非负矩阵分解

耿修瑞

中国科学院空天信息创新研究院

gengxr@sina.com.cn

2025.4

提纲

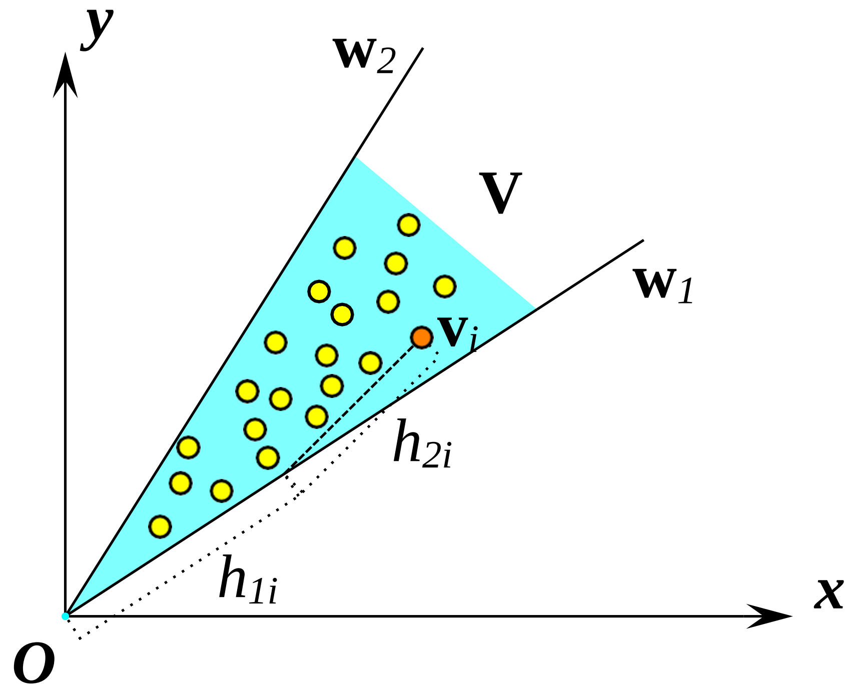
- 问题背景
- 非负矩阵分解
- 应用实例

问题背景

□ NMF示意图

$$\mathbf{V} = \mathbf{WH}$$

$$V_{ij}, W_{i\mu}, H_{\mu j} \geq 0$$



$$\mathbf{v}_i = h_{1i}\mathbf{w}_1 + h_{2i}\mathbf{w}_2 = \begin{bmatrix} \mathbf{w}_1 & \mathbf{w}_2 \end{bmatrix} \begin{bmatrix} h_{1i} \\ h_{2i} \end{bmatrix}$$

