次元削減

福嶋竜希

Agenda

- 次元削減の概論 (60分)

  主成分分析(PCA)
  因子分析(FA)
  オートエンコーダー(AE)

- 実装編(60分)
  ハンズオンを通して理解を深め、
  モデルの比較を行いましょう！

Agenda

- 次元削減の概論 (60分)

  主成分分析(PCA)
  因子分析(FA)
  オートエンコーダー(AE)

- 実装編(60分)

  ハンズオンを通して理解を深め、
  モデルの比較を行いましょう！

PCA

PCAがやること

データを新たな軸(達)に写す。

- 2 次元データを 1 次元に写すことを考える。(次元圧縮のため)
- このとき、次元を減らすわけだから、情報の損失が起きる。

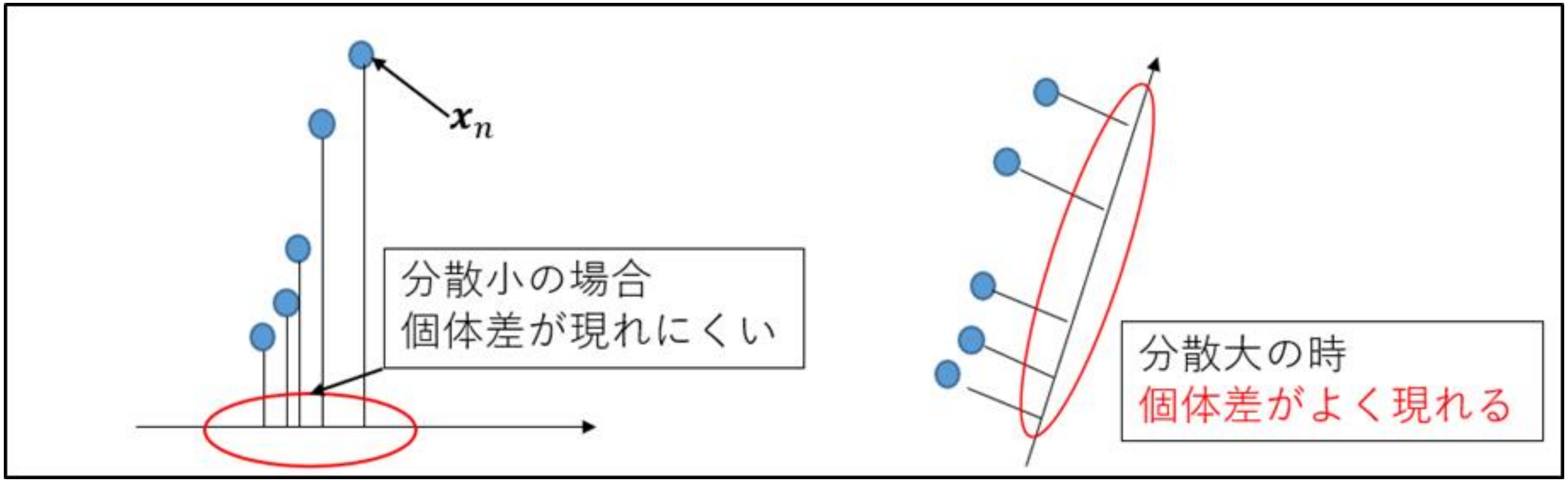
可能な限り、情報は保ちたい。

- 正規分布は多くのデータが従う分布であり、正規分布の情報量を表す数値に、エントロピーというものがある。
- このエントロピーを計算すると、分散の対数を用いて表せる。

情報を多く保つためには、写した後の分散を最大化するとよい。

PCA

- 直感的な解釈



- 分散を大きくすると、個体差がよく現れることがわかる。
- データを移す方向のことを主成分と呼ぶ。

- 例えば、100 次元あるデータを 1 次元で説明するには無理があるのでは…？

さらに主成分を追加していくとよい。

- ただし、既存のすべての主成分に直交するように新たな主成分を取り入れる。

PCA

考え方: 射影後の分散を大きくするように、射影方向 $\boldsymbol{u}_1$ をとる。

$\boldsymbol{x}$ が正規分布に従う時、$\boldsymbol{u}_1^T\boldsymbol{x}$ も正規分布に従い、情報量を表す微分エントロピーが次で与えられえる。

$$H[\boldsymbol{u}_1^T\boldsymbol{x}] = (\ln\sigma^2 + 1 + \ln(2\pi))/2 \quad (\sigma^2 \text{は分散パラメータ})$$

