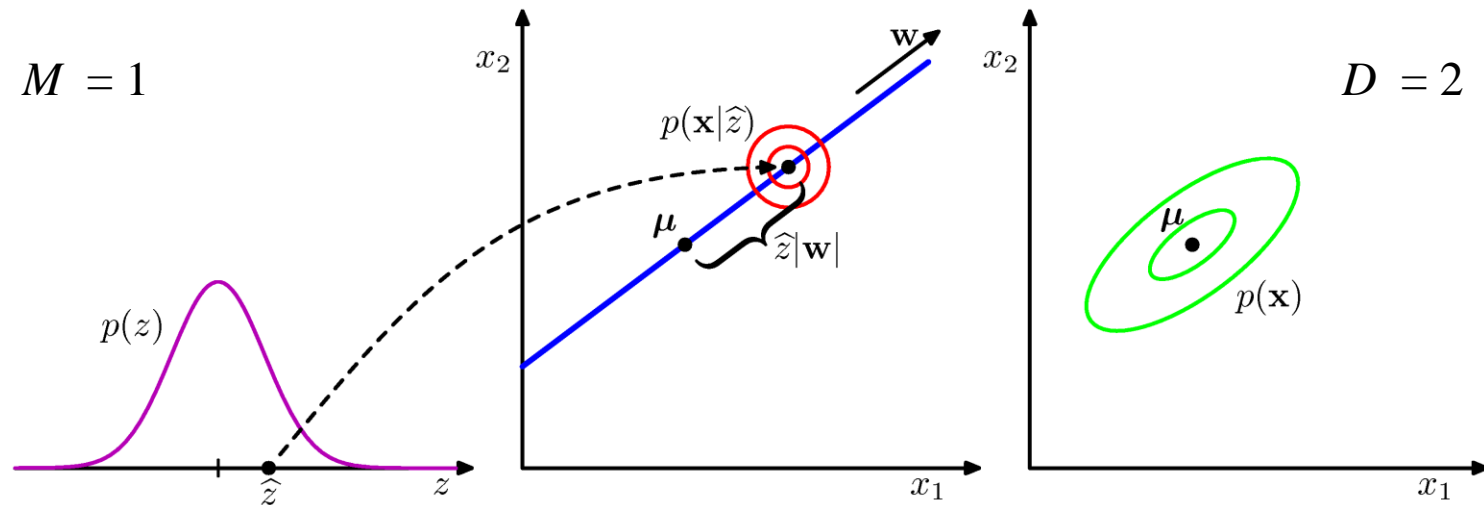
尤度関数が得られるので他の確率密度モデルとの直接の比較ができる。
The existence of a likelihood function allows direct comparison with other probability density models.

確率的主成分分析はデータサンプルを分布から得るための生成モデルとしても利用できる。
A probabilistic PCA model can be run generatively to provide samples from the distribution.

確率的潜在変数 Probabilistic latent variables

生成モデル Generative model

確率の主成分分析
Probabilistic PCA



$$p(\mathbf{z}) = \mathcal{N}(\mathbf{z}|\mathbf{0}, \mathbf{I}) \xrightarrow{\quad} p(\mathbf{x}|\mathbf{z}) = \mathcal{N}(\mathbf{x}|\mathbf{W}\mathbf{z} + \mu, \sigma^2\mathbf{I}) \quad p(\mathbf{x}) = \int p(\mathbf{x}|\mathbf{z})p(\mathbf{z}) d\mathbf{z}$$

\uparrow
 $\mathbf{x} = \mathbf{W}\mathbf{z} + \mu + \epsilon$

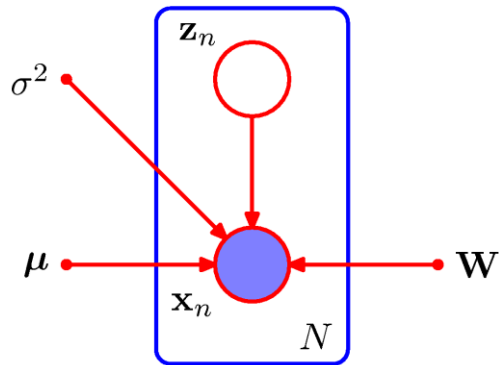
確率的潜在変数 Probabilistic latent variables

尤度関数 Likelihood function

確率の主成分分析
Probabilistic PCA

尤度関数 Likelihood function

$$p(\mathbf{X}|\boldsymbol{\mu}, \mathbf{W}, \sigma^2) \quad \mathbf{X} = \{\mathbf{x}_n\}$$



