行列分解

宗政一舟

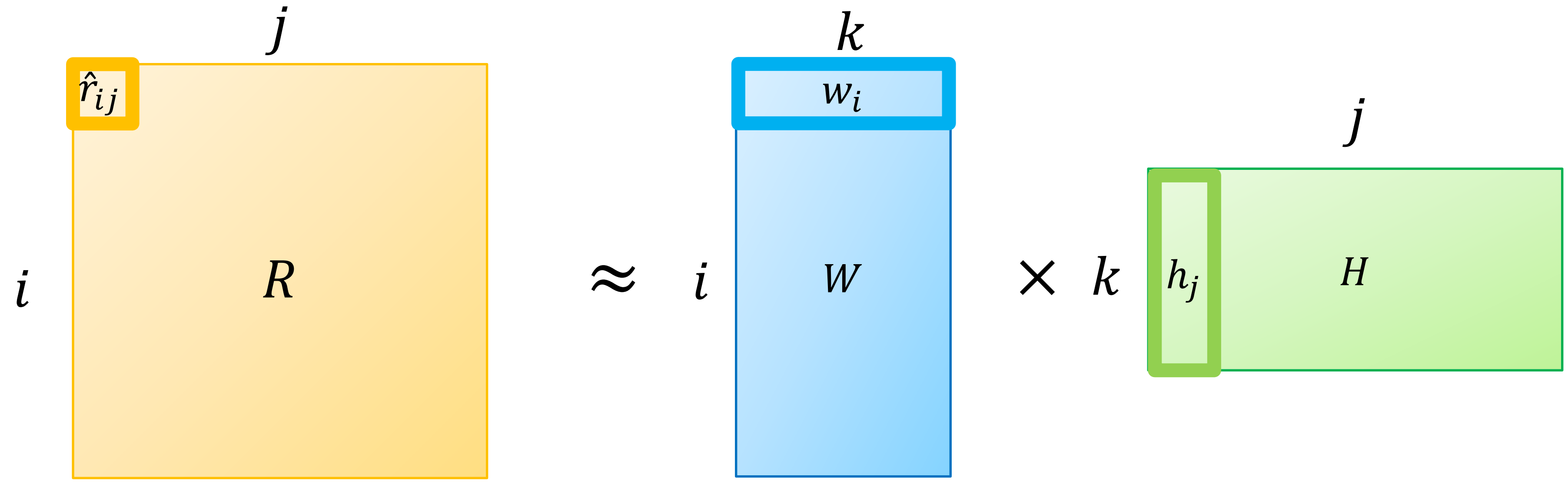
もくじ

1. 行列分解とは
2. 自然言語処理では
3. 行列分解の例
    1. 確率的勾配降下法(SGD)
    2. 非負値行列分解(NMF)
    3. 特異値分解(SVD)
4. 欠損値の扱い

1.行列分解とは

行列分解とは

- 高次元の行列がある時にデータスパースネスが発生してしまうことがある
- 行列Rを行列Wと行列Hに分解することでこの問題の解決する



$$\boxed{\hat{r}_{ij} \approx w_i h_j}$$

今回

- 紹介する行列分解
  1. 確率的勾配降下法(SGD)
  2. 非負値行列分解(NMF)
  3. 特異値分解(SVD)
- 細かい証明などはやらず、具体的にどのようにやるかが中心

2.自然言語処理では

自然言語処理では

- 例えば、ある商品レビューの[レビュー × 語彙] 行列を作成したい
- レビュー数が増えたら語彙数も爆発的に増える可能性がある → 行列分解

レビュー 1 : I have a very good time .
レビュー 2 : I do not like this goods.
レビュー 3 : I want to buy this again.