ゆえに、分散を最大化するような射影方向をとるべきである。

正規分布に従う確率変数 (ベクトル) の線型変換による像 (ここでは $\boldsymbol{u}_1^T\boldsymbol{x}$) も正規分布に従う。$\boldsymbol{u}_1^T\boldsymbol{x}$ の分散を $\boldsymbol{\sigma}^2$ と定義している。

連続確率変数 (ベクトル) $\boldsymbol{z}$ の微分エントロピーは、次で定義される。

$$H[\boldsymbol{z}] = -\int_{dom\, p(\boldsymbol{z})} p(\boldsymbol{z}) \ln p(\boldsymbol{z})\, d\boldsymbol{z}$$

$dom\, f(\cdot)$: 関数 f の定義域

以降、この考え方 (分散最大化) に則って、PCA を定式化していく。

PCA

$\overline{\boldsymbol{x}} = \sum_{n=1}^{N} \boldsymbol{x}_n$， $\mathrm{S} = \frac{1}{N}\sum_{n=1}^{N}(\boldsymbol{x}_n - \overline{\boldsymbol{x}})(\boldsymbol{x}_n - \overline{\boldsymbol{x}})^T$ とする。

射影後の分散は次のように導かれる。

$$\frac{1}{N}\sum_{n=1}^{N}(\boldsymbol{u}_1^T\boldsymbol{x}_n - \boldsymbol{u}_1^T\overline{\boldsymbol{x}})^2 = \boldsymbol{u}_1^T\boldsymbol{S}\boldsymbol{u}_1 \to max$$

$\|\boldsymbol{u}_1\|=\infty$ なる $\boldsymbol{u}_1$ に対して、 $\boldsymbol{u}_1^T\boldsymbol{S}\boldsymbol{u}_1$ は発散。
いわゆる、不良設定問題なので、制約 $\|\boldsymbol{u}_1\|^2=1$ を導入する。

ラグランジュの未定乗数法を用いて、次を最大化する。(λ_1 はラグランジュ乗数)

$$\boldsymbol{u}_1^T\boldsymbol{S}\boldsymbol{u}_1 + \lambda_1(1 - \|\boldsymbol{u}_1\|^2)$$

$\boldsymbol{u}_1$ で微分し、$\mathbf{0}$ とおくと、固有値問題に帰着する。

$$\boldsymbol{S}\boldsymbol{u}_1 = \lambda_1\boldsymbol{u}_1 \ \ (\Rightarrow \boldsymbol{u}_1^T\boldsymbol{S}\boldsymbol{u}_1 = \lambda_1)$$

ゆえに、最大固有値に対応する固有ベクトルが第 1 主成分である。
第 2 主成分は 2 番目に最も大きい固有値に対応する固有ベクトル等、
大きさが上位 M までの固有値に対応する固有ベクトルが主部分空間をなす。

<u>注意</u>　**S** は半正定値なので、固有値はすべて非負であり、不適な値は無い。
S は対称行列であることを考えると、主部分空間は正規直交系である。

PCA(まとめ)

考え方

与えられたデータの射影後の情報量を大きくするような射影方向を求める

- 射影方向を $\boldsymbol{u_1}$ とすると、射影後の情報量（エントロピー）は次のように書ける。

$$H[\boldsymbol{u}_1^T \boldsymbol{x}] = \frac{1}{2}(\ln \sigma^2 + 1 + \ln(2\pi)) \quad (\sigma^2 \text{は分散})$$

射影後の情報量の最大化は、射影後の分散 $\boldsymbol{\sigma}^2$ の最大化と同値である

- $\sigma^2 = \frac{1}{N}\sum_{n=1}^{N}(\boldsymbol{u}_1^T\boldsymbol{x}_n - \boldsymbol{u}_1^T\overline{\boldsymbol{x}})^2 = \boldsymbol{u}_1^T\boldsymbol{S}\boldsymbol{u}_1$ と書けて、これを最大化する。
- ラグランジュ乗数 λ_1 を導入し、ラグランジュ関数を $\boldsymbol{u_1}$ で微分すれば、

$$\boldsymbol{S}\boldsymbol{u}_1 = \lambda_1 \boldsymbol{u}_1 \ (\Rightarrow \boldsymbol{u}_1^T\boldsymbol{S}\boldsymbol{u}_1 = \lambda_1)$$

