本レジュメの巻末には、Appendix のページを付けている。必要に応じて参照する。

1　12 章まえがき

ここはワケわからなかった。

2　主成分分析

読み合わせ。

2.1　分散最大化による定式化

データ点 $\boldsymbol{x}_n$ を $\mathbb{R}^D \to \mathbb{R}^M$ $(M < D)$ として射影することを考える。簡単のため $M = 1$ の部分空間への射影について考える。この部分空間の方向を、D 次元空間の単位ベクトル $\boldsymbol{u}_1$ で表す。このベクトルを用いるとデータ点 $\boldsymbol{x}_n$ は $\boldsymbol{u}_1^T \boldsymbol{x}_n$ としてスカラー値に射影される。

他のデータ点についても同様の射影をおこなうことで、明らかに射影後のデータの平均値は $\boldsymbol{u}_1^T \bar{\boldsymbol{x}}$ となる。ここで $\bar{\boldsymbol{x}} = \frac{1}{N}\sum_{n=1}^{N} \boldsymbol{x}_n$ である。

ここで、射影されたデータの標本分散は次のようになる。

$$\begin{aligned}
\frac{1}{N}\sum_{n=1}^{N}(\boldsymbol{u}_1^T\boldsymbol{x}_n - \boldsymbol{u}_1^T\bar{\boldsymbol{x}})^2 &= \frac{1}{N}\sum_{n=1}^{N}(\boldsymbol{u}_1^T\boldsymbol{x}_n - \boldsymbol{u}_1^T\bar{\boldsymbol{x}})(\boldsymbol{u}_1^T\boldsymbol{x}_n - \boldsymbol{u}_1^T\bar{\boldsymbol{x}})^T \\
&= \frac{1}{N}\sum_{n=1}^{N}\boldsymbol{u}_1^T(\boldsymbol{x}_n - \bar{\boldsymbol{x}})(\boldsymbol{x}_n^T\boldsymbol{u}_1 - \bar{\boldsymbol{x}})^T\boldsymbol{u}_1) \\
&= \frac{1}{N}\sum_{n=1}^{N}\boldsymbol{u}_1^T(\boldsymbol{x}_n - \bar{\boldsymbol{x}})(\boldsymbol{x}_n - \bar{\boldsymbol{x}})^T\boldsymbol{u}_1 \\
&= \boldsymbol{u}_1^T\boldsymbol{S}\boldsymbol{u}_1
\end{aligned} \tag{1}$$

ただし $\boldsymbol{S} = \frac{1}{N}\sum_{n=1}^{N}(\boldsymbol{x}_n - \bar{\boldsymbol{x}})(\boldsymbol{x}_n - \bar{\boldsymbol{x}})^T$ として共分散行列を定義している。

この射影された分散を最大化することを考えよう。これはベクトル $\boldsymbol{u}_1$ が単位ベクトルであるという制約条件 $\boldsymbol{u}_1^T\boldsymbol{u}_1 = 1$ における停留値問題であるから、ラグランジュ未定乗数法を用いれば良い。いまは λ_1 をラグランジュ未定乗数として

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

の最大化を実行すればいい。L を $\boldsymbol{u}_1$ で偏微分すると

$$\frac{\partial L}{\partial \boldsymbol{u}_1} = \boldsymbol{S}\boldsymbol{u}_1 + \lambda_1\dot{(} - \boldsymbol{u}_1) \tag{3}$$

ここで $\frac{\partial L}{\partial \boldsymbol{u}_1} = 0$ とすれば次の方程式が導かれる。

$$\boldsymbol{S}\boldsymbol{u}_1 = \lambda_1\boldsymbol{u}_1 \tag{4}$$

これが L についての停留条件であり、$\boldsymbol{S}$ についての固有方程式である。なお、$\boldsymbol{u}_1$ は S の固有ベクトル、λ_1 はそのベクトルに対応する固有値である。

式 4 の両辺に左から $\boldsymbol{u}_1^T$ をかけると

$$\boldsymbol{u}_1^T \boldsymbol{S} \boldsymbol{u}_1 = \lambda_1 \tag{5}$$

となり、射影後の分散は $\boldsymbol{u}_1$ を最大固有値 λ_1 に属する固有ベクトルを選んだときに最大になることがわかる。この固有ベクトルは第一主成分と呼ばれる。