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

$$p(\mathbf{z}) = \mathcal{N}(\mathbf{z}|\mathbf{0}, \mathbf{I})$$

$$p(\mathbf{x}) = \int p(\mathbf{x}|\mathbf{z})p(\mathbf{z}) d\mathbf{z}$$

$$\Rightarrow p(\mathbf{x}) = \mathcal{N}(\mathbf{x}|\boldsymbol{\mu}, \mathbf{C})$$
$$\mathbf{C} = \mathbf{W}\mathbf{W}^T + \sigma^2\mathbf{I}$$

$$\begin{aligned}\mathbb{E}[\mathbf{x}] &= \mathbb{E}[\mathbf{W}\mathbf{z} + \boldsymbol{\mu} + \boldsymbol{\epsilon}] = \boldsymbol{\mu} \\ \text{cov}[\mathbf{x}] &= \mathbb{E}[(\mathbf{W}\mathbf{z} + \boldsymbol{\epsilon})(\mathbf{W}\mathbf{z} + \boldsymbol{\epsilon})^T] \\ &= \mathbb{E}[\mathbf{W}\mathbf{z}\mathbf{z}^T\mathbf{W}^T] + \mathbb{E}[\boldsymbol{\epsilon}\boldsymbol{\epsilon}^T] \\ &= \mathbf{W}\mathbf{W}^T + \sigma^2\mathbf{I}\end{aligned}$$

確率的潜在変数 Probabilistic latent variables

尤度関数 Likelihood function

確率の主成分分析
Probabilistic PCA

潜在変数空間の座標の回転不変性に関する冗長性

Redundancy in the parameterization corresponding to rotations of the latent space coordinates

$$\widetilde{\mathbf{W}}\widetilde{\mathbf{W}}^T = \mathbf{W}\mathbf{R}\mathbf{R}^T\mathbf{W}^T = \mathbf{W}\mathbf{W}^T$$

$$\widetilde{\mathbf{W}} = \mathbf{W}\mathbf{R}$$

$$\mathbf{R}\mathbf{R}^T = \mathbf{I}$$

確率的潜在変数 Probabilistic latent variables

尤度関数 Likelihood function

確率の主成分分析
Probabilistic PCA

効率的な C^{-1} の計算 Efficient computation of C^{-1} $\mathcal{O}(D^3) \rightarrow \mathcal{O}(M^3)$

$$C^{-1} = \sigma^{-2} \mathbf{I} - \sigma^{-2} \mathbf{W} \mathbf{M}^{-1} \mathbf{W}^T$$

$$\mathbf{M} = \mathbf{W}^T \mathbf{W} + \sigma^2 \mathbf{I}$$

$$\begin{aligned} & (\mathbf{A} + \mathbf{B} \mathbf{D}^{-1} \mathbf{C})^{-1} \\ &= \mathbf{A}^{-1} - \mathbf{A}^{-1} \mathbf{B} (\mathbf{D} + \mathbf{C} \mathbf{A}^{-1} \mathbf{B})^{-1} \mathbf{C} \mathbf{A}^{-1} \end{aligned}$$

事後分布 Posterior distribution

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

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

$$p(\mathbf{z}) = \mathcal{N}(\mathbf{z} | \mathbf{0}, \mathbf{I})$$

確率的潜在変数 Probabilistic latent variables

最尤推定 Maximum likelihood

確率の主成分分析
Probabilistic PCA

対数尤度関数 Log likelihood function

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

$$p(\mathbf{x}) = \mathcal{N}(\mathbf{x}|\boldsymbol{\mu}, \mathbf{C})$$

$\boldsymbol{\mu}$ の最尤解 Maximum likelihood solution of $\boldsymbol{\mu}$: $\boldsymbol{\mu} = \bar{\mathbf{x}} = \frac{1}{N} \sum_{n=1}^N \mathbf{x}_n$

$$\Rightarrow \ln p(\mathbf{X}|\mathbf{W}, \boldsymbol{\mu}, \sigma^2) = -\frac{N}{2} \{ D \ln(2\pi) + \ln |\mathbf{C}| + \text{Tr}(\mathbf{C}^{-1} \mathbf{S}) \}$$

$$\mathbf{S} = \frac{1}{N} \sum_{n=1}^N (\mathbf{x}_n - \bar{\mathbf{x}})(\mathbf{x}_n - \bar{\mathbf{x}})^T$$

