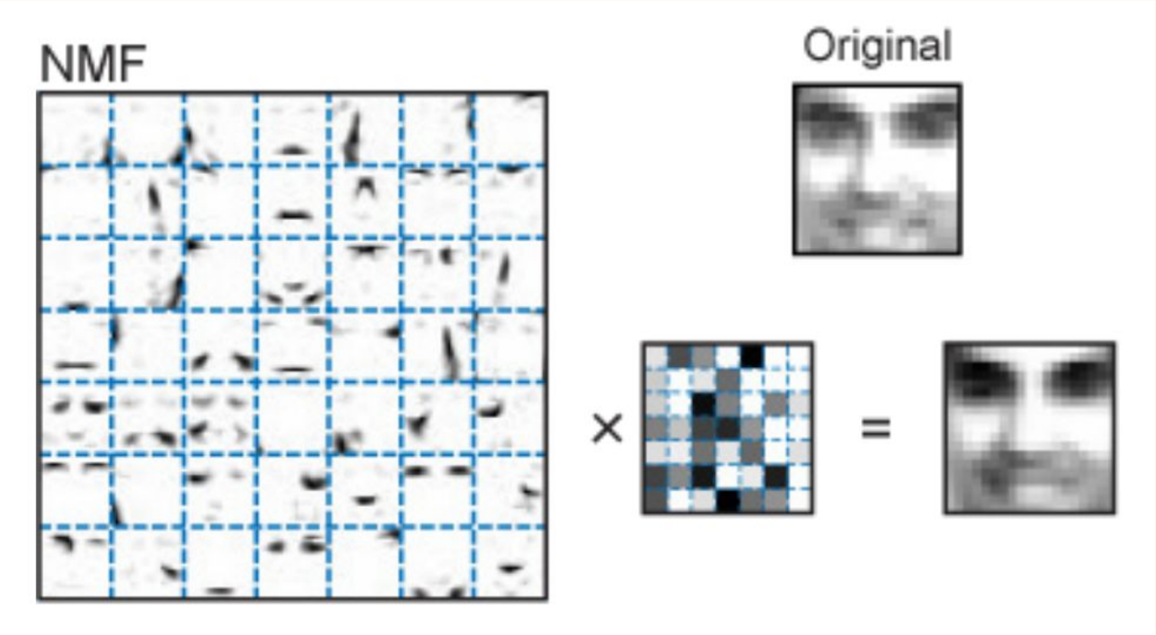
NMF

Non Negative Matrix Factorization

非負値行列因子分解

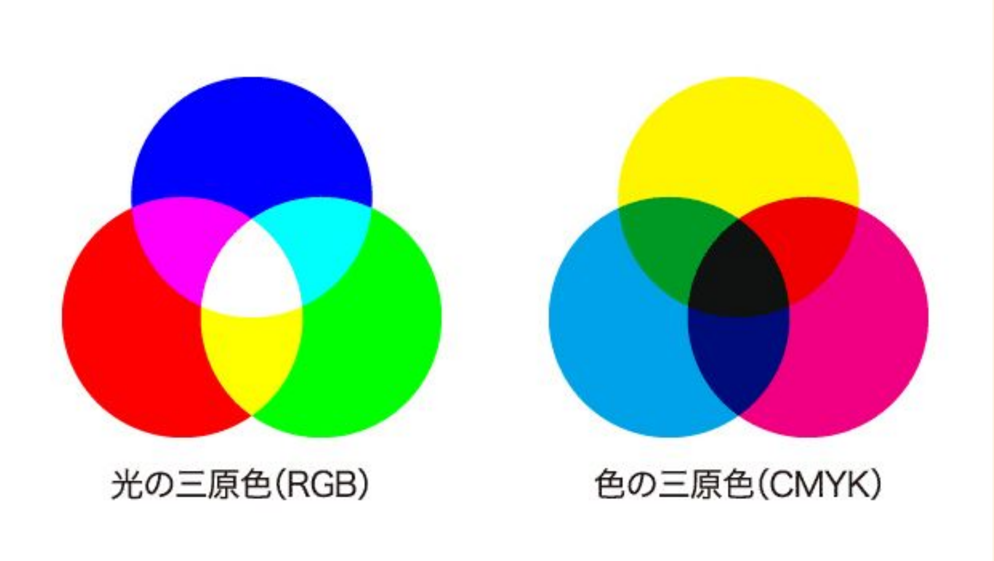
NMFの背景

画像処理や音声処理分野で活用
→画素値、スペクトル、頻度（単語の頻度）、個数（語彙数）etc
→実世界データは非負値であることが多い



観測データの性質
・観測データの非負性→RGB、単語頻度、音の周波数は全て非負
・基本となる特徴の非負性→波長が負の値である光は観測不可
・重ね合わせにおける非負性→観測データは特徴の重ね合わせで表現

eg)人間が知覚可能な色は光の三原色の
重ね合わせによって表現される



NMFの仮定

仮定：観測情報yは基底hの重みu付き和で表現可能（加法性）
　　→NMFの目的は基底Hと重みUを求めること

仮定：基底Hおよび重みUは非負

eg) 任意の色＝赤＊A＋青＊B＋緑＊C（A,B,Cは重み ）
※画素や周波数等は近似的に加法性が成立としてNMFを適用している

N個の観測ベクトル

$\boldsymbol{y}_1, \boldsymbol{y}_2, \cdots, \boldsymbol{y}_N$

基底がM個あると仮定し、基底ベクトル$\boldsymbol{h}_m$、結合係数$u_{m,n}$で表すと、

$$\boldsymbol{y}_n \simeq \sum_{m=1}^{M} \boldsymbol{h}_m u_{m,n} (n = 1, \cdots, N)$$

NMFの行列表現

観測ベクトルを並べたデータ行列を