なお、その他の主成分を得るためには

・求めるベクトルに対して、既に得られている主成分（ベクトル）と直交するという条件を課す

・上記の直交条件のもとで、射影分散を最大化する

という手続きを踏めばいい。

一般の M 次元への射影についても次の事実を確認したい：射影されたデータの分散を最大化するような M 次元部分空間の上への線形写像が、データ共分散行列 $\boldsymbol{S}$ の上位 M 個の固有値に属する M 本の固有ベクトルにより定義される（空間が張られる）。の議論および結果については読み合わせとする[*1]。

以上をまとめると、主成分分析ではデータ集合の平均 $\bar{\boldsymbol{x}}$ と共分散行列 $\boldsymbol{S}$ が必要であり、さらに $\boldsymbol{S}$ の上位 M 個の固有値に対応する M 個の固有ベクトルを求める必要がある。

計算負荷のところはよくわからんので読み合わせ。

2.2 誤差最小化による定式化

ここでは先ほどの分散最大化とは異なるアプローチで定式化をおこなう。まず定式化のための準備として必要な量を定義し、問題の設定をおこなう。

いまデータ点 $\boldsymbol{x}_n$ を正規直交基底 $\boldsymbol{u}_i$ により

$$\boldsymbol{x}_n = \sum_{i=1}^{D} \alpha_{nj} \boldsymbol{u}_i \tag{6}$$

と書くことにする。なお $\{\boldsymbol{u}_i\}$ は D 次元のベクトルである。いま低次元の部分空間（次元数は M とする）に限られた数の $\boldsymbol{u}_i$ を用いてデータ点 $\boldsymbol{x}_n$ を近似し、射影をおこないたい。近似の際に、M 次元の線形部分空間は（元の空間の初めの）M 個の基底ベクトルを用いてやれば良い。各データ点 $\boldsymbol{x}_n$ は次のように近似することができる。

$$\tilde{\boldsymbol{x}}_{\boldsymbol{n}} = \sum_{i=1}^{M} z_{ni} \boldsymbol{u}_i + \sum_{i=M+1}^{D} b_i \boldsymbol{u}_i \tag{7}$$

以下において我々が解きたい問題は、この近似された $\tilde{\boldsymbol{x}}_{\boldsymbol{n}}$ と元のデータ点 $\boldsymbol{x}_n$ との距離について、全データで平均を取った量を最小化するというものになる。その量は次の式で表される。

$$J = \frac{1}{N} \sum_{n=1}^{N} \|\boldsymbol{x}_n - \tilde{\boldsymbol{x}}_{\boldsymbol{n}}\|^2 \tag{8}$$

これは $\boldsymbol{x}_n$ と $\tilde{\boldsymbol{x}}_{\boldsymbol{n}}$ との誤差と考えられ、この J を最小化することで $\{z_{ni}\}, \{b_i\}, \{\boldsymbol{u}_i\}$ を決定するというのが、誤差最小化による定式化である。最小化の手続きはシンプルであり、J を $\{z_{ni}\}, \{b_i\}, \{\boldsymbol{u}_i\}$ について微分し停留条件を考えるだけである。詳細な計算過程については App を参照してほしい。

[*1] 演習問題解答.pdf p.85-86 参照。数学的帰納法を用いて証明をおこなう。

J をあとで計算しやすくするために少々展開を施しておく。まずは（12.10）を（12.11）に代入してやる。

$$J = \frac{1}{N}\sum_{n=1}^{N}\|\boldsymbol{x}_n - (\sum_{i=1}^{M} z_{ni}\boldsymbol{u}_i + \sum_{i=M+1}^{D} b_i\boldsymbol{u}_i)\|^2 \tag{9}$$

次に式 9 を展開し、転置の記号を括弧内に入れ込む。

$$\begin{aligned} J &= \frac{1}{N}\sum_{n=1}^{N}(\boldsymbol{x}_n - \sum_{i=1}^{M} z_{ni}\boldsymbol{u}_i - \sum_{i=M+1}^{D} b_i\boldsymbol{u}_i)^T(\boldsymbol{x}_n - \sum_{j=1}^{M} z_{nj}\boldsymbol{u}_j - \sum_{j=M+1}^{D} b_j\boldsymbol{u}_j) \\ &= \frac{1}{N}\sum_{n=1}^{N}(\boldsymbol{x}_n^T - \sum_{i=1}^{M} z_{ni}\boldsymbol{u}_i^T - \sum_{i=M+1}^{D} b_i\boldsymbol{u}_i^T)(\boldsymbol{x}_n - \sum_{j=1}^{M} z_{nj}\boldsymbol{u}_j - \sum_{j=M+1}^{D} b_i\boldsymbol{u}_j) \end{aligned} \tag{10}$$