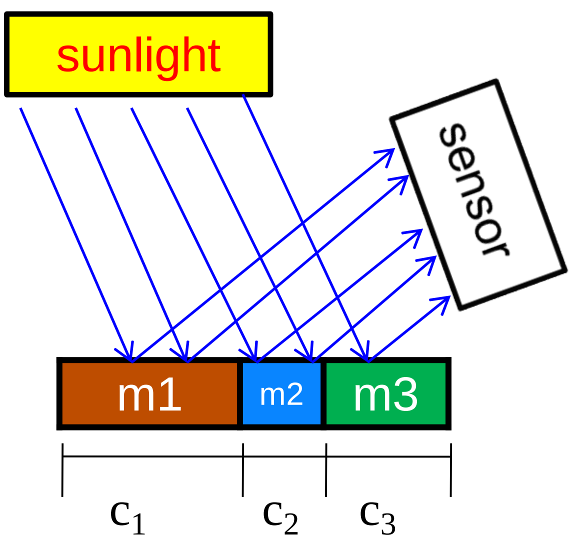
只有小黄点都在内部，才能保证h_{1i}, h_{2i}都非负。

问题背景

矩阵表达

$$\mathbf{R} = \mathbf{E}\mathbf{C}$$

把非负观测数据矩阵分解为两个非负矩阵的乘积！

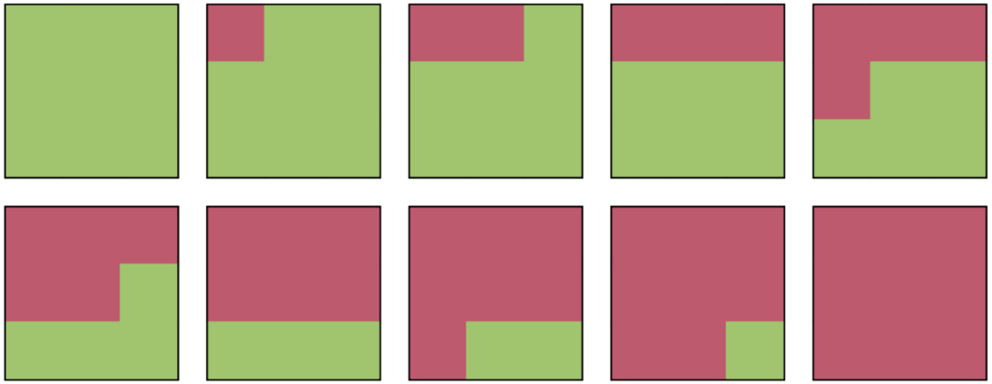


$$\mathbf{r}_i = \sum_{j=1}^{M} c_{ij}\mathbf{e}_j = \mathbf{E}\mathbf{c}_i$$

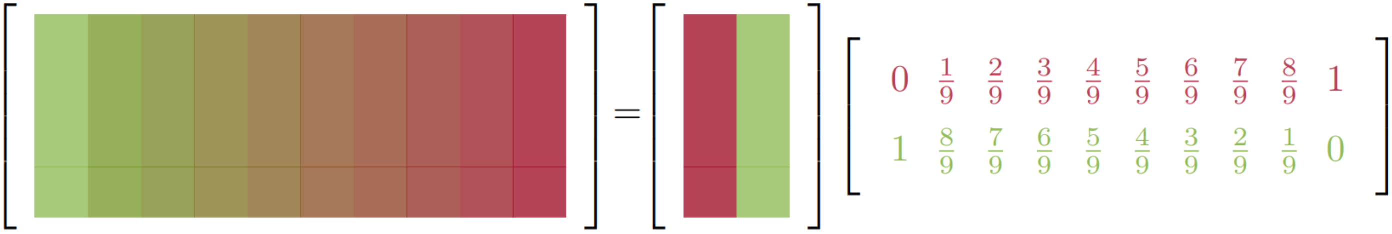
$$\mathbf{R} = \left[\mathbf{r}_1, \cdots, \mathbf{r}_N\right] = \left[\mathbf{E}\mathbf{c}_1, \cdots, \mathbf{E}\mathbf{c}_N\right] = \mathbf{E}\mathbf{C}$$

问题背景

混合像元示意图



非负矩阵分解示意图



问题背景

优化模型

$$\begin{cases} \min_{\mathbf{E},\mathbf{C}} f(\mathbf{E},\mathbf{C}) = \frac{1}{2}\|\mathbf{R}-\mathbf{E}\mathbf{C}\|_F^2 \\ s.t.\ \mathbf{E} \geq \mathbf{0}, \mathbf{C} \geq \mathbf{0} \end{cases}$$

找$\mathbf{E},\mathbf{C}$，使得$f(\mathbf{E},\mathbf{C})$达到极小值

问题背景

- 目标函数关于自变量的导数

$$\frac{\partial f\left(\mathbf{E},\mathbf{C}\right)}{\partial \mathbf{E}} = \mathbf{E}\mathbf{C}\mathbf{C}^T - \mathbf{R}\mathbf{C}^T$$

$$\frac{\partial f\left(\mathbf{E},\mathbf{C}\right)}{\partial \mathbf{C}} = \mathbf{E}^T\mathbf{E}\mathbf{C} - \mathbf{E}^T\mathbf{R}$$

问题背景

无约束最小二乘

$$\frac{\partial f\left(\mathbf{E},\mathbf{C}\right)}{\partial \mathbf{E}} = \mathbf{E}\mathbf{C}\mathbf{C}^T - \mathbf{R}\mathbf{C}^T \overset{令}{=} \mathbf{0} \Rightarrow \mathbf{E} = \mathbf{R}\mathbf{C}^T\left(\mathbf{C}\mathbf{C}^T\right)^{-1}$$

$$\frac{\partial f\left(\mathbf{E},\mathbf{C}\right)}{\partial \mathbf{C}} = \mathbf{E}^T\mathbf{E}\mathbf{C} - \mathbf{E}^T\mathbf{R} \overset{令}{=} \mathbf{0} \Rightarrow \mathbf{C} = \left(\mathbf{E}^T\mathbf{E}\right)^{-1}\mathbf{E}^T\mathbf{R}$$

令偏导数等0，即可得到常见的无约束最小二乘解，
无约束最小二乘的问题？

问题背景

梯度下降法

$$\mathbf{E} = \mathbf{E} - \alpha \frac{\partial f\left(\mathbf{E},\mathbf{C}\right)}{\partial \mathbf{E}}$$

$$\mathbf{C} = \mathbf{C} - \beta \frac{\partial f\left(\mathbf{E},\mathbf{C}\right)}{\partial \mathbf{C}}$$

梯度下降法的问题？

问题背景

投影梯度法

$$\mathbf{E} = \max\left(\mathbf{0}, \mathbf{E} - \alpha \frac{\partial f\left(\mathbf{E}, \mathbf{C}\right)}{\partial \mathbf{E}}\right)$$

$$\mathbf{C} = \max\left(\mathbf{0}, \mathbf{C} - \beta \frac{\partial f\left(\mathbf{E}, \mathbf{C}\right)}{\partial \mathbf{C}}\right)$$

