情報数理科学 VII　教師なし学習

松島　慎

2018 年 11 月 2 日

講義で扱う教師なし学習の問題設定

$(\mathbf{x}_i, \mathbf{z}_i)$ の組がいくつか生成されたが、$(i = 1, 2, \ldots, n)$、$\mathbf{z}_i$ は観測できない。どのように推定すればよいか？
-潜在変数モデル（$\mathbf{z}_i$ を潜在変数、$\mathbf{x}_i$ を顕在変数とした確率モデルを考える）
-情報圧縮（$\mathbf{z}_i$ は $\mathbf{x}_i$ の情報を圧縮した値と考える）

1　クラスタリング問題

1.1　ハードクラスタリング

$z_i \in \{1, .., K\}$ のとき、クラスタリングまたはハードクラスタリングという。ハードクラスタリングは教師変数のない多クラス分類問題と言い換えることもできる。ハードクラスタリングは $\mathbf{z}_i = (z_{i1}, z_{ik}, \ldots, z_{iK}) \in \left\{ \mathbf{z} \in \{0,1\}^K \middle| \|\mathbf{z}\|_1 = 1 \right\}$ であるということもできる (one-hot encoding, 1-of-K 表記)。

- K-平均法/K-means clustering
  - 予測
    各 $k \in [K]$ でクラスタ中心 $\mu_k \in \mathbb{R}^d$ を定めて、各 $\mathbf{x}_i \in \mathbb{R}^d$ について
    $$\hat{z_k} = \begin{cases} 1 & \text{if } k = \operatorname{argmin}_{k'} \|\mathbf{x} - \mu_{k'}\| \\ 0 & \text{otherwise} \end{cases}$$
    と定める。
  - 目的関数
    $\mu = (\mu_k)_k$, $Z = (z_{ik})_{ik}$ とし、
    $$\begin{aligned} J(\mu) &= \sum_i \min_k \|\mathbf{x}_i - \mu_k\|^2 \\ &= \min_Z \underbrace{\sum_i \sum_k z_{ik} \|\mathbf{x}_i - \mu_k\|^2}_{\mathcal{L}(Z,\mu)} \end{aligned}$$
    Z にかかわらず $J(\mu) \leq \mathcal{L}(Z, \mu)$ となる。
  - EM アルゴリズム
    1. μ を適当に定める。
    2. E ステップ: $\mathcal{L}(Z, \mu)$ の Z についての最小化
    $$\begin{aligned} Z &= \operatorname*{argmin}_Z \mathcal{L}(Z, \mu) \\ \Rightarrow z_{ik} &= \begin{cases} 1 & \text{if } k = \operatorname{argmin}_{k'} \|\mathbf{x}_i - \mu_{k'}\| \\ 0 & \text{otherwise} \end{cases} \end{aligned}$$
    3. M ステップ:$\mathcal{L}(Z, \mu)$ の μ についての最小化
    $$\begin{aligned} \mu &= \operatorname*{argmin}_\mu \mathcal{L}(Z, \mu) \\ \Rightarrow \mu_k &= \frac{\sum_i z_{ik} \mathbf{x}_i}{\sum_i z_{ik}} \end{aligned}$$
    4. 値が変わらなくなるまで 2. と 3. を繰り返す。

1.2　ソフトクラスタリング

$\mathbf{z}_i = (z_{i1}, z_{ik}, \ldots, z_{iK}) \in \left\{ \mathbf{z} \in [0,1]^K \middle| \|\mathbf{z}\|_1 = 1 \right\} =: \triangle_K$ のとき、ソフトクラスタリングという。z_{ik} はデータ i が k に属する割合と解釈される。