$$\boldsymbol{Y} = [\boldsymbol{y}_1, \boldsymbol{y}_2, \cdots, \boldsymbol{y}_N] = (y_{k,n})_{K \times N}$$

基底ベクトルを並べて基底行列を

$$\boldsymbol{H} = [\boldsymbol{h}_1, \boldsymbol{h}_2, \cdots, \boldsymbol{h}_M] = (h_{k,m})_{K \times M}$$

結合係数を並べて係数行列を

$$\boldsymbol{U} = (u_{m,n})_{M \times N}$$

→観測データ行列は基底行列と結合係数行列の積で表現可能

$$\boldsymbol{Y} \simeq \boldsymbol{H}\boldsymbol{U}$$

$$(\mathbf{2} \times \mathbf{3}\text{行列}) \times (\mathbf{3} \times \mathbf{1}\text{行列}) = (\mathbf{2} \times \mathbf{1}\text{行列})$$

$$\begin{pmatrix} a1 & a2 & a3 \\ b1 & b2 & b3 \end{pmatrix} \times \begin{pmatrix} c \\ d \\ e \end{pmatrix} = \begin{pmatrix} a1 \cdot c & a2 \cdot d & a3 \cdot e \\ b1 \cdot c & b2 \cdot d & b3 \cdot e \end{pmatrix}$$