问题背景

非负最小二乘

1. 首先得到 $\mathbf{E}$

2. 然后根据非负最小二乘得到 $\mathbf{C}$

$$\begin{cases} \min\limits_{\mathbf{c}_i} \left\| \mathbf{r}_i - \mathbf{E}\mathbf{c}_i \right\|^2 \\ s.t. \qquad \mathbf{c}_i \geq \mathbf{0} \end{cases}$$

$$\mathbf{c}_i = lsqnonneg(\mathbf{E}, \mathbf{r}_i)$$

非负矩阵分解

- NMF

（Lee&Seung，1999，nature）

非负矩阵分解

- NMF

$$\mathbf{E}_{ij} = \mathbf{E}_{ij} - \alpha_{ij}\left(\frac{\partial f\left(\mathbf{E},\mathbf{C}\right)}{\partial \mathbf{E}}\right)_{ij}$$

$$\mathbf{E}_{ij} = \mathbf{E}_{ij} + \alpha_{ij}\left[\left(\mathbf{RC}^{T}\right)_{ij} - \left(\mathbf{ECC}^{T}\right)_{ij}\right]$$

$$\alpha_{ij} = \frac{\mathbf{E}_{ij}}{\left(\mathbf{ECC}^{T}\right)_{ij}}$$

非负矩阵分解

- NMF

$$\mathbf{C}_{ij} = \mathbf{C}_{ij} - \beta_{ij}\left(\frac{\partial f\left(\mathbf{E},\mathbf{C}\right)}{\partial \mathbf{C}}\right)_{ij}$$

$$\mathbf{C}_{ij} = \mathbf{C}_{ij} + \beta_{ij}\left[\left(\mathbf{E}^{T}\mathbf{R}\right)_{ij} - \left(\mathbf{E}^{T}\mathbf{E}\mathbf{C}\right)_{ij}\right]$$

$$\beta_{ij} = \frac{\mathbf{C}_{ij}}{\left(\mathbf{E}^{T}\mathbf{E}\mathbf{C}\right)_{ij}}$$

非负矩阵分解

- NMF的乘式迭代

$$\mathbf{C}_{ij}=\mathbf{C}_{ij}\frac{\left(\mathbf{E}^{T}\mathbf{R}\right)_{ij}}{\left(\mathbf{E}^{T}\mathbf{E}\mathbf{C}\right)_{ij}}$$

$$\mathbf{E}_{ij}=\mathbf{E}_{ij}\frac{\left(\mathbf{R}\mathbf{C}^{T}\right)_{ij}}{\left(\mathbf{E}\mathbf{C}\mathbf{C}^{T}\right)_{ij}}$$

非负矩阵分解

- NMF的乘式迭代

$$\mathbf{E} = \mathbf{E}. \times \left(\mathbf{R}\mathbf{C}^T\right). / \left(\mathbf{E}\mathbf{C}\mathbf{C}^T\right)$$

$$\mathbf{C} = \mathbf{C}. \times \left(\mathbf{E}^T\mathbf{R}\right). / \left(\mathbf{E}^T\mathbf{E}\mathbf{C}\right)$$

讨论：该迭代能保证目标函数一直下降么？

非负矩阵分解

目标函数的负梯度方向：

$$-\frac{\partial f}{\partial \mathbf{E}} = \mathbf{RC}^{\mathrm{T}} - \mathbf{ECC}^{\mathrm{T}}$$

乘式迭代中的迭代方向方向：

$$\mathrm{A}.\times\left(-\frac{\partial f}{\partial \mathbf{E}}\right) = \mathrm{A}.\times\left(\mathbf{RC}^{\mathrm{T}} - \mathbf{ECC}^{\mathrm{T}}\right)$$

他们的内积：

$$\left\langle \mathrm{A}.\times\left(\mathbf{RC}^{\mathrm{T}} - \mathbf{ECC}^{\mathrm{T}}\right), \mathbf{RC}^{\mathrm{T}} - \mathbf{ECC}^{\mathrm{T}} \right\rangle \geq 0$$

因此乘式迭代的方向必然是指向目标函数下降的方向，但这并不意味着目标函数必然下降

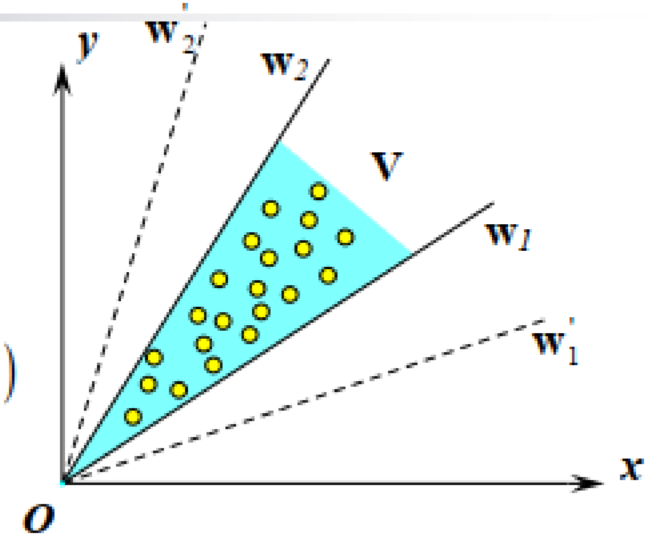
非负矩阵分解

NMF的全局收敛性

$$f(\mathbf{W},\mathbf{H})=\frac{1}{2}\left\|\mathbf{V}-\mathbf{W}\mathbf{H}\right\|_F^2 \qquad \mathbf{V}\in R^{m\times n},\mathbf{W}\in R^{m\times p},\ \mathbf{H}\in R^{p\times n}$$