ソフト K-平均法

- 予測
$$\hat{z_k} = \frac{\exp(-\beta \|\mathbf{x} - \mu_k\|^2)}{\sum_{k'} \exp(-\beta \|\mathbf{x} - \mu_{k'}\|^2)}$$
- 目的関数
$$\begin{aligned} J(\mu) &= \sum_i -\log \left(\sum_{k'} \exp(-\beta \|\mathbf{x}_i - \mu_{k'}\|^2) \right) \\ &= \sum_i -\log \left(\frac{\sum_{k'} \exp\left(-\beta \|\mathbf{x}_i - \mu_{k'}\|^2\right)}{\exp\left(-\beta \|\mathbf{x}_i - \mu_k\|^2\right)} \right) \\ &\quad + \sum_i -\log \left(\exp\left(-\beta \|\mathbf{x}_i - \mu_k\|^2\right) \right) \end{aligned}$$

よって $\gamma = (\gamma_{ik})_{ik} \in \triangle_K^n$ として、

$$J(\mu) = \underbrace{\sum_i \sum_k \gamma_{ik} \log \left(\frac{\exp(-\beta \left\| \mathbf{x}_i - \mu_k \right\|^2}{\sum_{k'} \exp(-\beta \left\| \mathbf{x}_i - \mu_{k'} \right\|^2)} \right) - \gamma_{ik} \log \gamma_{ik}}_{-\mathcal{K}(\gamma,\mu)} + \underbrace{\sum_i \sum_k \beta \gamma_{ik} \left\| \mathbf{x}_i - \mu_k \right\|^2 + \gamma_{ik} \log \gamma_{ik}}_{\mathcal{L}(\gamma,\mu)}$$

ギブスの不等式
$(\gamma_k) \in \triangle_K$,$(\gamma'_k) \in \triangle_K$ に対し、

$$\sum_k \gamma_k \log \gamma'_k \leq \sum_k \gamma_k \log \gamma_k$$

(∵ イェンゼンの不等式より

$$\sum_k \gamma_k \log \left(\gamma'_k / \gamma_k \right) \leq \log \left(\sum_k \gamma_k \left(\gamma'_k / \gamma_k \right) \right) = 0$$

)

$\mathcal{K}(\gamma,\mu)$ は γ, μ にかかわらず常に非負。すなわち γ にかかわらず $J(\mu) \leq \mathcal{L}(\gamma,\mu)$ となる。

- EM アルゴリズム
  1. μ を適当に定める。
  2. E ステップ: $\mathcal{L}(\gamma,\mu)$ を γ について最小化 ($\mathcal{K}(\gamma,\mu)$ を γ について最小化 $\to \mathcal{K}(\gamma,\mu) = 0$)

$$\gamma = \operatorname*{argmin}_{\gamma} \mathcal{K}(\gamma,\mu)$$
$$\Rightarrow \gamma_{ik} = \frac{\exp(-\beta \left\| \mathbf{x}_i - \mu_k \right\|^2)}{\sum_{k'} \exp(-\beta \left\| \mathbf{x}_i - \mu_{k'} \right\|^2)}$$

  目的関数の値は変わらない

  3. M ステップ: $\mathcal{L}(\gamma,\mu)$ を μ について最小化

$$\mu = \operatorname*{argmin}_{\mu} \mathcal{L}(\gamma,\mu)$$
$$\Rightarrow \mu_k = \frac{\sum_i \gamma_{ik} \mathbf{x}_i}{\sum_i \gamma_{ik}}$$

  第一項、第二項ともに減少する。
  4. 値が変わらなくなるまで 2. と 3. を繰り返す。

2 次元削減問題

$\mathbf{z}_i \in \mathbb{R}^K (K < d)$ のとき、次元削減、次元圧縮、特徴抽出などという。ここでは主に線形変換を用いる方法を考える。

2.1 線形代数の復習

Definition 1 (半正定値行列)**.** 実対称行列 $A = (a_{jj'})_{jj'} \in \mathbb{R}^{n \times n}$ が半正定値行列であるとは任意のベクトル $\mathbf{x} = (x_j)_j \in \mathbb{R}^n$ に対して、$\mathbf{x}^\top A \mathbf{x} = \sum_j \sum_{j'} a_{jj'} x_j x_{j'} \geq 0$ が成立すること。