一般に

$(A \times B)$行列 $\times$ $(B \times C)$行列

$= (A \times C)$行列

NMFの原理図

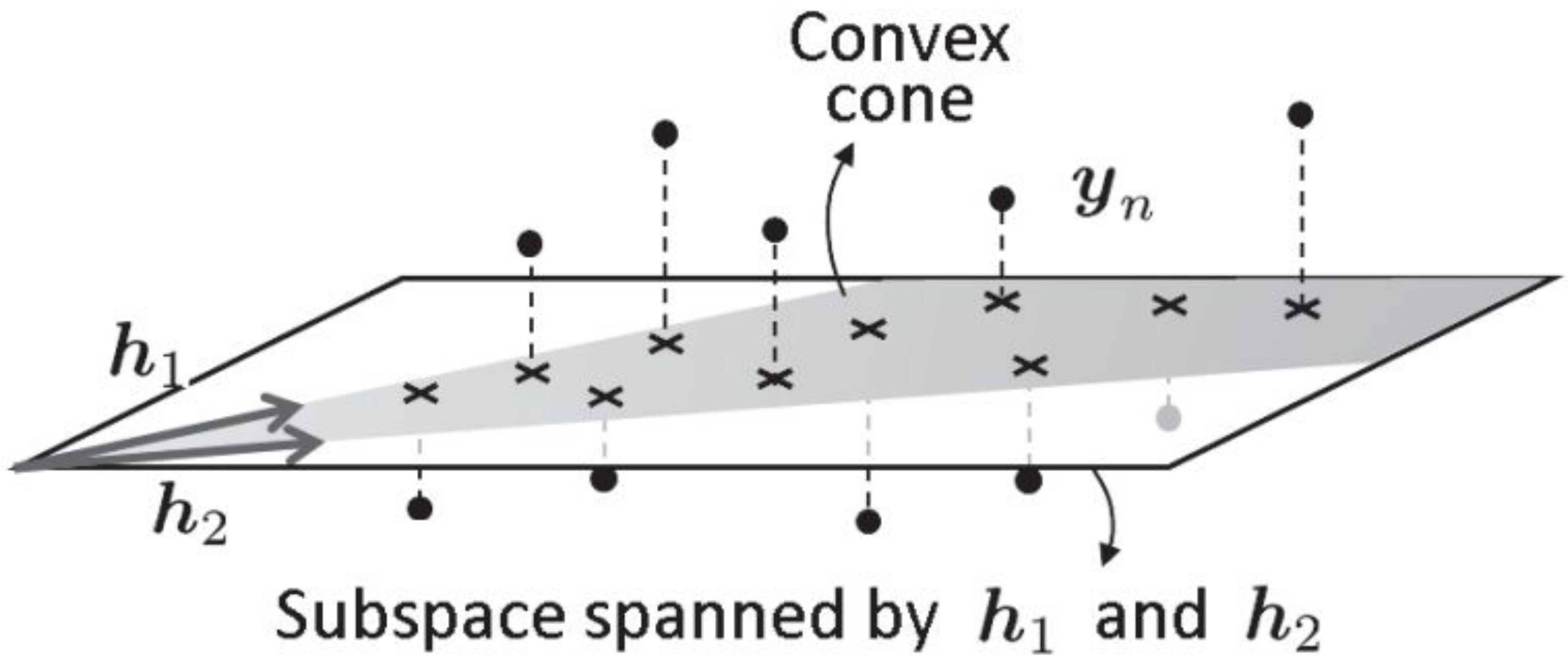


図 2　非負値基底の非負結合

任意の観測データが基底ベクトルh1、h2から構成される凸錐上に射影
→次元を落とした近似として解釈可能

$$y_n \simeq u_1 \boldsymbol{h}_1 + u_2 \boldsymbol{h}_2$$

閑話:凸錐

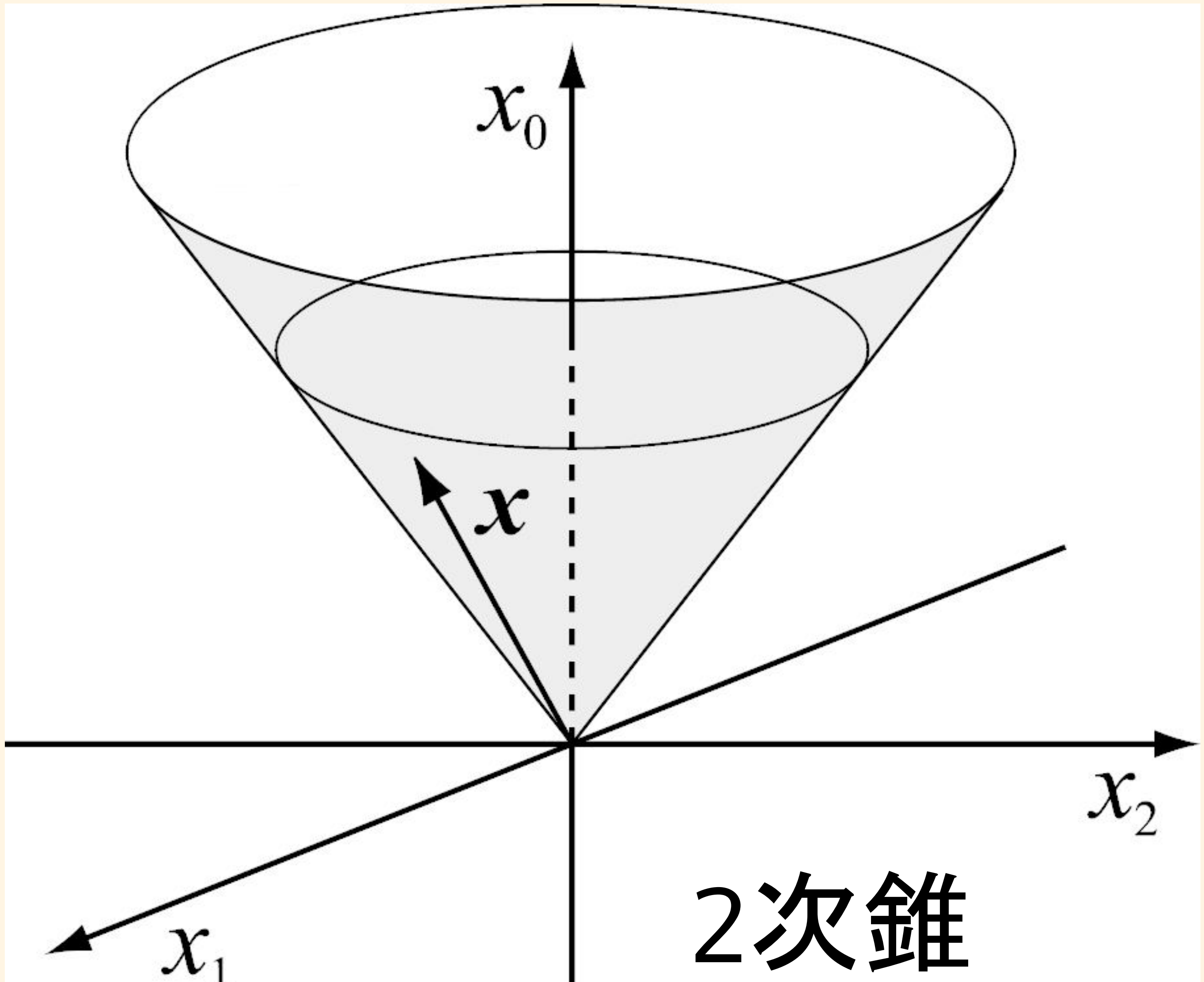
錐（$cone$）とは次のような集合を指す

$\boldsymbol{x} \in C, \quad \alpha \in [0, \infty) \rightarrow \alpha \boldsymbol{x} \in C$

錐が凸集合であるならば、それは凸錐である。

$eg)C = \{\boldsymbol{x} \in \boldsymbol{R}^n | \boldsymbol{x} = \sum_{i=1}^{m} \alpha_i \boldsymbol{\alpha}^i, \alpha \geq 0\}$