$$=\frac{1}{2}vec\left(\mathbf{W}^T\right)^T\begin{bmatrix}\mathbf{H}\mathbf{H}^T & & \\ & \ddots & \\ & & \mathbf{H}\mathbf{H}^T\end{bmatrix}vec\left(\mathbf{W}^T\right)-trace\left(\mathbf{H}^T\mathbf{W}^T\mathbf{V}\right)+\frac{1}{2}trace\left(\mathbf{V}\mathbf{V}^T\right)$$

$$=\frac{1}{2}vec\left(\mathbf{H}\right)^T\begin{bmatrix}\mathbf{W}^T\mathbf{W} & & \\ & \ddots & \\ & & \mathbf{W}^T\mathbf{W}\end{bmatrix}vec\left(\mathbf{H}\right)-trace\left(\mathbf{W}\mathbf{H}\mathbf{V}^T\right)+\frac{1}{2}trace\left(\mathbf{V}^T\mathbf{V}\right)$$



当一个变量固定时，该目标函数为凸函数；而整体而言，目标函数是非凸的。

$$\nabla^2 f=\begin{bmatrix}\dfrac{\partial^2 f}{\partial vec\left(\mathbf{W}^T\right)\partial vec\left(\mathbf{W}^T\right)^T} & \dfrac{\partial^2 f}{\partial vec\left(\mathbf{W}^T\right)\partial vec\left(\mathbf{H}\right)^T}\\ \dfrac{\partial^2 f}{\partial vec\left(\mathbf{H}\right)\partial vec\left(\mathbf{W}^T\right)^T} & \dfrac{\partial^2 f}{\partial vec\left(\mathbf{H}\right)\partial vec\left(\mathbf{H}\right)^T}\end{bmatrix}=\begin{bmatrix}\tilde{\mathbf{H}} & \mathbf{A}\text{-}\mathbf{V}\otimes\mathbf{I}_p\\ \mathbf{A}^T\text{-}\mathbf{V}^T\otimes\mathbf{I}_p & \tilde{\mathbf{W}}\end{bmatrix}$$

$$\mathbf{K}_{pn}=\begin{bmatrix}\mathbf{I}_n\otimes e^1(p)\\ \vdots\\ \mathbf{I}_n\otimes e^p(p)\end{bmatrix}$$

$$\mathbf{A}=\left(\mathbf{W}\mathbf{H}\right)\otimes\mathbf{I}_p+\left(\mathbf{W}\otimes\mathbf{H}\right)\mathbf{K}_{pn}$$

$e^i\left(p\right)$为$\mathbf{I}_p$的第i行

$$\tilde{\mathbf{H}}=\begin{bmatrix}\mathbf{H}\mathbf{H}^T & & \\ & \ddots & \\ & & \mathbf{H}\mathbf{H}^T\end{bmatrix}\qquad \tilde{\mathbf{W}}=\begin{bmatrix}\mathbf{W}^T\mathbf{W} & & \\ & \ddots & \\ & & \mathbf{W}^T\mathbf{W}\end{bmatrix}$$

$$\frac{\partial tr\left(\mathbf{A}\mathbf{X}\mathbf{B}\right)}{\partial\mathbf{X}}=\mathbf{A}^T\mathbf{B}^T$$

$$vec\left(\mathbf{A}\mathbf{B}\mathbf{C}\right)=\left(\mathbf{C}^T\otimes\mathbf{A}\right)vec\left(\mathbf{B}\right)$$

非负矩阵分解

NMF的全局收敛性（A的推导）

$$g(\mathbf{W},\mathbf{H}) = tr\left(\mathbf{W}\mathbf{H}\mathbf{H}^T\mathbf{W}^T\right)$$

$$\frac{\partial g(\mathbf{W},\mathbf{H})}{\partial vec\left(\mathbf{W}^T\right)} = \left(I_m \otimes \mathbf{H}\mathbf{H}^T\right)vec\left(\mathbf{W}^T\right) = \tilde{\mathbf{H}}vec\left(\mathbf{W}^T\right) \qquad \frac{\partial g(\mathbf{W},\mathbf{H})}{\partial \mathbf{W}^T} = \mathbf{H}\mathbf{H}^T\mathbf{W}^T = \mathbf{I}_p\mathbf{H}\mathbf{H}^T\mathbf{W}^T$$

$$\frac{\partial^2 g(\mathbf{W},\mathbf{H})}{\partial vec\left(\mathbf{W}^T\right)\partial vec(\mathbf{H})^T} = \frac{\partial\left(I_m \otimes \mathbf{H}\mathbf{H}^T\right)vec\left(\mathbf{W}^T\right)}{\partial vec(\mathbf{H})^T} = D_{\mathbf{H}}\left(\mathbf{I}_p\mathbf{H}\mathbf{H}^T\mathbf{W}^T\right)$$

$$= (\mathbf{W}\mathbf{H}) \otimes \mathbf{I}_p + (\mathbf{W} \otimes \mathbf{H})\mathbf{K}_{pm} \overset{\text{记为}}{=} \mathbf{A}$$