Definition 2 (固有値分解)**.** 半正定値行列 $A \in \mathbb{R}^{n \times n}$ に対して、直交行列 ($Q^\top Q = I_n$ を満たす n 次正方行列) Q と非負対角行列 D が存在して $AQ = QD$ を満たす。(Q, D) を A の固有値分解という

すなわち、($Q = \left[\mathbf{q}_1 \cdots \mathbf{q}_n \right]$ かつ $D = \operatorname{diag}(\lambda_j)$ と考えると) 互いに直交するベクトル $(\mathbf{q}_j)_j \in (\mathbb{R}^n)^n$ と非負の実数 $\lambda_1 \geq \lambda_2 \ldots \geq \lambda_n \geq 0$ が存在し、$1 \leq j \leq n$ で、

$$A \mathbf{q}_j = \lambda_j \mathbf{q}_j$$

となる。これらを固有値、固有ベクトルという。

Proposition 1 (レーリー商)**.** 半正定値行列 $A \in \mathbb{R}^{n \times n}$ に対して、$\lambda_1 \geq \lambda_2 \ldots \geq \lambda_n \geq 0$ をその固有値とする。$\mathcal{U}$ を各列が直交する n 行 K 列の行列の集合とする。この時、

$$\max_{U \in \mathcal{U}} \operatorname{tr} U^\top A U = \sum_{j=1}^{K} \lambda_j.$$

A の固有ベクトルを K 本並べた行列は上式の最大値を達成する。

Definition 3 (特異値分解)**.** 実行列 $A \in \mathbb{R}^{n \times m}$ に対して、直交行列 $U \in \mathbb{R}^{n \times n}$ と $V \in \mathbb{R}^{m \times m}$ と非対角成分が 0 で対角成分が非負である行列 $D \in \mathbb{R}^{n \times m}$ が存在し、$AV = UD$ となる。(U, V, D) を A の特異値分解という。

すなわち、($U = \left[\mathbf{u}_1 \cdots \mathbf{u}_K \right]$ かつ $V = \left[\mathbf{v}_1 \cdots \mathbf{v}_K \right]$ かつ $D = \operatorname{diag}((\sigma_j)_j)$ と考えると) 互いに直互いに直交するベクトル $(\mathbf{u}_j)_j \in (\mathbb{R}^n)^n$ と互いに直交するベクトル $(\mathbf{v}_j)_j \in (\mathbb{R}^m)^m$ と非負の実数 $\sigma_1 \geq \sigma_2 \ldots \geq \sigma_{\min\{n,m\}} \geq 0$ が存在し、$1 \leq j \leq \min\{n, m\}$ に対し、

$$A \mathbf{v}_j = \sigma_j \mathbf{u}_j,\ A^\top \mathbf{u}_j = \sigma_j \mathbf{v}_j$$

となる。これらを右特異ベクトル、左特異ベクトル、特異値という。

Proposition 2 (特異値分解と固有値分解の関係)**.** $X \in \mathbb{R}^{n\times m}$ の特異値分解が $(U, V, \mathrm{diag}\,((\sigma_j)_j))$ であり、$X^\top X$ の固有値分解が $(Q, \mathrm{diag}\,((\lambda_j)_j))$ であるとする。このとき $1 \le j \le m$ で

$$V = Q,\ \sqrt{\lambda_j} = \sigma_j.$$

Proposition 3 (フロベニウスノルム)**.** 実行列 $A = (a_{jj'})_{jj'} \in \mathbb{R}^{n\times m}$ のフロベニウスノルムを $\|A\|_F$ と書き、$\|A\|_F = \sqrt{\mathrm{tr}\,(A^\top A)} = \sqrt{\mathrm{tr}\,(AA^\top)} = \sqrt{\sum_j \sum_{j'} (a_{jj'})^2}$ で定義する。A の特異値分解が (U, V, D) のとき $\|A\|_F^2 = \|D\|_F^2$.