確率的潜在変数 Probabilistic latent variables

最尤推定 Maximum likelihood

確率的主成分分析
Probabilistic PCA

\mathbf{W} の最尤解 Maximum likelihood solution of \mathbf{W} Tipping and Bishop (1999)

$$\mathbf{W}_{\text{ML}} = \mathbf{U}_M (\mathbf{L}_M - \sigma^2 \mathbf{I})^{1/2} \mathbf{R}$$

\mathbf{U}_M : 列は, データ共分散行列 \mathbf{S} の, 上位 M 個の固有値に対応する M 個の固有ベクトル

Columns are M eigenvectors of the data covariance matrix \mathbf{S} whose eigenvalues are the M largest.

\mathbf{L}_M : データ共分散行列 \mathbf{S} の, 上位 M 個の固有値を要素にもつ対角行列
The diagonal matrix whose elements are the M largest eigenvalues of the data covariance matrix \mathbf{S}

\mathbf{R} : 任意の直交行列 An arbitrary orthogonal matrix

確率的潜在変数 Probabilistic latent variables

最尤推定 Maximum likelihood

確率の主成分分析
Probabilistic PCA

\mathbf{W} の最尤解 Maximum likelihood solution of \mathbf{W} Tipping and Bishop (1999)

$$\mathbf{W}_{\text{ML}} = \mathbf{U}_M (\mathbf{L}_M - \sigma^2 \mathbf{I})^{1/2} \mathbf{R}$$

$\mathbf{u}_1, \dots, \mathbf{u}_M$: 上位 M 個の固有ベクトル
The M principal eigenvectors

σ^2 の最尤解 Maximum likelihood solution of σ^2

$$\sigma_{\text{ML}}^2 = \frac{1}{D - M} \sum_{i=M+1}^D \lambda_i$$

確率的潜在変数 Probabilistic latent variables

最尤推定 Maximum likelihood

確率の主成分分析
Probabilistic PCA

事後分布 Posterior distribution

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

事後平均 Posterior mean

$$\mathbb{E}[\mathbf{z}|\mathbf{x}] = \mathbf{M}^{-1}\mathbf{W}_{\text{ML}}^T(\mathbf{x} - \bar{\mathbf{x}})$$

$$\sigma^2 \rightarrow 0 \quad (\mathbf{W}_{\text{ML}}^T\mathbf{W}_{\text{ML}})^{-1}\mathbf{W}_{\text{ML}}^T(\mathbf{x} - \bar{\mathbf{x}})$$

データ点の潜在変数空間上への直交射影
An orthogonal projection of the data
point onto the latent space

確率的潜在変数 Probabilistic latent variables

最尤推定 Maximum likelihood

共分散行列の自由度

The degrees of freedom of the covariance matrix

一般のガウス分布の共分散行列

The covariance matrix of
a general Gaussian distribution

$$D(D + 1)/2$$

確率的主成分分析

Probabilistic PCA

$$\mathbf{C} = \mathbf{W}\mathbf{W}^T + \sigma^2\mathbf{I}$$

$$DM + 1 - \frac{M(M - 1)}{2}$$

直交行列 \mathbf{R} の自由度

The degrees of freedom
of the orthogonal matrix \mathbf{R}

確率的潜在変数 Probabilistic latent variables

因子分析 Factor analysis

確率の主成分分析
Probabilistic PCA

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

$$p(\mathbf{z}) = \mathcal{N}(\mathbf{z}|\mathbf{0}, \mathbf{I})$$

$$p(\mathbf{x}) = \mathcal{N}(\mathbf{x}|\boldsymbol{\mu}, \mathbf{C})$$

$$\mathbf{C} = \mathbf{W}\mathbf{W}^T + \sigma^2\mathbf{I}$$

因子分析
Factor analysis

$$p(\mathbf{x}|\mathbf{z}) = \mathcal{N}(\mathbf{x}|\mathbf{W}\mathbf{z} + \boldsymbol{\mu}, \boldsymbol{\Psi})$$

$$p(\mathbf{z}) = \mathcal{N}(\mathbf{z}|\mathbf{0}, \mathbf{I})$$

$$p(\mathbf{x}) = \mathcal{N}(\mathbf{x}|\boldsymbol{\mu}, \mathbf{C})$$

$$\mathbf{C} = \mathbf{W}\mathbf{W}^T + \boldsymbol{\Psi}$$

$\boldsymbol{\Psi}$: 対角行列
A diagonal matrix

確率的潜在変数 Probabilistic latent variables

独立成分分析 Independent component analysis (ICA)