$$\mathbf{F}(\mathbf{H}) = \mathbf{I}_p\mathbf{H}\mathbf{H}^T\mathbf{W}^T \qquad d\mathbf{F}(\mathbf{H}) = \mathbf{I}_p(d\mathbf{H})\mathbf{H}^T\mathbf{W}^T + \mathbf{I}_p\mathbf{H}\left(d\mathbf{H}^T\right)\mathbf{W}^T$$

$$dvec(\mathbf{F}(\mathbf{H})) = vec\left(\mathbf{I}_p(d\mathbf{H})\mathbf{H}^T\mathbf{W}^T\right) + vec\left(\mathbf{I}_p\mathbf{H}\left(d\mathbf{H}^T\right)\mathbf{W}^T\right)$$

$$= \left((\mathbf{W}\mathbf{H}) \otimes \mathbf{I}_p\right)vec(d\mathbf{H}) + (\mathbf{W} \otimes \mathbf{H})vec\left(d\mathbf{H}^T\right) = \left((\mathbf{W}\mathbf{H}) \otimes \mathbf{I}_p\right)vec(d\mathbf{H}) + (\mathbf{W} \otimes \mathbf{H})\mathbf{K}_{pm}vec(d\mathbf{H})$$

$$D_{\mathbf{H}}\left(\mathbf{I}_p\mathbf{H}\mathbf{H}^T\mathbf{W}^T\right) = (\mathbf{W}\mathbf{H}) \otimes \mathbf{I}_p + (\mathbf{W} \otimes \mathbf{H})\mathbf{K}_{pm}$$

定理：假设$\mathbf{F}(\mathbf{X}) : R^{m\times n} \to R^{p\times q}$，若$d\mathbf{F}(\mathbf{X}) = \mathbf{A}(d\mathbf{X})\mathbf{B} + \mathbf{C}d\mathbf{X}^T\mathbf{D}$，$D_{\mathbf{X}}\mathbf{F}(\mathbf{X}) = \mathbf{B}^T \otimes \mathbf{A} + \left(\mathbf{D}^T \otimes \mathbf{C}\right)K_{mn}$

非负矩阵分解

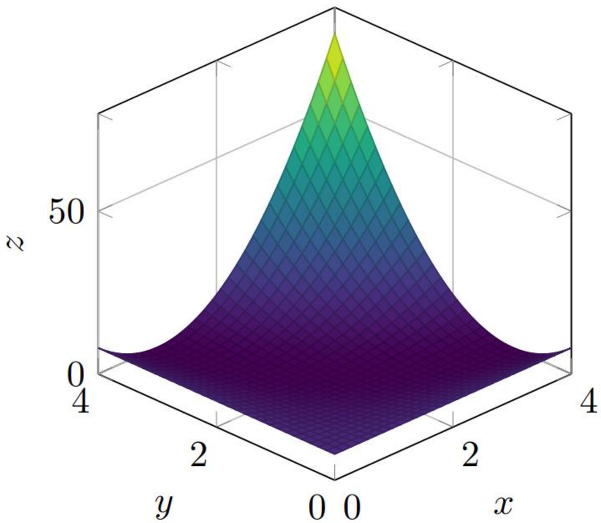
NMF的全局收敛性

$f(\mathbf{W},\mathbf{H})=\frac{1}{2}\|\mathbf{V}-\mathbf{WH}\|_F^2$ 未必是凸函数的例子

$$f(x,y)=\frac{1}{2}(4-xy)^2$$

取 $\mathbf{a}=[1\quad 4]^{\mathrm{T}},\mathbf{b}=[4\quad 1]^{\mathrm{T}}$

$$\frac{81}{32}=f\left(\frac{\mathbf{a}+\mathbf{b}}{2}\right)>\frac{f(\mathbf{a})+f(\mathbf{b})}{2}=0$$



定义1（**凸函数**）：凸函数是定义在某个线性空间的凸子集上的实值函数，且对该子集上的任意两个向量 $\mathbf{a},\mathbf{b}$，都有：

$$f\left(\frac{\mathbf{a}+\mathbf{b}}{2}\right)\le\frac{f(\mathbf{a})+f(\mathbf{b})}{2}$$

非负矩阵分解

NMF的概率解释

1.高斯分布情形。当模型误差满足高斯分布，且各个分量之间独立时，模型的误差函数为$\|\mathbf{V}-\mathbf{WH}\|_F^2$，相应的乘式迭代如前所述

2.泊松分布情形。当模型误差满足泊松分布时，对应的模型误差函数为

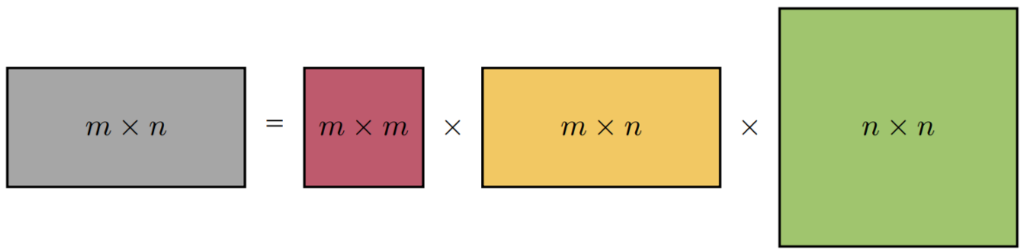
$$D\left(\mathbf{V} \| \mathbf{WH}\right)=\sum_{ij}\left(\mathbf{V}_{ij}\ln\frac{\mathbf{V}_{ij}}{\left(\mathbf{WH}\right)_{ij}}-\mathbf{V}_{ij}+\left(\mathbf{WH}\right)_{ij}\right)$$