Definition 4 (直交射影)**.** $n > K$ について、$n \times K$ 行列 V の列が互いに直交している、すなわち、$V^\top V = I_K$ とする。この時 $\mathbf{x} \mapsto VV^\top \mathbf{x}$ という変換を直交射影という。

2.2 特異値分解による低ランク近似

自然言語処理では LSI (Latent Semantic Indexing) または LSA(Latent Semantic Analysis) と呼ばれる手法（と同じ)。

- データ：$X = \begin{bmatrix} \mathbf{x}_1^\top \\ \vdots \\ \mathbf{x}_n^\top \end{bmatrix} = (x_{ij})_{ij} \in \mathbb{R}^{n\times d}$
- 低ランク表現：$Z = \begin{bmatrix} \mathbf{z}_1^\top \\ \vdots \\ \mathbf{z}_n^\top \end{bmatrix} = (z_{ik})_{ik} \in \mathbb{R}^{n\times K}$
- 辞書行列・基底行列 $V = \begin{bmatrix} \mathbf{v}_1 \cdots \mathbf{v}_K \end{bmatrix} \in \mathbb{R}^{d\times K}$ を用いた $\mathbf{z}_i = (z_{ik})_k$ からの $\mathbf{x}_i$ の復元: $\hat{\mathbf{x}} = V\mathbf{z}_i = \sum_k z_{ik}\mathbf{u}_k$
- 目的関数

$$\begin{aligned} \left\|X - ZV^\top\right\|_F^2 &= \sum_i \|\mathbf{x}_i - V\mathbf{z}_i\|^2 \\ &= \sum_i \left\|\mathbf{x}_i - \sum_k z_{ik}\mathbf{v}_k\right\|^2 \end{aligned}$$

- 解は一意ではないが X の特異値分解を (U', V', D') とすれば、$Z^* = U'D'$, $V^* = V'$ は解の一つ。

Eckart-Young-Mirsky の定理
$X \in \mathbb{R}^{n\times m}$ の特異値分解を $(U, V, \mathrm{diag}((\sigma_j)_j))$ とする。階数 $k \le \min\{n, m\}$ の行列の集合 $\mathbb{S}_k$ と $\hat{X} = \sum_{j=1}^k \sigma_j \mathbf{u}_j \mathbf{v}_j^\top$ に対し、

$$\left\|X - \hat{X}\right\|_F^2 = \min_{X' \in \mathbb{S}_k} \|X - X'\|_F^2 = \sum_{j=k+1}^{\min\{n,m\}} \sigma_j$$

2.3 主成分分析

簡単のため $X^\top \mathbf{1} = \mathbf{0}$ を満たすように前処理されているとする。

- z の予測（V の列は互いに直交） $\mathbf{z}_i = V^\top \mathbf{x}_i$
- z からの予測 $\hat{\mathbf{x}} = V\mathbf{z}_i + \mathbf{b}$
- 目的関数：$\hat{\mathbf{x}}_i(V, \mathbf{b}) = VV^\top \mathbf{x}_i + \mathbf{b}$ とし、

$$\begin{aligned} J(V, \mathbf{b}) &= \sum_i \|\mathbf{x}_i - \hat{\mathbf{x}}_i(V, \mathbf{b})\|_2^2 \\ &= \sum_i \left\|\mathbf{x}_i - VV^\top \mathbf{x}_i - \mathbf{b}\right\|_2^2 \\ &= \sum_i \left\|X - XVV^\top - \mathbf{1}_n \mathbf{b}^\top\right\|_F^2 \end{aligned}$$

$$\mathrm{minimize}_{\mathbf{b}} \sum_i \left\|\mathbf{x}_i - VV^\top \mathbf{x}_i - \mathbf{b}\right\|^2$$

$$\Rightarrow \mathbf{b}^* = \frac{\sum \mathbf{x}_i - VV^\top \mathbf{x}_i}{n} (= \mathbf{0})$$