独立成分分析では潜在変数の分布が以下のように分解されていると考える.
ICA arises when we consider a distribution over the latent variables that factorizes, so that

$$p(\mathbf{z}) = \prod_{j=1}^M p(z_j)$$

潜在変数は非ガウス分布に従うことが必要.

The latent variables are required to have non-Gaussian distributions.

例 e.g.:
$$p(z_j) = \frac{1}{\pi \cosh(z_j)} = \frac{2}{\pi(e^{z_j} + e^{-z_j})}$$

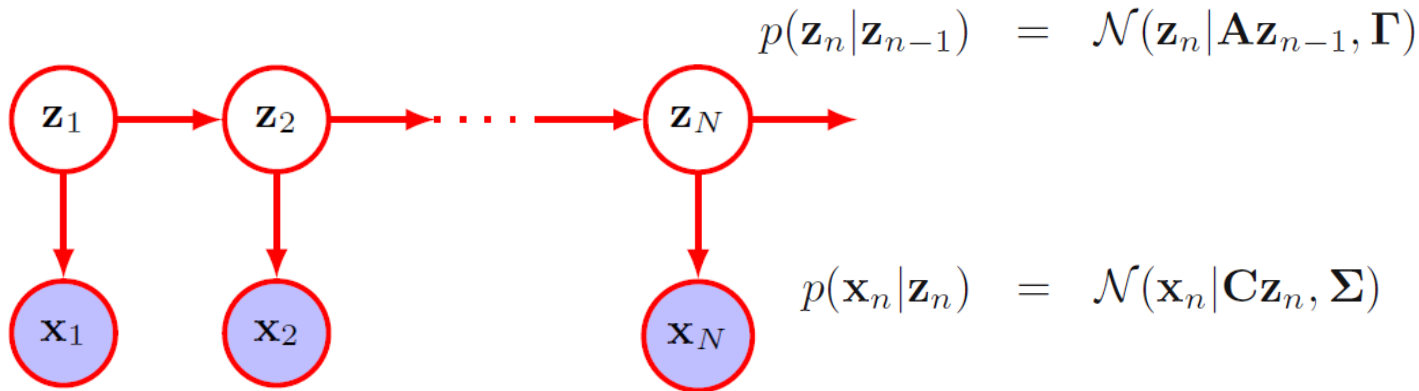
確率的潜在変数 Probabilistic latent variables

カルマンフィルタ Kalman filters

線形動的システム Linear dynamical system
or カルマンフィルタ Kalman filter

潜在変数
Latent
variables

観測変数
Observed
variables



エビデンスの下限 Evidence lower bound (ELBO)

モデル A model: $p(\mathbf{x}, \mathbf{z}|\mathbf{w})$

対数尤度関数 Log likelihood function

$$\ln p(\mathbf{x}|\mathbf{w}) = \mathcal{L}(q, \mathbf{w}) + \text{KL}(q(\mathbf{z})\|p(\mathbf{z}|\mathbf{x}, \mathbf{w}))$$

ただし where $\mathcal{L}(q, \mathbf{w}) = \int q(\mathbf{z}) \ln \left\{ \frac{p(\mathbf{x}, \mathbf{z}|\mathbf{w})}{q(\mathbf{z})} \right\} d\mathbf{z}$

$$\text{KL}(q(\mathbf{z})\|p(\mathbf{z}|\mathbf{x}, \mathbf{w})) = - \int q(\mathbf{z}) \ln \left\{ \frac{p(\mathbf{z}|\mathbf{x}, \mathbf{w})}{q(\mathbf{z})} \right\} d\mathbf{z}$$

$$\text{KL}(q(\mathbf{z})\|p(\mathbf{z}|\mathbf{x}, \mathbf{w})) \geq 0 \quad \Rightarrow \quad \ln p(\mathbf{x}|\mathbf{w}) \geq \mathcal{L}(q, \mathbf{w}) : \text{エビデンスの下限}$$