$D\left(\mathbf{V} \| \mathbf{WH}\right)$称之为$\mathbf{V}$ 相对于$\mathbf{WH}$的散度。

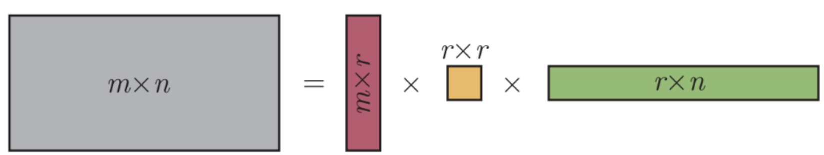
非负矩阵分解

NMF与SVD

对于观测数据 $\mathbf{V}=[\mathbf{v}_1 \quad \mathbf{v}_2 \quad \cdots \quad \mathbf{v}_n]$，设其秩为$r$，利用奇异值分解，可将其分解为$\mathbf{V}=\tilde{\mathbf{W}}\tilde{\mathbf{D}}\tilde{\mathbf{H}}^{\mathrm{T}}$，



或 $\mathbf{V}=\tilde{\mathbf{W}}(:,1:r)\tilde{\mathbf{D}}(1:r,1:r)\tilde{\mathbf{H}}(:,1:r)^{\mathrm{T}}$



如果不对 $\mathbf{W},\mathbf{H}$做任何约束的话，只需令 $\mathbf{WH}=\tilde{\mathbf{W}}\tilde{\mathbf{D}}\tilde{\mathbf{H}}^{\mathrm{T}}$，即可使 $\|\mathbf{V}-\mathbf{WH}\|_F^2$ 的模型误差达到最小。

非负矩阵分解

NMF与SVD

将目标函数的模型重新表述为

$$\|\mathbf{V}-\mathbf{W}\mathbf{H}\|_F^2=\sum_{i=1}^{n}\left\|\mathbf{v}_i-\sum_{j=1}^{r}h_{ji}\mathbf{w}_j\right\|^2$$

假设 $\mathbf{W}=[\mathbf{w}_1\quad \mathbf{w}_2\quad \cdots\quad \mathbf{w}_r]$已知，则上述目标函数可以转化为如下$n$ 个最优化问题。

$$\begin{cases}\min\limits_{\mathbf{h}_i}\left\|\mathbf{v}_i-\sum_{j=1}^{r}h_{ji}\mathbf{w}_j\right\|^2\\ s.t.\ \ \mathbf{h}_i\geq\mathbf{0}\end{cases}$$

当 $\mathbf{v}_i$ 在$\mathbf{W}$ 的列空间的投影在其r个向量张成的“锥体”内时，上面优化模型的解的非负性自动满足，此时目标函数值可以达到最小，即为$\mathbf{v}_i$ 到$\mathbf{W}$ 的各个列向量张成的超平面距离的平方。

非负矩阵分解

NMF与SVD

遗憾的是，只有当 **w** 的 r 个列向量张成的“锥体”的“锥角”足够大时，才能使得所有的非负观测数据在其上的投影都位于这个“锥体”之内。对于某些数据，其需要的“锥角”可能会大到第一象限无法容纳下这样的“锥体”，此时 **w** 的某些元素会由于非负性的约束而被强行设置为0。相应的目标函数也只可能在可行域的边界达到最小。

定理1 当非负矩阵分解的模型中的基矩阵的元素都大于零时，该矩阵的各个列张成的子空间与非负观测数据前 r 个左奇异向量张成的子空间相同。

非负矩阵分解

NMF与K-means的关系

假设观测数据为 $\mathbf{V}=[\mathbf{v}_1 \quad \mathbf{v}_2 \quad \cdots \quad \mathbf{v}_n]$ ，类别数为 r。

则K-means主要步骤为：

（1）初始化：给出 r 个初始种子点（或者叫聚类中心），记为 $\boldsymbol{\mu}_1$，$\boldsymbol{\mu}_2,\cdots$，$\boldsymbol{\mu}_r$。

（2）分类：分别计算每个观测点到各个种子点的欧氏距离，将其分配给距离其最近的中心点，从而得到 r 个类别。分别记为 C_1，$C_2,\cdots$，C_r

（3）类别中心更新：更新这 r 个类别的中心

（4）重复以上两个步骤直至迭代停止。

非负矩阵分解

NMF与K-means的关系

上述K-means聚类等价于如下目标函数的优化问题：

$$\min_{C,\boldsymbol{\mu}_j} J(C,\boldsymbol{\mu}_j)=\sum_{j=1}^{r}\sum_{\mathbf{v}_i\in C_j}\left\|\mathbf{v}_i-\boldsymbol{\mu}_j\right\|^2$$