[レビュー × 語彙] 行列 作成

	I	have	a	very	good	time	do	not	like	this	goods	want	to	buy	again
レビュー 1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0
レビュー 2	1	0	0	0	0	0	1	1	1	1	1	0	0	0	0
レビュー 3	1	0	0	0	0	0	0	0	0	1	0	0	1	1	1

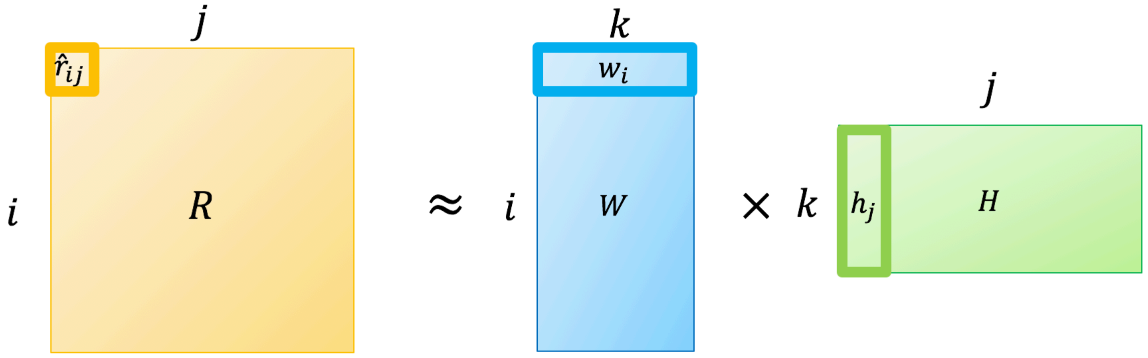
3. 行列分解の例

行列分解の例

- 確率的勾配降下法(SGD)
- 非負値行列分解(NMF)
- 特異値分解(SVD)

3.1 確率的勾配降下法(SGD)

SGDの更新式



$$\boxed{\hat{r}_{ij} \approx w_i h_j}$$

- 式(1)を$\boldsymbol{w}_i$と$\boldsymbol{h}_j$で偏微分をしてあげる

$$f(\boldsymbol{W},\boldsymbol{H}) = \frac{1}{2}\sum_{(i,j)\in D}\left(r_{ij} - \boldsymbol{w}_i\boldsymbol{h}_j\right)^2 + \frac{\lambda}{2}\left(\|\boldsymbol{W}\|_F^2 + \|\boldsymbol{V}\|_F^2\right) - (1)$$

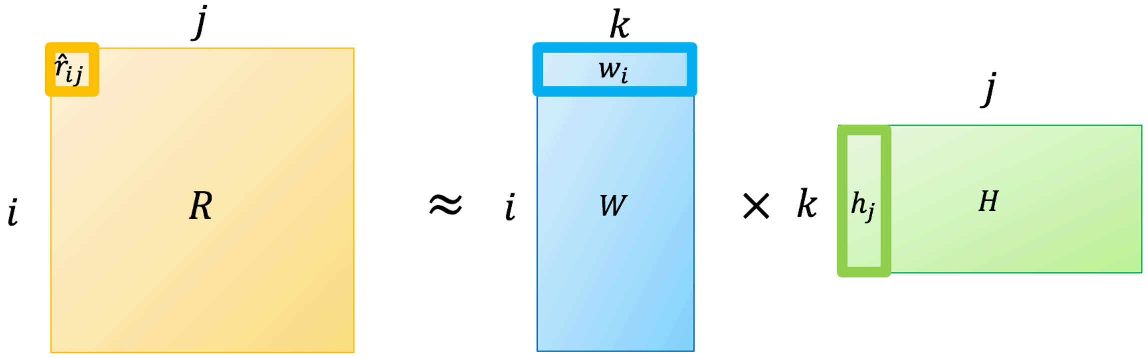
- 1度の更新で$\boldsymbol{w}_i$と$\boldsymbol{h}_j$から式(1)偏微分後を引いてあげる

$$\boldsymbol{w}_i \leftarrow \boldsymbol{w}_i - \eta\{\underbrace{-\left(r_{ij} - \boldsymbol{w}_i\boldsymbol{h}_j\right)\boldsymbol{h}_j + \lambda\boldsymbol{w}_i}_{}\} - (2) \qquad \boldsymbol{h}_j \leftarrow \boldsymbol{h}_j - \eta\{\underbrace{-\left(r_{ij} - \boldsymbol{w}_i\boldsymbol{h}_j\right)\boldsymbol{w}_i + \lambda\boldsymbol{h}_j}_{}\} - (3)$$