Evidence lower bound
(ELBO)

エビデンスの下限 Evidence lower bound (ELBO)

EMアルゴリズムによる対数尤度最大化

Maximization of the log likelihood function by the EM algorithm

Eステップ E step $\mathcal{L}(q, \mathbf{w})$ の $q(\mathbf{z})$ についての最大化
Maximization of $\mathcal{L}(q, \mathbf{w})$ w.r.t. $q(\mathbf{z})$

$$\Rightarrow q(\mathbf{z}) = p(\mathbf{z}|\mathbf{x}, \mathbf{w}^{(\text{old})})$$

Mステップ M step $\mathcal{L}(q, \mathbf{w})$ の \mathbf{w} についての最大化
Maximization of $\mathcal{L}(q, \mathbf{w})$ w.r.t. \mathbf{w}

$$\begin{aligned}\mathcal{L}(q, \mathbf{w}) &= \int p(\mathbf{z}|\mathbf{x}, \mathbf{w}^{(\text{old})}) \ln p(\mathbf{x}, \mathbf{z}|\mathbf{w}) d\mathbf{z} \\ &\quad - \int p(\mathbf{z}|\mathbf{x}, \mathbf{w}^{(\text{old})}) \ln p(\mathbf{z}|\mathbf{x}, \mathbf{w}^{(\text{old})}) d\mathbf{z}\end{aligned}$$

エビデンスの下限 Evidence lower bound (ELBO)

i.i.d.観測値データ集合 $\mathbf{x}_1, \dots, \mathbf{x}_N$ に対しては,
If we have a data set $\mathbf{x}_1, \dots, \mathbf{x}_N$ of i.i.d. observations,

対数尤度関数 Log likelihood function

$$\ln p(\mathbf{X}|\mathbf{w}) = \sum_{n=1}^N \ln p(\mathbf{x}_n|\mathbf{w})$$

エビデンスの下限 ELBO

$$\mathcal{L}(q, \mathbf{w}) = \sum_{n=1}^N \int q(\mathbf{z}_n) \ln \left\{ \frac{p(\mathbf{x}_n, \mathbf{z}_n|\mathbf{w})}{q(\mathbf{z}_n)} \right\} d\mathbf{z}_n$$

エビデンスの下限 Evidence lower bound (ELBO)

EM Expectation maximization

確率的主成分分析
Probabilistic PCA

完全データ対数尤度関数 Complete-data log likelihood function

$$\ln p(\mathbf{X}, \mathbf{Z} | \boldsymbol{\mu}, \mathbf{W}, \sigma^2) = \sum_{n=1}^N \{\ln p(\mathbf{x}_n | \mathbf{z}_n) + \ln p(\mathbf{z}_n)\}$$

期待完全データ対数尤度関数
Expected complete-data
log likelihood function

$$\begin{aligned} p(\mathbf{x} | \mathbf{z}) &= \mathcal{N}(\mathbf{x} | \mathbf{W}\mathbf{z} + \boldsymbol{\mu}, \sigma^2 \mathbf{I}) & p(\mathbf{z}) &= \mathcal{N}(\mathbf{z} | \mathbf{0}, \mathbf{I}) \\ p(\mathbf{z} | \mathbf{x}) &= \mathcal{N}(\mathbf{z} | \mathbf{M}^{-1} \mathbf{W}^T (\mathbf{x} - \boldsymbol{\mu}), \sigma^2 \mathbf{M}^{-1}) & \boldsymbol{\mu} &= \bar{\mathbf{x}} \end{aligned}$$

$$\begin{aligned} \mathbb{E}[\ln p(\mathbf{X}, \mathbf{Z} | \boldsymbol{\mu}, \mathbf{W}, \sigma^2)] &= - \sum_{n=1}^N \left\{ \frac{D}{2} \ln(2\pi\sigma^2) + \frac{1}{2} \text{Tr}(\mathbb{E}[\mathbf{z}_n \mathbf{z}_n^T]) \right. \\ &\quad + \frac{1}{2\sigma^2} \|\mathbf{x}_n - \boldsymbol{\mu}\|^2 - \frac{1}{\sigma^2} \mathbb{E}[\mathbf{z}_n]^T \mathbf{W}^T (\mathbf{x}_n - \boldsymbol{\mu}) \\ &\quad \left. + \frac{1}{2\sigma^2} \text{Tr}(\mathbb{E}[\mathbf{z}_n \mathbf{z}_n^T] \mathbf{W}^T \mathbf{W}) + \frac{M}{2} \ln(2\pi) \right\} \end{aligned}$$

