# 8章　次元縮約

## NMFについて

7月8日（火）

豊川　幸秀

# NMFとは

- 次元縮約を利用したクラスタリング手法
- ベクトル空間モデルにおける文書データが対象
- データセットを表す行列をX、目標クラスタ数をKとした場合、Xは以下のように分解する。

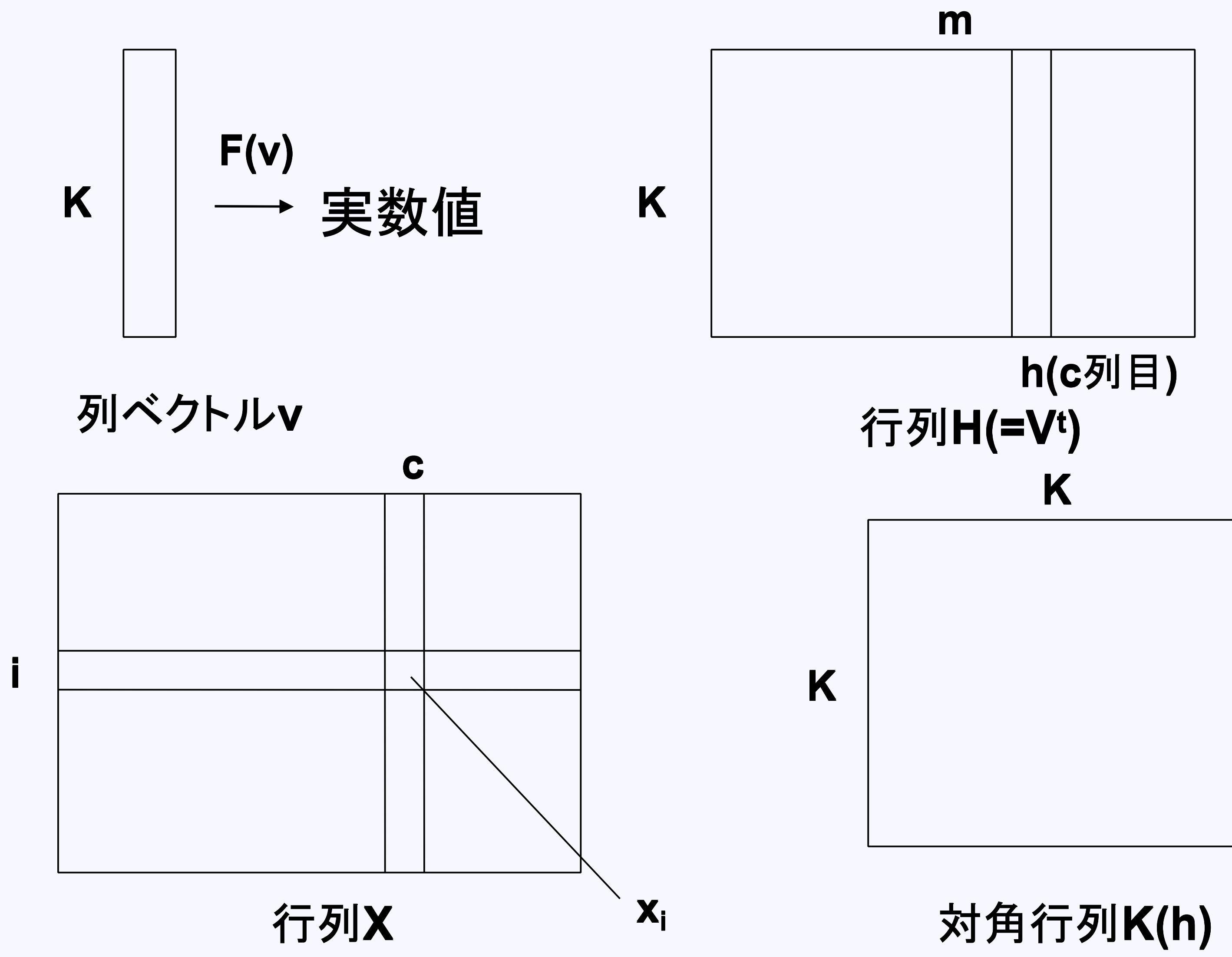
$$\underset{n \times m}{X} = \underset{n \times K}{U} \times \underset{K \times m}{V^t}$$