直交ベクトルによって形成される錐状の領域の例



NMFの性質

性質
・低次元近似
→基底ベクトルの次元Mは観測ベクトルの次元Kやデータ数Nに対して
　一般に小さく設定（M=Kの時H=E単位行列、M=Nの時U=E）

$$M < min(K, N)$$

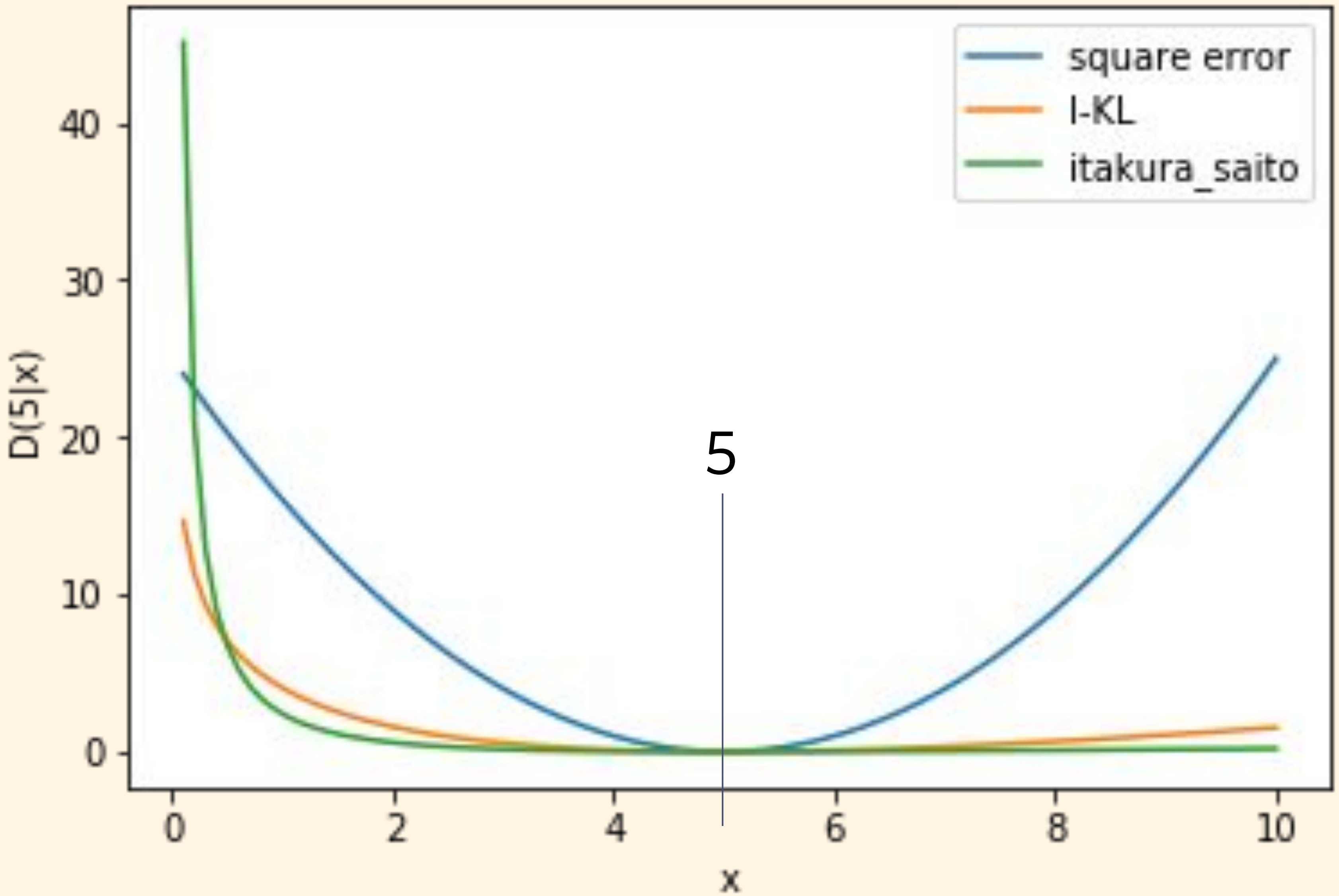
・観測データの共起成分をグルーピングする傾向（ここ謎）
→共起する成分をひとまとめにして基底と捉えるっぽい