という固有値問題に帰着する。

この固有値問題を解いて、
第 i 番目に大きい固有値に対応する固有ベクトルが第 i 主成分

主成分はいくつ必要か

- 最大で、入力データの次元だけ主成分が得られる。
  ∴いくつかは、対応する固有値 (情報量) が小さいかもしれないので捨ててよい

主成分をいくつ採用するか

基準 1 : 累積寄与率が8割以上になるように採用
基準 2 : 固有値が 1 以上の主成分のみを採用 (必ず、データの標準化を行う)

<u>基準 1 を採用した実装を行う</u>

累積寄与率

PCA の対角化の結果、固有値 $\lambda_1, \lambda_2, \ldots, \lambda_D$ を得たとし、$\lambda_1 > \lambda_2 > \cdots > \lambda_D$ とする
λ_i の累積寄与率は、$(\lambda_1 + \lambda_2 + \cdots + \lambda_i)/\sum_j \lambda_j$ で与えられる。

$$\frac{\lambda_1 + \lambda_2 + \cdots + \lambda_5}{\sum_j \lambda_j} < 0.80, \qquad \frac{\lambda_1 + \lambda_2 + \cdots + \lambda_6}{\sum_j \lambda_j} > 0.80 \;\Rightarrow\; \text{主成分は 6 個}$$

非線形への拡張

- 先ほどの PCA では、非線形データは扱えない。

- 非線形データの次元削減方法としては、次の 3 つが挙げられる。

  (1) データそのものを工夫して、線形にする

  (2) カーネル PCA

  (3) オートエンコーダ

- 本セミナーでは、オートエンコーダを扱う。

- ちなみに、(1) は、実際にやるならば、途方もない作業になる。

- (2) はカーネル法が絡むので複雑になるが、洗練された理論である。

因子分析

- 潜在因子の存在を仮定する。z で潜在因子を表すものとすると、観測データ x との関係は次のようになる。

$$
\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_D \end{bmatrix} = \begin{bmatrix} w_{11}z_1 + w_{12}z_2 + \cdots + w_{1M}z_M \\ w_{21}z_1 + w_{22}z_2 + \cdots + w_{2M}z_M \\ \vdots \\ w_{D1}z_1 + w_{D2}z_2 + \cdots + w_{DM}z_M \end{bmatrix}
$$

$D > M$を仮定するので、次元削減になる。

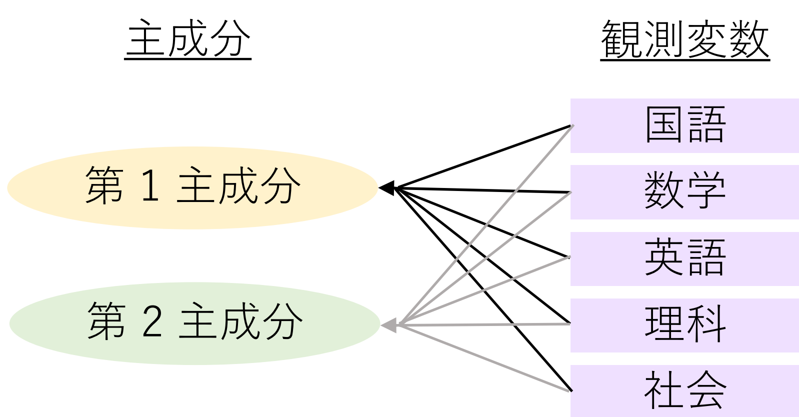
式から、潜在因子の加重和で観測データが表現されることがわかる。

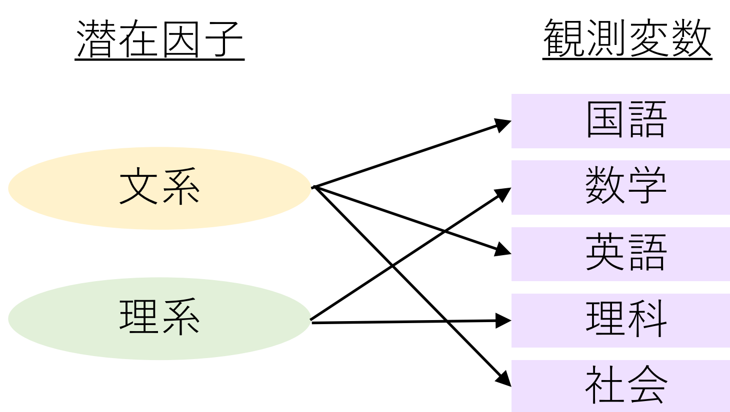
因子分析は、似た傾向を持つ変数をPCAよりわかりやすくまとめてくれる。
∴機械学習よりデータ分析に向く。