ここからは $\{z_{ni}\}, \{b_i\}$ による微分計算となる。App を参照してほしい（Go to App)。

最後に $\{\boldsymbol{u}_i\}$ で J を微分したいが、その前に J を $\{\boldsymbol{u}_i\}$ で表しておく必要がある。そのために式（12.12-13）を式（12.10）に代入し $\boldsymbol{x}_n - \tilde{\boldsymbol{x}}_{\boldsymbol{n}}$ を変形したい。

$$\begin{aligned} \boldsymbol{x}_n - \tilde{\boldsymbol{x}}_{\boldsymbol{n}} &= \boldsymbol{x}_n - (\sum_{i=1}^{M} \boldsymbol{x}_n^T\boldsymbol{u}_i\boldsymbol{u}_i + \sum_{i=M+1}^{D} \bar{\boldsymbol{x}}^T\boldsymbol{u}_i\boldsymbol{u}_i) \\ &= \boldsymbol{x}_n - (\sum_{i=1}^{D} \boldsymbol{x}_n^T\boldsymbol{u}_i\boldsymbol{u}_i - \sum_{i=M+1}^{D} \boldsymbol{x}_n^T\boldsymbol{u}_i\boldsymbol{u}_i + \sum_{i=M+1}^{D} \bar{\boldsymbol{x}}^T\boldsymbol{u}_i\boldsymbol{u}_i) \\ &= \boldsymbol{x}_n - \boldsymbol{x}_n + \sum_{i=M+1}^{D} \boldsymbol{x}_n^T\boldsymbol{u}_i\boldsymbol{u}_i - \sum_{i=M+1}^{D} \bar{\boldsymbol{x}}^T\boldsymbol{u}_i\boldsymbol{u}_i \ (\because \text{式 } 12.9) \\ &= \sum_{i=M+1}^{D} \boldsymbol{x}_n^T\boldsymbol{u}_i\boldsymbol{u}_i - \sum_{i=M+1}^{D} \bar{\boldsymbol{x}}^T\boldsymbol{u}_i\boldsymbol{u}_i \\ &= \sum_{i=M+1}^{D} \{(\boldsymbol{x}_n - \bar{\boldsymbol{x}})^T\boldsymbol{u}_i\}\boldsymbol{u}_i \end{aligned} \tag{11}$$

大変申し訳ないが、ここから最後まで読み合わせ。

App 誤差 J の微分計算

(12.11) で与えられる J について $\{z_{ni}\}$ と $\{b_i\}$ による微分を求め

(12.12-13) を導きたい。

$\{z_{ni}\}$ による微分

式(10) を展開し z に依らない項を落としていく.

$$J \sim \frac{1}{N}\sum_{n=1}^{N}\left(\boxed{-\sum_{i=1}^{M} z_{ni} u_i^T x_n - x_n^T \sum_{j=1}^{M} z_{nj} u_j + \sum_{i=1}^{M} z_{ni} u_i^T \sum_{j=1}^{M} z_{nj} u_j} \right.$$

$$\left. + \sum_{i=M+1}^{D} b_i u_i^T \sum_{j=1}^{M} z_{nj} u_j + \sum_{i=1}^{M} z_{ni} u_i^T \sum_{j=M+1}^{D} b_j u_j \right)$$

$$= \frac{1}{N}\sum_{n=1}^{N}\left(\boxed{\cdots} + \sum_{i=M+1}^{D}\sum_{j=1}^{M} b_i z_{nj} \delta_{ij} + \sum_{i=1}^{M}\sum_{j=M+1}^{D} z_{ni} b_j \delta_{ij} \right) \quad (\because (12.7))$$

常に $i \neq j$ なので $\delta_{ij} = 0$

$$= \frac{1}{N}\sum_{n=1}^{N}\left(\boxed{-\sum_{i=1}^{M} z_{ni} u_i^T x_n - x_n^T \sum_{j=1}^{M} z_{nj} u_j + \sum_{i=1}^{M}\sum_{j=1}^{M} z_{ni} z_{nj} \delta_{ij}} \right)$$

$$\sum_{j=1}^{M} z_{nj}^2$$

この J を z_{nj} で偏微分すると

$$\frac{\partial J}{\partial z_{nj}} \sim \frac{1}{N}\sum_{n=1}^{N}\left(-\sum_{i=1}^{M} \delta_{ij} u_i^T x_n - x_n^T u_j + 2 z_{nj} \right)$$

$$\sim u_j^T x_n - x_n^T u_j + 2 z_{nj}$$

$$= -2 x_n^T u_j + 2 z_{nj}$$

$\dfrac{\partial J}{\partial z_{nj}} = 0$ とすると $z_{nj} = x_n^T u_j$ 、式 (12.12) が導かれる.