- 目的関数：$J(V) = \sum_i \left\|\mathbf{x}_i - VV^\top \mathbf{x}_i - \mathbf{0}\right\|^2 = \left\|X(I_n - VV^\top)\right\|_F^2$
- 解：$X^\top X$ の固有値分解を $\left(\begin{bmatrix} \mathbf{q}_1 \cdots \mathbf{q}_n \end{bmatrix}, D\right)$ として、

$$V^* = \begin{bmatrix} \mathbf{q}_1 \cdots \mathbf{q}_K \end{bmatrix}$$

- 特異値分解による低ランク近似の解との関係
  $X^\top X$ の固有値分解における Q と X の特異値分解における V は一致する
  前処理の前提があるかないかに注意

$$\begin{aligned} &\left\|X(I_n - VV^\top)\right\|_F^2 \\ &= \mathrm{tr}\left(X(I_n - VV^\top)(I_n - VV^\top)X^\top\right) \\ &= \mathrm{tr}\left(X^\top X\right) - \mathrm{tr}\left(V^\top X^\top XV\right) \end{aligned}$$

よって解の一つは $X^\top X$ の固有ベクトルを K 本並べた行列。

$X^\top \mathbf{1} = \mathbf{0}$ を満たさない場合

$$\mathbf{b}^* = \frac{\sum \mathbf{x}_i - VV^\top \mathbf{x}_i}{n} = n^{-1}\left(X - XVV^\top\right)^\top \mathbf{1}$$

$$\begin{aligned} J(V) &= \left\| X(I - VV^\top) - \mathbf{1}\left(n^{-1}\left(X - XVV^\top\right)^\top \mathbf{1}\right)^\top \right\|_F^2 \\ &= \left\|(I - n^{-1}\mathbf{1}\mathbf{1}^\top)X(I - VV^\top)\right\|_F^2 \\ &= \operatorname{tr}\left(X^\top(I - n^{-1}\mathbf{1}\mathbf{1}^\top)X\right) - \operatorname{tr}\left(V^\top \underbrace{X^\top(I - n^{-1}\mathbf{1}\mathbf{1}^\top)X}_{\sum_i \mathbf{x}_i\mathbf{x}_i^\top - n^{-1}(\sum \mathbf{x}_i)(\sum \mathbf{x}_i)^\top} V\right) \end{aligned}$$

解は共分散行列の固有値ベクトルを並べたもの

2.4 主成分分析の拡張

- 正則化つき主成分分析
  - 目的関数:

$$\begin{aligned} \left\|X - ZV^\top\right\|_F^2 + \lambda_1 \|Z\|_{1,1} + \frac{\lambda_2}{2}\|Z\|_F^2 &= \sum_i \|\mathbf{x}_i - V\mathbf{z}_i\|^2 + \lambda_1 \|\mathbf{z}_i\|_1 + \frac{\lambda_2}{2}\|\mathbf{z}_i\|_2^2 \\ &= \sum_i \left\|\mathbf{x}_i - \sum_k z_{ik}\mathbf{v}_k\right\|^2 + \sum_k \left(\lambda_1 |z_{ik}| + \frac{\lambda_2}{2}\|\mathbf{z}_i\|_2^2\right) \end{aligned}$$

  - 制約：
    $V^\top V = I$
- 非負値行列分解
  - 目的関数:

$$\begin{aligned} \left\|X - ZV^\top\right\|_F^2 &= \sum_i \|\mathbf{x}_i - V\mathbf{z}_i\|^2 \\ &= \sum_i \left\|\mathbf{x}_i - \sum_k z_{ik}\mathbf{v}_k\right\|^2 \end{aligned}$$

  - 制約:
    Z と V は非負行列