・係数行列のスパース性
→あるベクトルの近似に有効でない基底の影響を除外したい
→有効でない基底の係数は0

・分解の一意性
→一般的に**H**、**U**は一意に定まらない（HSS^(-1)Uも解になりえる）
→最適化の際、次のような対策を取る
・スパースネス（行列の全要素に対する零要素の割合）の条件を付ける
・初期値をランダムに設定し、複数回アルゴリズム回す

NMFのアルゴリズム

基底行列Hと係数行列Uを求める際の方針
→データ行列Yと行列積HUの乖離度を最小化する最適化問題
→乖離度：二乗誤差、一般化KLダイバージェンス、板倉斎藤距離



二乗誤差

$$D(y|x) = (y - x)^2$$

一般化KLダイバージェンス

$$D(y|x) = y\ln\left(\frac{y}{x}\right) - (y - x)$$

板倉 − 斎藤距離

$$D(y|x) = \frac{y}{x} - \ln\left(\frac{y}{x}\right) - 1$$

NMFのアルゴリズム

基底行列Hと係数行列Uを求める際の方針
→データ行列Yと行列積HUの乖離度を最小化する最適化問題
→乖離度:二乗誤差、一般化KLダイバージェンス、板倉斎藤距離
→行列の要素計算:フロベニウスノルム

$$||A||_F = \sqrt{\sum_{i,j} a_{i,j}^2}$$

→フロベニウスノルムを使い、データ行列と行列積の差分を乖離度と仮定

$$\begin{aligned} D(\boldsymbol{Y}|\boldsymbol{HU}) &= ||\boldsymbol{Y} - \boldsymbol{HU}||_F^2 \\ &= \sum_{k,n} ||y_{k,n} - \sum_m h_{k,m} u_{m,n}||^2 \\ &= \sum_{k,n} \left(\left|y_{k,n}\right|^2 - 2y_{k,n} \sum_m h_{k,m} u_{m,n} - \left|\sum_m h_{k,m} u_{m,n}\right|^2 \right) \end{aligned}$$

閑話:イェンセンの不等式

$$
\begin{aligned}
D(\boldsymbol{Y}|\boldsymbol{HU}) &= ||\boldsymbol{Y} - \boldsymbol{HU}||_F^2 \\
&= \sum_{k,n} ||y_{k,n} - \sum_m h_{k,m} u_{m,n}||^2 \\
&= \sum_{k,n} \left(\left|y_{k,n}\right|^2 - 2y_{k,n} \sum_m h_{k,m} u_{m,n} - \left|\sum_m h_{k,m} u_{m,n}\right|^2 \right)
\end{aligned}
$$

和の絶対値を取り扱うことは難しい→イェンセンの不等式

$f(\cdot)$が凸関数の時、

$$\sum_{i=1}^n \lambda_i f(x_i) \geq f\left(\sum_{i=1}^n \lambda_i x_i\right) \qquad \begin{cases} \sum_i \lambda_i = 1 \\ \lambda_i \geq 0 \end{cases}$$

凸関数の性質

性質1：任意の$x1, x2, \lambda (0 \leq \lambda \leq 1)$に対して，$\lambda f(x1) + (1 - \lambda) f(x2) \geq f(\lambda x1 (1 - \lambda) x2)$を満たす。

性質2：任意の$x1, x2$に対して，2点$(x1, f(x1)), (x2, f(x2))$を結ぶ線分が関数の上側にある。

性質3：二階微分$f''(x)$が存在して0以上である

NMFのアルゴリズム

$$\left|\sum_m h_{k,m} u_{m,n}\right|^2 \leq \sum_i \frac{h_{k,m}^2 u_{m,n}^2}{\lambda_{k,m,n}}$$

$$\begin{aligned}
D(\boldsymbol{Y}|\boldsymbol{HU}) &= ||\boldsymbol{Y} - \boldsymbol{HU}||_F^2 \\
&= \sum_{k,n} ||y_{k,n} - \sum_m h_{k,m} u_{m,n}||^2 \\
&= \sum_{k,n} \left(||y_{k,n}||^2 - 2y_{k,n} \sum_m h_{k,m} u_{m,n} + \underline{||\sum_m h_{k,m} u_{m,n}||^2} \right) \\
&\leq \sum_{k,n} \left(||y_{k,n}||^2 - 2y_{k,n} \sum_m h_{k,m} u_{m,n} + \underline{\sum_m \frac{h_{k,m}^2 u_{m,n}^2}{\lambda_{k,m,n}}} \right)
\end{aligned}$$