Eステップ E step

$$\begin{aligned} \mathbb{E}[\mathbf{z}_n] &= \mathbf{M}^{-1} \mathbf{W}^T (\mathbf{x}_n - \bar{\mathbf{x}}) \\ \mathbb{E}[\mathbf{z}_n \mathbf{z}_n^T] &= \sigma^2 \mathbf{M}^{-1} + \mathbb{E}[\mathbf{z}_n] \mathbb{E}[\mathbf{z}_n]^T \end{aligned}$$

エビデンスの下限 Evidence lower bound (ELBO)

EM Expectation maximization

確率的主成分分析
Probabilistic PCA

期待完全データ対数尤度関数 Expected complete-data log likelihood function

$$\begin{aligned}\mathbb{E}[\ln p(\mathbf{X}, \mathbf{Z} | \boldsymbol{\mu}, \mathbf{W}, \sigma^2)] = & - \sum_{n=1}^N \left\{ \frac{D}{2} \ln(2\pi\sigma^2) + \frac{1}{2} \text{Tr}(\mathbb{E}[\mathbf{z}_n \mathbf{z}_n^T]) \right. \\ & + \frac{1}{2\sigma^2} \|\mathbf{x}_n - \boldsymbol{\mu}\|^2 - \frac{1}{\sigma^2} \mathbb{E}[\mathbf{z}_n]^T \mathbf{W}^T (\mathbf{x}_n - \boldsymbol{\mu}) \\ & \left. + \frac{1}{2\sigma^2} \text{Tr}(\mathbb{E}[\mathbf{z}_n \mathbf{z}_n^T] \mathbf{W}^T \mathbf{W}) + \frac{M}{2} \ln(2\pi) \right\}\end{aligned}$$

Mステップ M step

$$\begin{aligned}\mathbf{W}_{\text{new}} &= \left[\sum_{n=1}^N (\mathbf{x}_n - \bar{\mathbf{x}}) \mathbb{E}[\mathbf{z}_n]^T \right] \left[\sum_{n=1}^N \mathbb{E}[\mathbf{z}_n \mathbf{z}_n^T] \right]^{-1} \\ \sigma_{\text{new}}^2 &= \frac{1}{ND} \sum_{n=1}^N \left\{ \|\mathbf{x}_n - \bar{\mathbf{x}}\|^2 - 2 \mathbb{E}[\mathbf{z}_n]^T \mathbf{W}_{\text{new}}^T (\mathbf{x}_n - \bar{\mathbf{x}}) + \text{Tr}(\mathbb{E}[\mathbf{z}_n \mathbf{z}_n^T] \mathbf{W}_{\text{new}}^T \mathbf{W}_{\text{new}}) \right\}\end{aligned}$$

エビデンスの下限 Evidence lower bound (ELBO)

EM Expectation maximization

計算量 The amount of computation

通常の主成分分析
Conventional PCA

固有ベクトル Eigenvectors
 $\mathcal{O}(D^3)$ or $\mathcal{O}(MD^2)$

共分散行列 Covariance matrix
 $\mathcal{O}(ND^2)$ or $\mathcal{O}(N^3)$

主成分分析に対するEMアルゴリズム
EM algorithm for PCA

$\mathcal{O}(NDM)$

エビデンスの下限 Evidence lower bound (ELBO)

主成分分析に対するEM EM for PCA

$$\begin{aligned} \mathbb{E}[\mathbf{z}_n] &= \mathbf{M}^{-1} \mathbf{W}_{\text{old}}^T (\mathbf{x}_n - \bar{\mathbf{x}}) \\ \mathbf{M} &= \mathbf{W}_{\text{old}}^T \mathbf{W}_{\text{old}} + \sigma^2 \mathbf{I} \end{aligned} \quad \xrightarrow{\sigma^2 \rightarrow 0} \quad \boldsymbol{\Omega} = (\mathbf{W}_{\text{old}}^T \mathbf{W}_{\text{old}})^{-1} \mathbf{W}_{\text{old}}^T \tilde{\mathbf{X}}^T$$

$\boldsymbol{\Omega}$: n 番目の列が $\mathbb{E}[\mathbf{z}_n]$.
 n th column is $\mathbb{E}[\mathbf{z}_n]$.