记 $\mathbf{W}=[\boldsymbol{\mu}_1 \quad \boldsymbol{\mu}_2 \quad \cdots \quad \boldsymbol{\mu}_r]$，而聚类结果用元素取值只能为0或者1的矩阵 $\mathbf{H}=(h_{ji})$来表示，其中 $\mathbf{v}_i\in C_j$时，$h_{ji}=1$，而当$\mathbf{v}_i\notin C_j$ 时，$h_{ji}=0$。那么，上面的优化问题的目标函数可以重新表示为

$$J(C,\boldsymbol{\mu}_j)=\sum_{i=1}^{n}\sum_{j=1}^{r}h_{ji}\left\|\mathbf{v}_i-\boldsymbol{\mu}_j\right\|^2=\sum_{i=1}^{n}\left\|\mathbf{v}_i-\sum_{j=1}^{r}h_{ji}\boldsymbol{\mu}_j\right\|^2=\left\|\mathbf{V}-\mathbf{W}\mathbf{H}\right\|_F^2=f\left(\mathbf{W},\mathbf{H}\right)$$

从中可以看出，当基矩阵由各个类别中心向量构成，系数矩阵为类别归属矩阵时，矩阵的非负矩阵分解正好对应聚类分析中的K-means方法。

NMF与KKT定理

KKT定理是优化领域解决非线性约束优化问题的核心定理。非线性约束优化模型的一般形式为（只考虑不等式约束）：

$$\begin{cases} \min \ f(\mathbf{x}) \\ s.t \quad g_j(\mathbf{x}) \le 0 \quad (j = 1,2,\cdots,m) \end{cases}$$

相应的拉格朗日函数为：

$$L(\mathbf{x}, \boldsymbol{\lambda}) = f(\mathbf{x}) + \sum_{j=1}^{m} \lambda_j g_j(\mathbf{x})$$

假设 $\mathbf{x}$ 为优化问题的局部极小解，则必存在$\lambda_1, \lambda_2, \cdots, \lambda_m$，使得：

$$\begin{cases} (1) \ \nabla f(\mathbf{x}) + \sum_{j=1}^{m} \lambda_j \nabla g_j(\mathbf{x}) = 0 & \text{（梯度条件）} \\ (2) \ g_j(\mathbf{x}) \le 0 & \text{（约束条件）} \\ (3) \ \lambda_j g_j(\mathbf{x}) = 0 & \text{（松弛互补条件）} \\ (4) \ \lambda_j \ge 0 & \text{（非负条件）} \\ \quad j = 1,2,\cdots,m \end{cases}$$

$\mathbf{x}$ 称为优化问题的Kuhn-Tucker点。
由于Karush也类似考虑了约束优化的最优条件，所以上面的定理也叫KKT定理。

非负矩阵分解

NMF与KKT定理

$$\begin{cases} \min\limits_{\mathbf{W},\mathbf{H}} f(\mathbf{W},\mathbf{H}) = \frac{1}{2}\|\mathbf{V}-\mathbf{WH}\|_F^2 \\ \mathbf{W} \ge \mathbf{0}, \mathbf{H} \ge \mathbf{0} \end{cases}$$

相应的拉格朗日函数为：

$$F(\mathbf{W},\mathbf{H},\mathbf{A},\mathbf{B}) = \frac{1}{2}\|\mathbf{V}-\mathbf{WH}\|_F^2 + trace(\mathbf{AW}^T) + trace(\mathbf{BH}^T)$$

对自变量求偏导：

$$\frac{\partial F(\mathbf{W},\mathbf{H})}{\partial \mathbf{W}} = (\mathbf{WH}-\mathbf{V})\mathbf{H}^T + \mathbf{A} = \mathbf{0}$$

$$\frac{\partial F(\mathbf{W},\mathbf{H})}{\partial \mathbf{H}} = \mathbf{W}^T(\mathbf{WH}-\mathbf{V}) + \mathbf{B} = \mathbf{0}$$

非负矩阵分解

NMF与KKT定理

根据KKT定理中的互补松弛条件，有：

$$\mathbf{A}.\times\mathbf{W}=\mathbf{0}\qquad \mathbf{B}.\times\mathbf{H}=\mathbf{0}$$