$\{b_i\}$ による微分

式(10)を展開し b に依らない項を落としていく。

$$J \simeq \frac{1}{N}\sum_{n=1}^{N}\Bigg(-x_n^T\sum_{j=M+1}^{D} b_j u_j + \sum_{i=1}^{M} z_{ni} u_i^T \sum_{j=M+1}^{D} b_j u_j$$

$$-\sum_{i=M+1}^{D} b_i u_i^T x_n + \sum_{i=M+1}^{D} b_i u_i^T \sum_{j=1}^{M} z_{nj} u_j + \sum_{i=M+1}^{D} b_i u_i^T \sum_{j=M+1}^{D} b_j u_j\Bigg)$$

$$= \frac{1}{N}\sum_{n=1}^{N}\Bigg(\boxed{\cdots} + \sum_{i=1}^{M}\sum_{j=M+1}^{D} z_{ni} b_j \delta_{ij} + \sum_{i=M+1}^{D}\sum_{j=1}^{M} b_i z_{nj} \delta_{ij}\Bigg)$$

常に $i \neq j$ なので $\delta_{ij} = 0$

$$= \frac{1}{N}\sum_{n=1}^{N}\Bigg(-x_n^T\sum_{j=M+1}^{D} b_j u_j - \sum_{i=M+1}^{D} b_i u_i^T x_n + \underbrace{\sum_{i=M+1}^{D}\sum_{j=M+1}^{D} b_i b_j \delta_{ij}}_{\to \sum_{j=M+1}^{D} b_j^2}\Bigg)$$

この J を b_j で偏微分すると

$$\frac{\partial J}{\partial b_j} \simeq \frac{1}{N}\sum_{n=1}^{N}\Bigg(-x_n^T u_j - \sum_{i=M+1}^{D} \delta_{ij} u_i^T x_n + 2b_j\Bigg)$$

$$= \frac{1}{N}\sum_{n=1}^{N}\left(-x_n^T u_j - u_j^T x_n + 2b_j\right)$$

$$= -2\bar{x}^T u_j + 2b_j$$

$\frac{\partial J}{\partial b_j} = 0$ とすると $b_j = \bar{x}^T u_j$、式(12.13)が導かれる。

App 誤差Jの微分計算

(12.11)で与えられるJについて$\{z_{ni}\}$と$\{b_i\}$による微分を求め

(12.12-13)を導きたい。

$\{z_{ni}\}$による微分

式(10)を展開しzに依らない項を落としていく。

$$J \sim \frac{1}{N}\sum_{n=1}^{N}\left(\boxed{-\sum_{i=1}^{M} z_{ni}u_i^T x_n - x_n^T\sum_{j=1}^{M} z_{nj}u_j + \sum_{i=1}^{M} z_{ni}u_i^T\sum_{j=1}^{M} z_{nj}u_j}\right.$$

$$\left. + \sum_{i=M+1}^{D} b_i u_i^T \sum_{j=1}^{M} z_{nj}u_j + \sum_{i=1}^{M} z_{ni}u_i^T \sum_{j=M+1}^{D} b_j u_j \right)$$

$$= \frac{1}{N}\sum_{n=1}^{N}\left(\boxed{\cdots} + \sum_{i=M+1}^{D}\sum_{j=1}^{M} b_i z_{nj}\delta_{ij} + \sum_{i=1}^{M}\sum_{j=M+1}^{D} z_{ni}b_j\delta_{ij}\right) \quad (\because (12.7))$$

常に$i \neq j$なので$\delta_{ij}=0$

$$= \frac{1}{N}\sum_{n=1}^{N}\left(\boxed{-\sum_{i=1}^{M} z_{ni}u_i^T x_n - x_n^T\sum_{j=1}^{M} z_{nj}u_j + \sum_{i=1}^{M}\sum_{j=1}^{M} z_{ni}z_{nj}\delta_{ij}}\right)$$

$$\sum_{j=1}^{M} z_{nj}^2$$

このJをz_{nj}で偏微分すると

$$\frac{\partial J}{\partial z_{nj}} \sim \frac{1}{N}\sum_{n=1}^{N}\left(-\sum_{i=1}^{M}\delta_{ij}u_i^T x_n - x_n^T u_j + 2z_{nj}\right)$$

$$\sim u_j^T x_n - x_n^T u_j + 2z_{nj}$$

$$= -2x_n^T u_j + 2z_{nj}$$

$\dfrac{\partial J}{\partial z_{nj}} = 0$ とすると $z_{nj} = x_n^T u_j$、式(12.12)が導かれる。