変形手順

$$\left(\sum_i x_i\right)^2 = \left(\sum_i \lambda_i \frac{x_i}{\lambda_i}\right)^2 \leq \sum_i \lambda_i \left(\frac{x_i}{\lambda_i}\right)^2 = \sum_i \frac{x_i^2}{\lambda_i}$$

閑話：補助関数法

目的関数の上界を定める関数Gが求まった
→上界を定める関数Gを反復的に降下させることで解を求める
→反復降下の補助を行う関数（補助関数）として上界を定める関数Gを使う
→補助関数の反復降下による目的関数の降下方法を補助関数法

H と U の行列要素 hk,1,...,hk,M , u1,n,...,uM,n を含んだ非線形関数項であ ること に気付く. この項に対し, 行列要素 hk,m, um,n ご との関数の和に分離した形を した上限関数を設けたい. 2 次関数は凸関数であるため,

NMFのアルゴリズム

$$D(\boldsymbol{Y}|\boldsymbol{HU}) \leq \underbrace{\sum_{k,n}\left(||y_{k,n}||^2 - 2y_{k,n}\sum_{m} h_{k,m}u_{m,n} + \sum_{m}\frac{h_{k,m}^2 u_{m,n}^2}{\lambda_{k,m,n}}\right)}_{G}$$

→補助変数　$\lambda_{k,m,n}$（イェンセンの不等式における係数）

→係数の条件より

$$\lambda_{k,m,n} = \frac{h_{k,m}u_{m,n}}{\sum_l h_{k,l}u_{l,n}}$$

→補助関数**G**をH、Uに関して偏微分を行い停留点を探す

$$\frac{\partial G}{\partial h_{k,m}} = -2y_{k,n}\sum_m u_{m,n} + \sum_m 2\frac{h_{km}u_{m,n}^2}{\lambda_{k,m,n}} = 0$$

$$\frac{\partial G}{\partial u_{m,n}} = -2y_{k,n}\sum_m h_{k.m} + \sum_m 2\frac{h_{k,m}^2 u_{m,n}}{\lambda_{k.m.n}} = 0$$

NMFのアルゴリズム

$$\frac{\partial G}{\partial h_{k,m}} = -2y_{k,n}\sum_m u_{m,n} + \sum_m 2\frac{h_{km}u_{m,n}^2}{\lambda_{k,m,n}} = 0$$

$$\frac{\partial G}{\partial u_{m,n}} = -2y_{k,n}\sum_m h_{k.m} + \sum_m 2\frac{h_{k,m}^2 u_{m,n}}{\lambda_{k.m.n}} = 0$$

整理途中

参考文献

・非負値行列因子分解/亀岡弘和/計測と制御 第 51 巻 第 9 号 2012 年 9 月号

・チュートリアル: 非負値行列因子分解/亀岡弘和

・非負値行列因子分解/亀岡弘和/
http://www.kecl.ntt.co.jp/people/kameoka.hirokazu/publications/Kameoka2012SICE09published.pdf

・NMFアルゴリズムの導出/https://r9y9.github.io/blog/2013/07/27/nmf-euclid/

・NMFでMovieLensレコメンド/https://ohke.hateblo.jp/entry/2017/12/26/230000

・非負値行列因子分解を改めてやり直してみた/
https://qiita.com/sumita_v09/items/d22850f41257d07c45ea

http://www.sanographix.net/