因此：

$$(\mathbf{WH}-\mathbf{V})\mathbf{H}^{T}.\times\mathbf{W}=\mathbf{0}$$

$$\mathbf{W}^{T}(\mathbf{WH}-\mathbf{V}).\times\mathbf{H}=\mathbf{0}$$

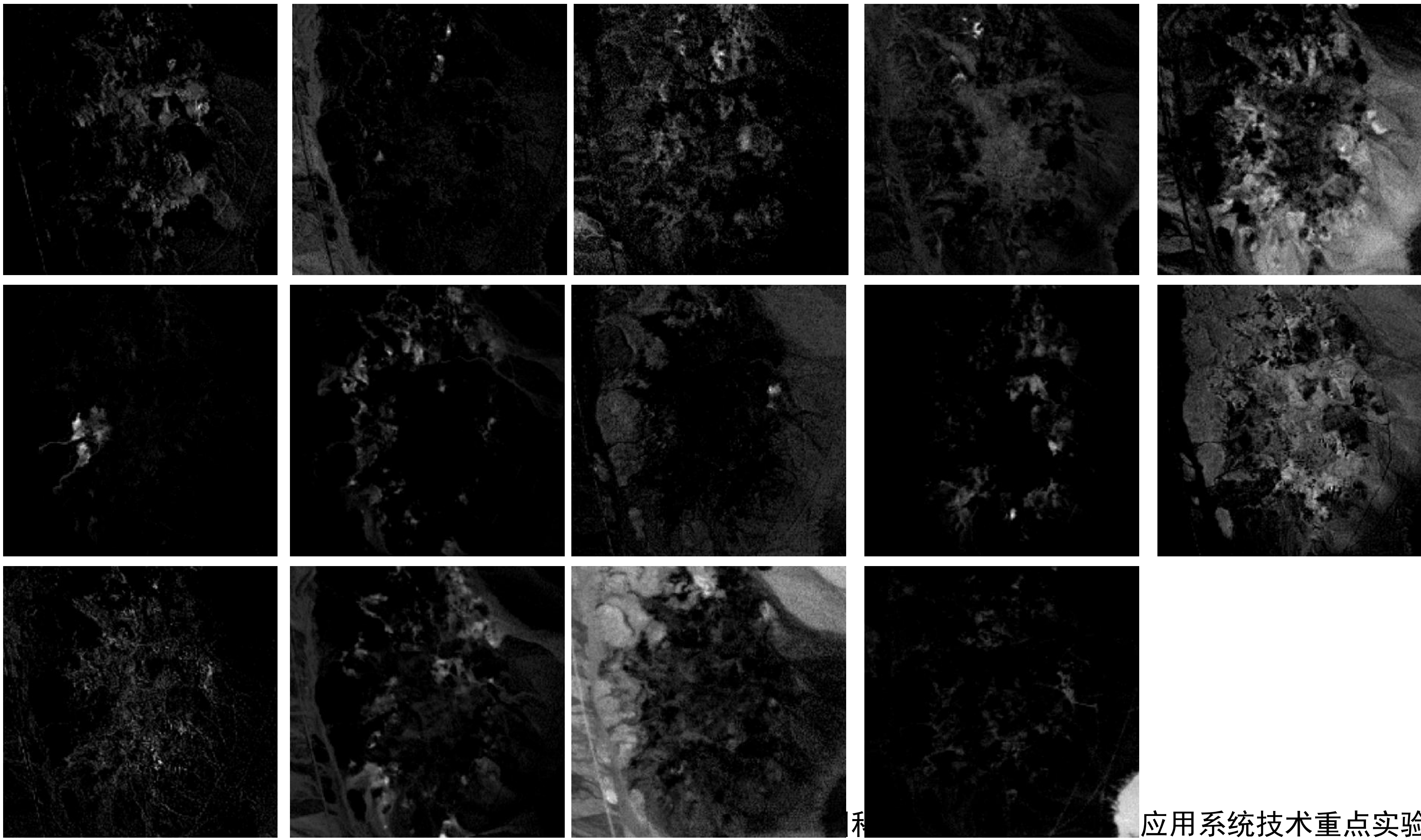
整理即得乘式迭代公式

$$\mathbf{W}=\mathbf{W}.\times\mathbf{VH}^{T}./\,\mathbf{WHH}^{T}$$

$$\mathbf{H}=\mathbf{H}.\times\mathbf{W}^{T}\mathbf{V}./\,\mathbf{W}^{T}\mathbf{WH}$$

非负矩阵分解

应用1：混合像元分析



非负矩阵分解

应用2：人脸识别

源图像

混合

NMF分离效果

非负矩阵分解

讨论

1. NMF的适用条件是什么？
2. 如果不满足非负条件，NMF会失效么？
3. 大张角情况下NMF可以使用么？
4. 非负张量分解？

非负矩阵分解

课堂练习

1. 试给出如下优化问题的NMF的乘式迭代公式？

$$\begin{cases} \min\limits_{\mathbf{E},\mathbf{C}} f\left(\mathbf{E},\mathbf{C}\right) = \dfrac{1}{2}\left\|\mathbf{R}-\mathbf{EC}\right\|_F^2 + \dfrac{1}{2}\lambda V\left(\mathbf{E}\right) \\ s.t.\ \mathbf{E} \geq \mathbf{0}, \mathbf{C} \geq \mathbf{0} \end{cases}$$

非负矩阵分解

总结

1. 非负矩阵分解的乘式迭代是一种特殊的梯度下降法，它巧妙地规避了常规优化方法中迭代方向和迭代步长的选取难题。
2. 从概率角度，对噪声概率密度分布不同的假设，会导致不同的目标函数。
3. 一定条件下，非负矩阵分解的基矩阵与奇异值分解的前几个相应的左奇异向量所张成的空间相同。也就是说，对于满足一定条件的数据，非负矩阵分解和SVD 或者 PCA 并没有本质的区别。
4. K-means 可以看做一种特殊的非负矩阵分解。
5. 非负矩阵分解可以看做最优化理论中 KKT 条件的一个应用。
6. 根据不同的应用场景，要对非负矩阵分解中的矩阵变量添加相应的约束，以使得最终的解具有相应的物理内涵。

谢　谢

耿修瑞

中国科学院空天信息创新研究院

gengxr@sina.com.cn