$\tilde{\mathbf{X}}$: n 番目の行が $\mathbf{x}_n - \bar{\mathbf{x}}$.
 n th row is $\mathbf{x}_n - \bar{\mathbf{x}}$.

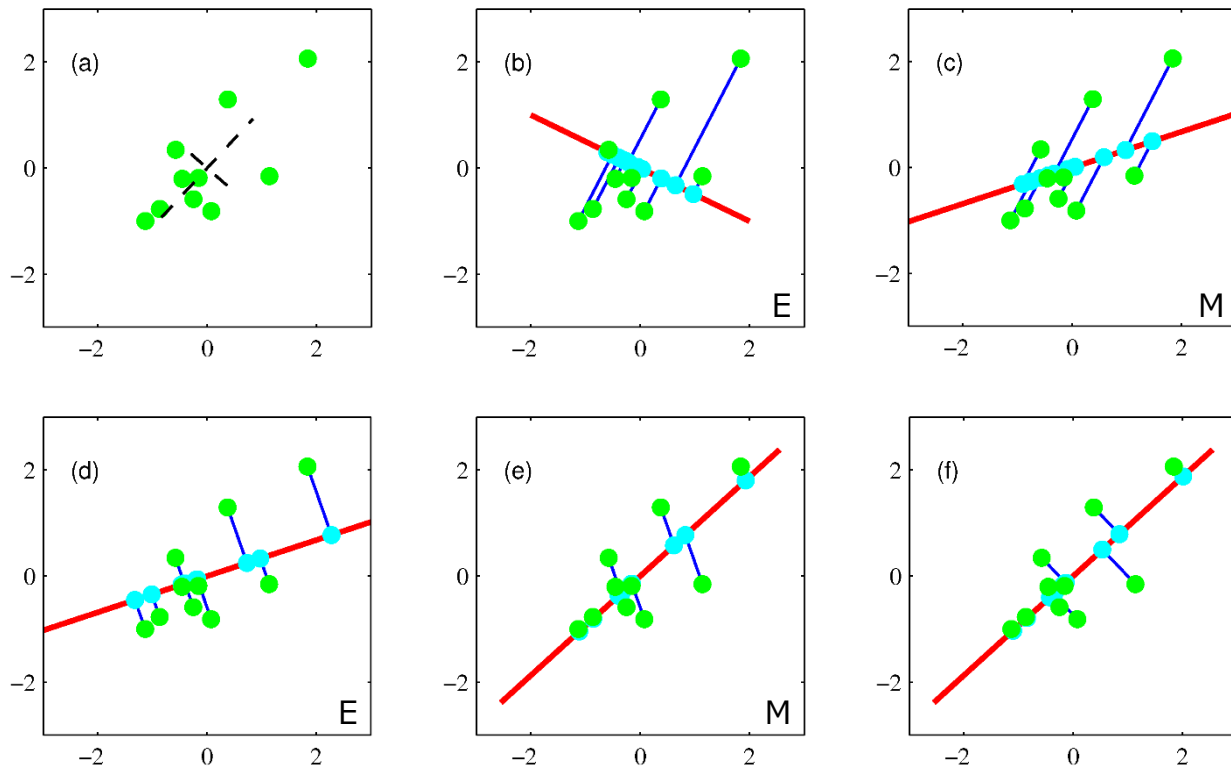
$$\mathbf{W}_{\text{new}} = \left[\sum_{n=1}^N (\mathbf{x}_n - \bar{\mathbf{x}}) \mathbb{E}[\mathbf{z}_n]^T \right] \left[\sum_{n=1}^N \mathbb{E}[\mathbf{z}_n \mathbf{z}_n^T] \right]^{-1} \xrightarrow{\sigma^2 \rightarrow 0} \mathbf{W}_{\text{new}} = \tilde{\mathbf{X}}^T \boldsymbol{\Omega}^T (\boldsymbol{\Omega} \boldsymbol{\Omega}^T)^{-1}$$

$$\mathbb{E}[\mathbf{z}_n \mathbf{z}_n^T] = \sigma^2 \mathbf{M}^{-1} + \mathbb{E}[\mathbf{z}_n] \mathbb{E}[\mathbf{z}_n]^T$$

エビデンスの下限 Evidence lower bound (ELBO)