- EMアルゴリズムなどで w 達は推定されるため、
因子分析のアルゴリズムは非常に煩雑なため、省略する。

主成分分析と因子分析

主成分分析	因子分析
・情報をまとめることが目的 より少ない変数 (主成分) でデータを説明したい。 ・観測変数が原因で、 主成分とはそれから生じる結果 ・主成分が観測変数の線型結合	・潜在因子を見つけることが目的 ・潜在因子から、結果として観測値が生起すると考える ・観測変数が、潜在因子の線型結合



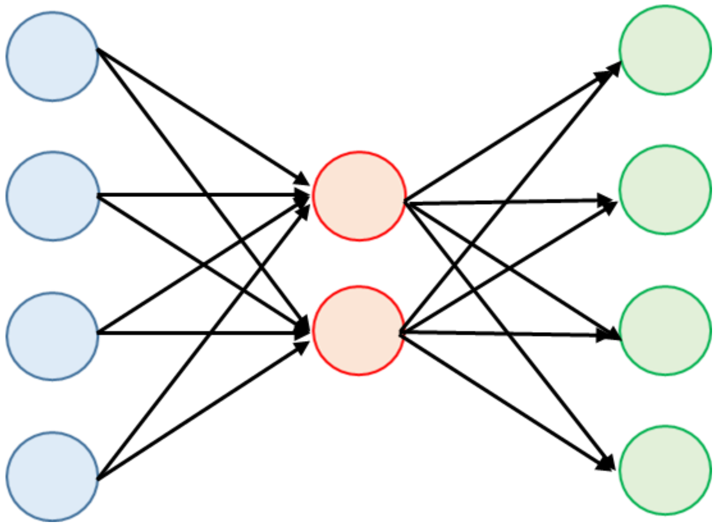


本章では、主成分、及び潜在因子を潜在変数として扱う。

ニューラルネットワーク

- ニューラルネットワークは、入力層、隠れ層、出力層と、各層をつなぐ重みと活性化関数が構成要素となる。

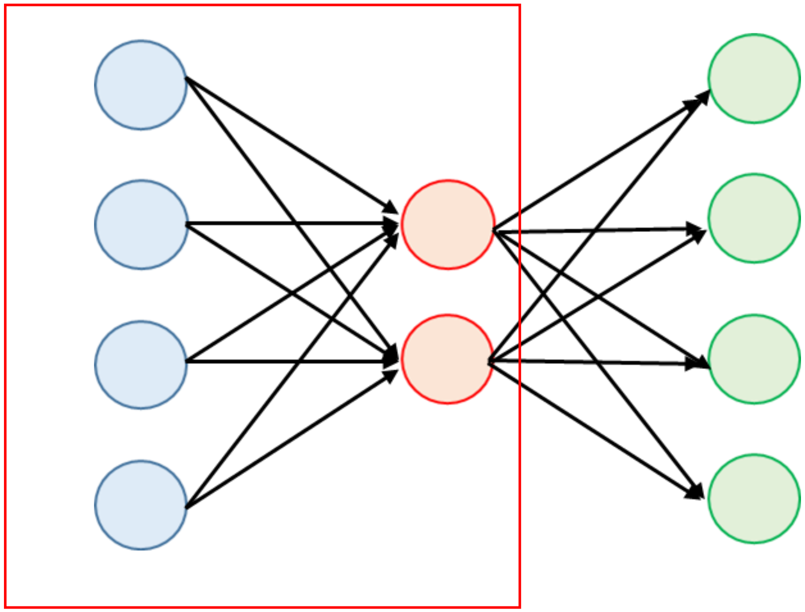
- ニューラルネットワークの構造の例を以下に示す。



ニューラルネットワーク

- ニューラルネットワークは、入力層、隠れ層、出力層と、各層をつなぐ重みと活性化関数が構成要素となる。

- ニューラルネットワークの構造の例を以下に示す。



- 図の赤枠内で見れば、次元圧縮が起きている。
- 出力に入力自身を与えて、ニューラルネットワーク回帰をすれば、次元圧縮の後、元の状態を復元するように訓練できるというのがオートエンコーダ。