$\boldsymbol{w}_i$で微分した項　　　　h_iで微分した項

η:学習率, λ: 正則化の度合を決めるパラメータ

具体例



$\hat{r}_{ij} \approx w_i h_j$

$$\boldsymbol{w}_i \leftarrow \boldsymbol{w}_i - \eta\{-(r_{ij} - \boldsymbol{w}_i\boldsymbol{h}_j)\boldsymbol{h}_j^T + \lambda\boldsymbol{w}_i\} \quad -(2)$$

* $\boldsymbol{W}, \boldsymbol{H}$行列の初期値はランダム値をいれたりする

$$w_i \leftarrow w_i - \eta\{(\hat{r}_{ij} - w_i * h_j) * h_j^T + \lambda w_i\}$$

$$= w_i - \eta\{(\hat{r}_{ij} - w_i h_j) * h_j^T + \lambda w_i\}$$

$$= w_i - \eta\{\hat{r}_{xy} - w_i h_j * h_j^T + \lambda w_i\}$$

$$= w_i - \eta\, (\hat{r}_{ij} - w_i h_j) * h_j^T + \lambda w_i$$

3.2 非負値行列分解(NMF)

NMFの更新式

- 式(4)を最小にするような$\boldsymbol{W},\boldsymbol{H}$を求める
  - $\boldsymbol{W},\boldsymbol{H}$は非負値になるため制約つき
  - Dは距離関数でユークリッド距離、KLダイバージェンスなど(今回はユークリッド距離の更新式を紹介)
  - イェンシェンの不等式を用いて算出する

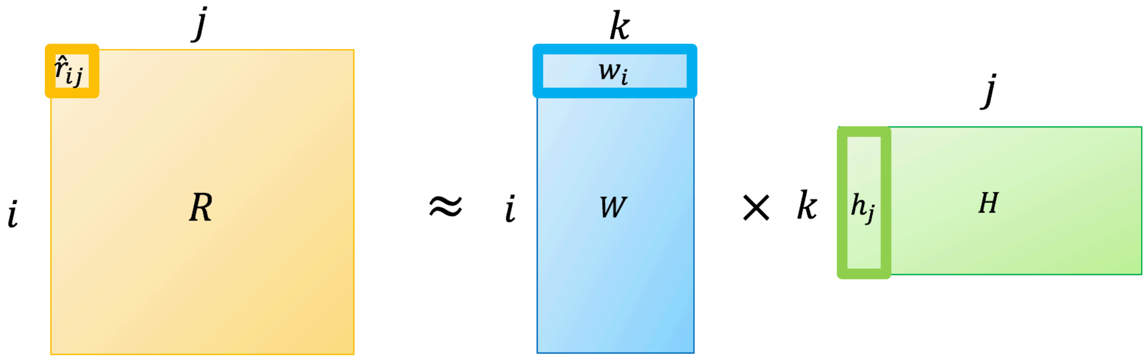
$$\boldsymbol{W},\boldsymbol{H} = \arg min_{\boldsymbol{W},\boldsymbol{H}}\ D(\boldsymbol{R}|\boldsymbol{W}\boldsymbol{H})\ ,s.t.\ \boldsymbol{W}_{i,k},\boldsymbol{H}_{kj} > 0\ -(4)$$

- 更新式

$$W_{i,k} \leftarrow W_{i,k}\frac{[RH^T]_{i,k}}{[WHH^T]_{i,k}} - (5) \qquad H_{k,j} \leftarrow H_{k,j}\frac{[W^TR]_{k,j}}{[W^TWH]_{k,j}} - (6)$$