主成分分析に対するEM EM for PCA

$$M = 1 \quad D = 2$$



エビデンスの下限 Evidence lower bound (ELBO)

主成分分析に対するEM EM for PCA

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

Eステップ E step

$$\begin{aligned}\mathbb{E}[\mathbf{z}_n] &= \mathbf{M}^{-1}\mathbf{W}^T(\mathbf{x}_n - \bar{\mathbf{x}}) & \mathbf{M} &= \mathbf{W}^T\mathbf{W} + \sigma^2\mathbf{I} \\ \mathbb{E}[\mathbf{z}_n\mathbf{z}_n^T] &= \sigma^2\mathbf{M}^{-1} + \mathbb{E}[\mathbf{z}_n]\mathbb{E}[\mathbf{z}_n]^T\end{aligned}$$

Mステップ M step

$$\begin{aligned}\mathbf{W}_{\text{new}} &= \left[\sum_{n=1}^N (\mathbf{x}_n - \bar{\mathbf{x}})\mathbb{E}[\mathbf{z}_n]^T \right] \left[\sum_{n=1}^N \mathbb{E}[\mathbf{z}_n\mathbf{z}_n^T] \right]^{-1} \\ \sigma_{\text{new}}^2 &= \frac{1}{ND} \sum_{n=1}^N \left\{ \|\mathbf{x}_n - \bar{\mathbf{x}}\|^2 - 2\mathbb{E}[\mathbf{z}_n]^T \mathbf{W}_{\text{new}}^T (\mathbf{x}_n - \bar{\mathbf{x}}) \right. \\ &\quad \left. + \text{Tr} \left(\mathbb{E}[\mathbf{z}_n\mathbf{z}_n^T] \mathbf{W}_{\text{new}}^T \mathbf{W}_{\text{new}} \right) \right\}\end{aligned}$$

エビデンスの下限 Evidence lower bound (ELBO)

因子分析に対するEM EM for factor analysis

$$p(\mathbf{x}|\mathbf{z}) = \mathcal{N}(\mathbf{x}|\mathbf{W}\mathbf{z} + \boldsymbol{\mu}, \boldsymbol{\Psi})$$

Eステップ E step

$$\begin{aligned}\mathbb{E}[\mathbf{z}_n] &= \mathbf{G}\mathbf{W}^T\boldsymbol{\Psi}^{-1}(\mathbf{x}_n - \bar{\mathbf{x}}) & \mathbf{G} &= (\mathbf{I} + \mathbf{W}^T\boldsymbol{\Psi}^{-1}\mathbf{W})^{-1} \\ \mathbb{E}[\mathbf{z}_n\mathbf{z}_n^T] &= \mathbf{G} + \mathbb{E}[\mathbf{z}_n]\mathbb{E}[\mathbf{z}_n]^T\end{aligned}$$

Mステップ M step

$$\begin{aligned}\mathbf{W}_{\text{new}} &= \left[\sum_{n=1}^N (\mathbf{x}_n - \bar{\mathbf{x}}) \mathbb{E}[\mathbf{z}_n]^T \right] \left[\sum_{n=1}^N \mathbb{E}[\mathbf{z}_n\mathbf{z}_n^T] \right]^{-1} \\ \boldsymbol{\Psi}_{\text{new}} &= \text{diag} \left\{ \mathbf{S} - \mathbf{W}_{\text{new}} \frac{1}{N} \sum_{n=1}^N \mathbb{E}[\mathbf{z}_n] (\mathbf{x}_n - \bar{\mathbf{x}})^T \right\}\end{aligned}$$



提出課題 VI : 連続潜在変数

Assignment VI: Continuous Latent Variables

提出期限 : **11月26日 (火曜日) 23:59:00** [日本標準時]

Submission deadline: **November 26 (Tuesday) 23:59:00** [Japan Standard Time]

提出課題は「一般」チャンネルの「ファイル」にアップロードされます。
同チャンネルに出現する通知のリンク先から解答を送信（提出）してください。
Assignments will be uploaded to "File" in the "General" channel. Send
(submit) your answers via the link that will appear in the same channel.

- 全6回の課題への解答をもとに成績評価を行います。
Your grades will be based on your answers to all six assignments.
- 解答送信後の解答再送信はできません。
Once submitted, answers cannot be resubmitted.
- 提出期限は一切延長しません。
The submission deadline will never be extended.