- UとVの各要素は非負
- Vのi行目のベクトル($V^t$のi列目)が、Xのi列目のベクトルをn次元からK次元に縮約した結果となる
- $V_{ik}$は、i番目のデータとk番目のトピックとの関連性の深さを表している
- U及びVは以下の式の繰り返しで求める

$$V'_{ij} \leftarrow V_{ij} \frac{(X^t U)_{ij}}{(VU^t U)_{ij}} \cdots ① \quad U'_{ij} \leftarrow U_{ij} \frac{(X^t V)_{ij}}{(UV^t V)_{ij}} \cdots ②$$

$\|X-UV^t\|$の値は小さくなっていくが、最終的な値は初期値$U^{(0)}$及び$V^{(0)}$に依存する。 → UとVは局所解

# 証明(定義)



- v:K次元の列ベクトル
- F(v):K次元から実数値へのある関数
- G(v,v'): G(v,v')≧F(v), G(v,v)=F(v)

  を満たすFの補助関数
- H:行列$V^t$と等しい、c番目の列ベクトルをhとする
- $x_i$:行列X上のi行c列目の要素
- K(h):i行i列の要素が

$$(K(h))_{ii} = \frac{(U^t U h)_i}{h_i}$$

  であるK×Kの対角行列

# ①の証明

F(h)及びG(h,h')を以下で定義

$$F(h) = \frac{1}{2}\sum_{i=1}^{n}(x_i - \sum_{j=1}^{K}U_{ij}h_j)^2 \quad (8.9)$$

$$G(h,h') = F(h') + (h-h')^t \nabla F(h') + \frac{1}{2}(h-h')^t K(h')(h-h') \quad (8.10)$$

(8.9)から、F(h)を最小にするhが||X-UH||を最小にするHのc番目の列ベクトルだとわかる。

(8.8)の更新式を利用するために、G(h,h')がF(h)の補助関数であることを示す。 → G(h,h')≧F(h)を示す。
F(h)は、

$$F(h) = F(h') + (h-h')^t \nabla F(h') + \frac{1}{2}(h-h')^t (U^t U)(h-h')$$

が成り立つことより、(8.10)とあわせて

$$(h-h')^t (K(h') - U^t U)(h-h') \geq 0$$

より、確かにG(h,h')はF(h)の補助関数である。

(8.10)のh'を$h^{(t)}$とし,F(h)の局所最小値を与える$h^{(t+1)}$を求めると、以下になる

$$h^{(t+1)} = h^{(t)} - (K(h^{(t)}))^{-1} \nabla F(h^{(t)})$$

変形して、

$$h_i^{(t+1)} = h_i^{(t)} \frac{(U^t x)_i}{(U^t U h^{(t)})_i}$$

すなわち、

$$H_{ic}^{(t+1)} = H_{ic}^{(t)} \frac{(U^t X)_{ic}}{(U^t U H^{(t)})_{ic}}$$

$(U^t X)_{ic} = (X^t U)_{ci}, (U^t U H^{(t)})_{ic} = (V^{(t)} U^t U)_{ci}$より、

$$V_{ij}^{(t+1)} = V_{ij}^{(t)} \frac{(X^t U)_{ij}}{(V^{(t)} U^t U)_{ij}} \qquad (8.15)$$

よって、①は証明できた。

# ②の証明

$\| X - U(V^{(t)})^{(t)} \| = \| (X - U(V^{(t)})^t)^t \| = \| X^t - V^{(t)} U^t \|$ より、

(8.15)を $X \to X^t, U^{(t)} \to V^{(t)}, V^t \to U^t$ とすればよい。

すなわち、

$$U_{ij}^{(t+1)} = U_{ij}^{(t)} \frac{(XV^t)_{ij}}{(U^{(t)}V^tV)_{ij}}$$

よって、②は証明できた。

# 正規化

U及びVは、①と②を繰り返すたびに確かに$||X-UV^t||$の値は減少するが、各々の行列の要素の値は増加したり減少したりする。
よって、繰り返しを行う度にどちらかの行列の大きさを以下の式で正規化する必要がある。

$$U_{ij} \leftarrow \frac{U_{ij}}{\sqrt{\sum_i U_{ij}^2}}$$

(Vについても同様)