* $W_{i,k}$のようにカンマで区切られている場合はi行k列の行列と解釈してください

具体例

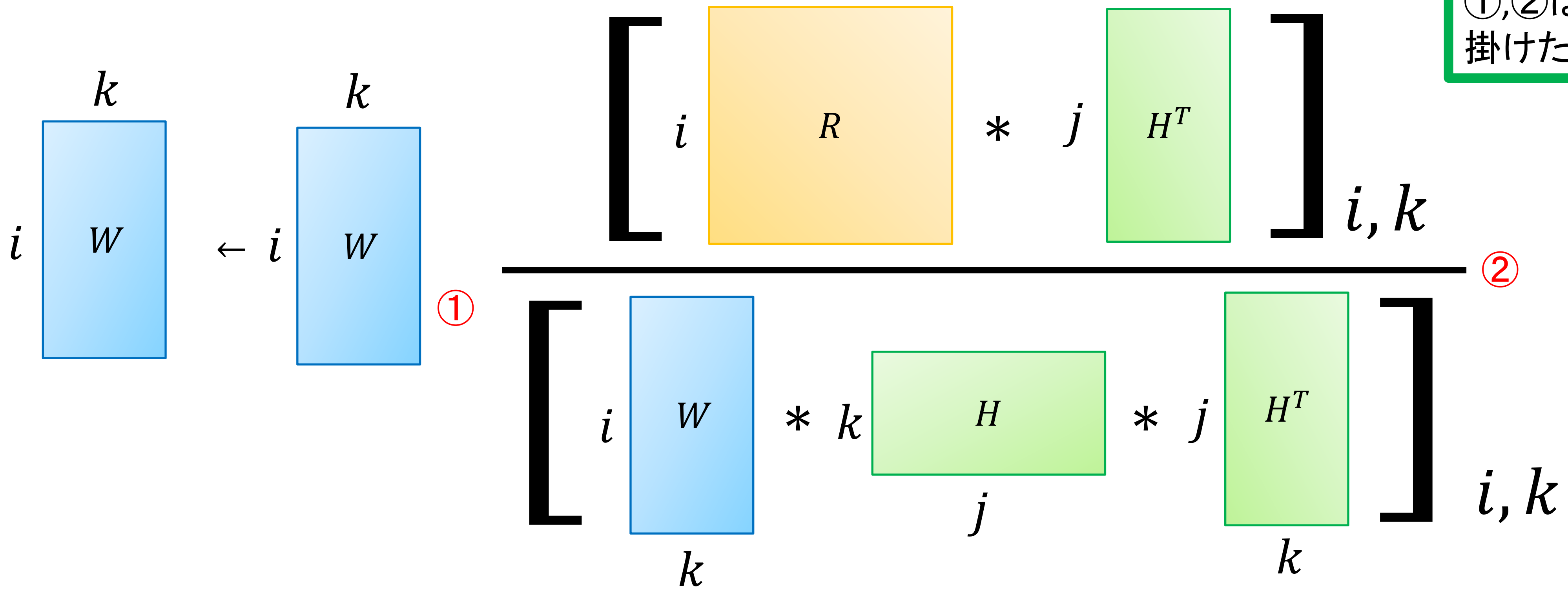


$\hat{r}_{ij} \approx w_i h_j$

$$W_{i,k} \leftarrow W_{i,k} \frac{[RH^T]_{i,k}}{[WHH^T]_{i,k}} - (5)$$

* $\boldsymbol{W}, \boldsymbol{H}$行列の初期値はランダム値にする

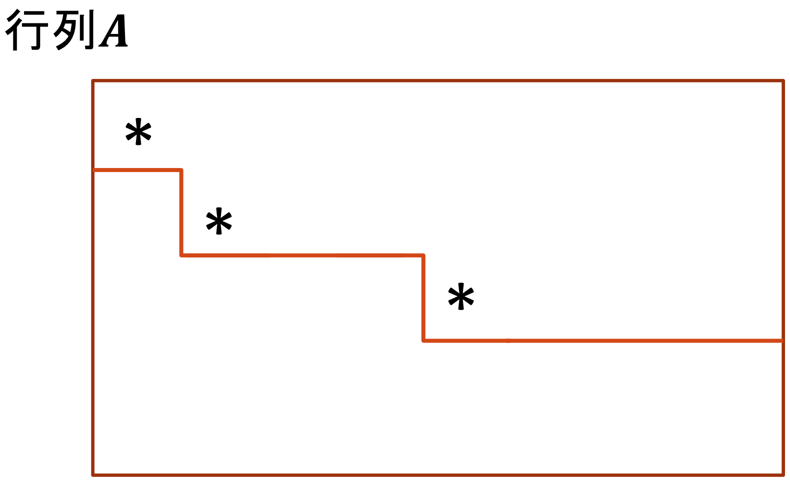
①,②はそれぞれの要素で掛けたり、割ったりする



3.3 特異値分解

ランク

- 任意の行列$\boldsymbol{A}$を階段形にしたときに0でない行が残る数



- *には0でない数が入っている。上の場合、$rank\boldsymbol{A} = 3$になります

固有値,固有ベクトル

- 正方行列$\boldsymbol{A}$に0でないベクトル$\boldsymbol{x}$,スカラーλの間に

$$\boldsymbol{A}\boldsymbol{x} = \lambda\boldsymbol{x}$$

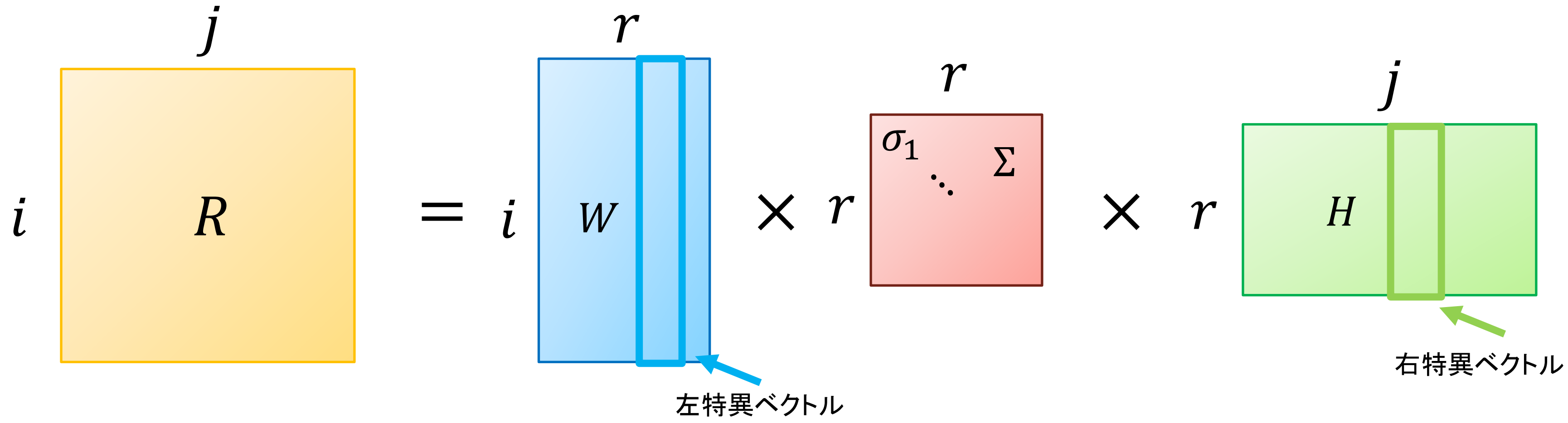
- 上式が成立するような$\boldsymbol{x}$を固有ベクトル,スカラーλを固有値と呼ぶ
- 固有値,固有ベクトルの計算

$$\lambda \text{が} \boldsymbol{A} \text{の固有値} \Leftrightarrow \det(\boldsymbol{A} - \lambda\boldsymbol{I}) = 0$$

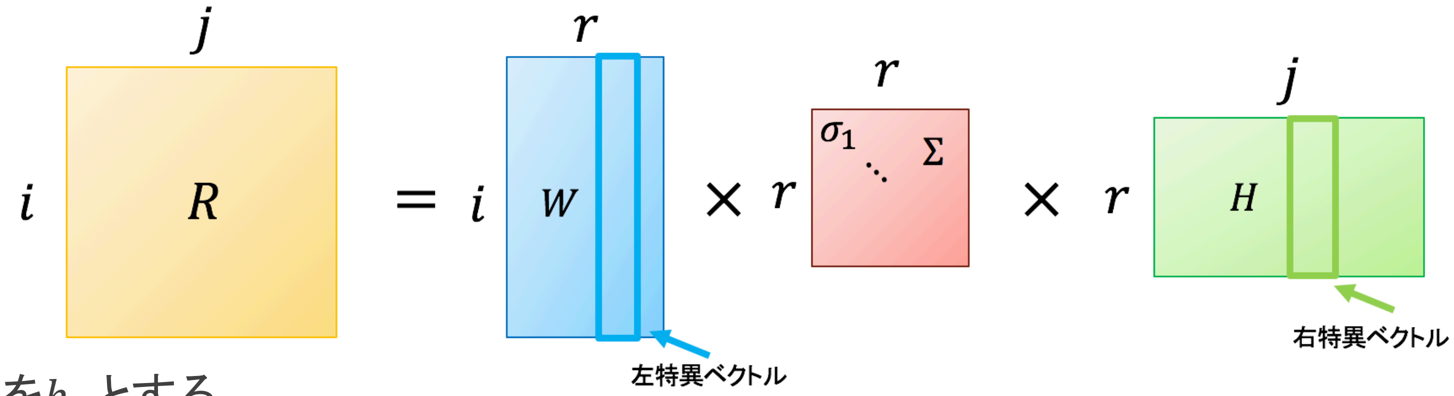
$\boldsymbol{I}$は単位行列

特異値分解

- 定義
  - 行列$\boldsymbol{R}_{i,j}$のランクをrとする
  - この時、$\boldsymbol{R}_{i,j} = \boldsymbol{W}_{i,r}\ \Sigma\ H_{r,j}$に分解することを特異値分解と呼ぶ
  - また、$\boldsymbol{W}^{\boldsymbol{T}}\boldsymbol{W} = \boldsymbol{H}^{\boldsymbol{T}}\boldsymbol{H} = \boldsymbol{I}$ になる。$\boldsymbol{I}$は単位行列
  - Σは対角行列で$\Sigma = diag(\sigma_1, \sigma_2, \dots \sigma_r)$である。$(\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r > 0)$
  - $\sigma_1, \sigma_2, \dots \sigma_r$を特異値といい、特異値$\sigma_k$の左特異ベクトルは$\boldsymbol{W}$の第$k$列ベクトル、右特異ベクトルは$\boldsymbol{H}$の第$k$列ベクトルである



特異値分解



- 特異値σ_kにおいて
  - 左特異ベクトルをw_k,右特異ベクトルをh_kとする
- $\boldsymbol{R}\boldsymbol{h}_k^T = \sigma_k \boldsymbol{w}_k, \boldsymbol{R}^T\boldsymbol{w}_k = \sigma_k \boldsymbol{h}_k^T$ ただし、$\boldsymbol{w}_k \neq \boldsymbol{0}, \boldsymbol{h}_k^T \neq \boldsymbol{0}$
- σ_k は $\boldsymbol{R}\boldsymbol{R}^T$または$\boldsymbol{R}^T\boldsymbol{R}$の固有値の正の平方根
- h_k は$\boldsymbol{R}^T\boldsymbol{R}$の固有ベクトル
- w_kは$\boldsymbol{R}\boldsymbol{R}^T$の固有ベクトル

具体例

- 特異値分解の具体例

行列$\mathbf{R} = \begin{pmatrix} 1 & 3 \\ 2 & 2 \\ 3 & 1 \end{pmatrix}$の特異値分解をしてみよう

- 特異値分解の流れ
  - 行列$\boldsymbol{R}$のランクを求める
  - $\boldsymbol{R}^T\boldsymbol{R}$の固有値の正の平方根を求める -> σ_k
  - $\boldsymbol{R}^T\boldsymbol{R}$の固有ベクトルを求める -> $\boldsymbol{h}_k$
  - $\boldsymbol{R}\boldsymbol{h}_k^T = \sigma_k \boldsymbol{w}_k$を変形$\frac{1}{\sigma_k}\boldsymbol{R}\boldsymbol{h}_k^T = \boldsymbol{w}_k$

- $Rh_k^T = \sigma_k w_k, R^T w_k = \sigma_k h_k^T$ ただし、$w_k \neq 0, h_k^T \neq 0$
- σ_k は RR^Tまたは$R^T R$の固有値の正の平方根
- h_k は$R^T R$の固有ベクトル
- w_kはRR^Tの固有ベクトル

具体例

- $Rh_k^T = \sigma_k w_k,\ R^T w_k = \sigma_k h_k^T$ ただし、$w_k \neq 0, h_k^T \neq 0$
- σ_k は RR^Tまたは$R^T R$の固有値の正の平方根
- h_k は$R^T R$の固有ベクトル
- w_kはRR^Tの固有ベクトル

- ランクを求める

$$\boldsymbol{R} = \begin{pmatrix} 1 & 3 \\ 2 & 2 \\ 3 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 3 \\ 0 & -4/3 \\ 0 & 0 \end{pmatrix} \text{なので} \ rank\boldsymbol{R} = 2$$

- $\boldsymbol{R}^T\boldsymbol{R}$の固有値,固有ベクトルを求める -> $\boldsymbol{h}_k$

$$\det(\boldsymbol{R}^T\boldsymbol{R} - \lambda\boldsymbol{I}) = 0$$

- 上式を計算すると$\lambda = 24,4$ となる。よって、特異値$\sigma_1 = 2\sqrt{6}, \sigma_2 = 2$
- また、固有ベクトル $\lambda = 24$の時 $(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}})$,$\lambda = 4$の時$(-\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}})$ それぞれ$\boldsymbol{h}_1, \boldsymbol{h}_2$に対応する

* 固有ベクトルの選び方が難しそう -> $\boldsymbol{W}^T\boldsymbol{W} = \boldsymbol{H}^T\boldsymbol{H} = \boldsymbol{I}$

具体例

- $Rh_k^T = \sigma_k w_k, R^T w_k = \sigma_k h_k^T$ ただし、$w_k \neq 0, h_k^T \neq 0$
- σ_k は RR^Tまたは$R^T R$の固有値の正の平方根
- h_k は$R^T R$の固有ベクトル
- w_kはRR^Tの固有ベクトル

- $\boldsymbol{Rh_k^T} = \sigma_k \boldsymbol{w}_k$を変形$\frac{1}{\sigma_k}\boldsymbol{Rh_k^T} = \boldsymbol{w_k}$
  - 上の式に今まで求めた値を代入すると求まる
  - $w_1 = (\frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}})$, $w_2 = (\frac{1}{\sqrt{2}}, 0, -\frac{1}{\sqrt{2}})$
- あとは,この形にすれば特異値分解終了$\boldsymbol{R}_{i,j} = \boldsymbol{W}_{i,r}\,\Sigma\,H_{r,j}$

$$R_{3,2} = (w_1, w_2) diag(\sigma_1, \sigma_2)(h_1, h_2)^T$$

$$= \begin{pmatrix} 1/\sqrt{3} & 1/\sqrt{2} \\ 1/\sqrt{3} & 0 \\ 1/\sqrt{3} & -1/\sqrt{2} \end{pmatrix} \begin{pmatrix} 2\sqrt{6} & 0 \\ 0 & 2 \end{pmatrix} \begin{pmatrix} 1/\sqrt{2} & 1/\sqrt{2} \\ -1/\sqrt{2} & 1/\sqrt{2} \end{pmatrix}$$

行列分解をする上での注意点

- 行列分解では行列をk次元に圧縮できる
  - SGDやNMFでは分解対象の行列の行数や列数よりも多く指定してはならない
  - SVDでは、ランク以上の行数や列数を指定することはできない(特異値分解は特異値の大きい値から使うらしい)
- 行列の要素が空の場合(欠損値)が発生する可能性
  - SGDは欠損値を無視して更新することは一応できる
  - NMFやSVDは欠損値を無視することはできない

4. 欠損値の扱い

欠損値の扱い

- 表のようなユーザ×アイテム行列
  - Eコマースを想定して、userに対してitemの評価値が入っているものとする
  - user1はitem4の評価値を付けていない
  - レビュー×語彙行列のように現れていないから「0」というようにするのは安易
- たとえば
  - (user1,item4)の要素にuser1の評価の平均値に入れてあげたりする(これも安易だが、、、)

	item1	item2	item3	item4	item5
user1	3	1	5	-	-
user2	4	-	1	2	1
user3	3	-	5	-	-

ご静聴ありがとうございました

参考文献

- [集合知プログラミング](オライリー本)
- [ビッグデータの計算科学 第3–4回 線形代数の基礎](http://www.iedu.i.kyoto-u.ac.jp/uploads/20141022.pdf#search=%27%E7%89%B9%E7%95%B0%E5%80%A4+%E6%B1%82%E3%82%81%E6%96%B9%27)
- [高校数学の美しい物語] [http://mathtrain.jp]