From Sparse Solutions of Systems of Equations to Sparse Modeling of Images

Minxuan Wang

2016 年 5 月 17 日

线性方程稀疏解

$$\min_{x} J(x) \ s.t. \ b = Ax \ A \in \mathbb{R}^{n\times m}$$

将J(x)理解为稀疏性(*sparsity*)，问题转化成线性方程的最稀疏解, 使用向量中非零值的个数来描述向量的稀疏性.

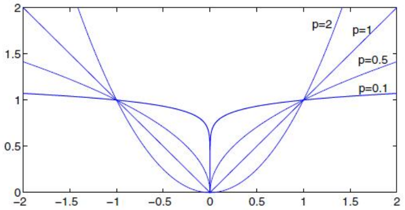
$$\|x\|_0 = \#\{i : x_i \neq 0\}$$

将矩阵A的列看作字典原子，b是原信号向量，x表示A中原子的线性组合，要求这个线性组合为最稀疏的

优化中的问题

NP-Hard 非凸问题，难以有效求解，但是假如A中的列不相关，*sparsity*(0范数问题)将会等价于1范数问题(凸函数)。另外对于某些种类的问题GA算法能够得到最稀疏解。

$$\|x\|_p = (\sum_i |x_i^p|)^{(1/p)}$$



最稀疏解的唯一性问题

什么时候存在最稀疏解?

spark定义: *The spark of a given matrix A is the smallest number of columns from A that are linearly dependent.*

$$\begin{bmatrix} 1 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix}$$

Rank=4 Spark=3

最稀疏解的唯一性问题

定理：*If a system of linear equations* $Ax = b$ *has a solution* x *obeying* $\|x\|_0$ <*spark(A)/2,, this solution is necessarily the sparsest possible.*

证明：若Ax=0，那么$\|x\|_0 \geqslant spark(A)$,因为Ax表示A中某几列的线性组合为0，说明A中这几列线性相关。故大于spark(A)

考虑另外一个解y，有A(x-y)=0,那么

$\|x\|_0 + \|y\|_0 \geqslant \|x - y\|_0 \geqslant spark(A)$

因为$\|x\|_0 < spark(A)/2$那么另外一个必然大于spark(A)/2所以没有比这个解更稀疏的解。

最稀疏解的唯一性问题

通常直接计算spark的比较困难。所以可以通过Mutual Coherence来刻画最稀疏解的唯一性。

定义:*The mutual coherence of a given matrix A is the largest absolute normalized inner product between different columns from A. Denoting the kth column in A by ak, the mutual coherence is given by*

$$\mu(A) = \max_{1 \leqslant k, j \leqslant m, k \neq j} \frac{|a_k^T a_j|}{\|a_k\|_2 \cdot \|a_j\|_2}$$

酉矩阵，$\mu(A)=0$

引理：*For any matrix* $A \in \mathbb{R}^{n\times m}$ *, the following relationship holds:*

$$spark(A) \geqslant 1 + \frac{1}{\mu(A)}$$

证明：首先normalize A矩阵的列向量，得到$\widetilde{A}$ *spark* μ都不变，Gram Matrix G=$\widetilde{A}^T\widetilde{A}$满足：

$\{G_{k,k} = 1 : 1 \leqslant k \leqslant m\}$ $\{|G_{k,j}| \leqslant \mu : 1 \leqslant k, j \leqslant m, k \neq j\}$

从$\widetilde{A}$中选取p列，并计算子Gram Matrix，如果此子矩阵是对角占优矩阵($\sum_{j\neq i}|G_{i,j}| < |G_{i,i}|$), 那么此矩阵是正定的(非奇异的)，故A中这p列线性无关。所以，如果有$p < 1 + 1/\mu$那么子矩阵就是正定的并且必然存在一个p满足介于$[1/\mu, 1 + 1/\mu]$之间(p任意取，可以取到$1 + 1/\mu$的下取整，而$1 + 1/\mu$的下取整大于$1/\mu$)所以存在p使$p + 1 \geqslant 1 + 1/\mu$,那么$spark(A) \geqslant 1 + 1/\mu$

求解最稀疏解

可以使用BP算法，求解1范数来代替0范数(凸优化问题)。

另外可以使用GA算法，例如MP，OMP。

目标：解优化$P_0 : min_x\|x\|_0 \ s.t.Ax = b$ 参数：A,b,错误阈值ϵ_0

初始化:$k = 0, x^0 = 0, r^0 = b - Ax^0 = b, S^0 = \varnothing$

迭代：(1)计算误差$\epsilon(j) = min_{z_j}\|a_j z_j - r^{k-1}\|_2^2$最优解为$z_j^* = a_j^T r^{k-1}\|a_j\|_2^2$

(2)找到最小的误差$\epsilon(j_0)$并将j_0保存进S，$S^k = S^{k-1} \cup \{j_0\}$

(3)计算x^k,根据S^k最小化$\|Ax - b\|_2^2$

(4)更新残差$r^k = b - Ax^k$

(5)残差小于阈值ϵ_0停止，最优解为x^k

可以证明如果满足$\|x\|_0 < \frac{1}{2}(1 + \frac{1}{\mu(A)})$则GA能收敛到最稀疏解

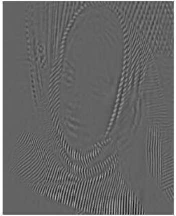
形态学成分分析MCA

假设将一副图像，考虑为两种信号的叠加。通过不同的稀疏模型来分离两种信号。

优化问题:$\min\limits_{x_1,x_2} \|x_1\|_0 + \|x_2\|_0 \ s.t. \|y - A_1x_1 - A_2x_2\|_2^2 \leqslant \epsilon_1^2 + \epsilon_2^2$

Original image

Cartoon part



Texture part

之前的介绍中字典都是固定的，K-SVD算法通过样本学习来更新字典。

目标：找到最佳的字典来表示样本$\{y_i\}_{i=1}^N$,优化目标：

$$\min_{D,X}\{\|Y-DX\|_F^2\}\ s.t.\ \forall i, \|x_i\|_0 < T0$$

设置$D^0 \in \mathbb{R}^{n\times K}$,J=1

使用之前的算法对每一个y_i计算x_i,i=1,2,...,N

$$\min_{x_i}\{\|y_i-Dx_i\|_2^2\}\ s.t.\ \|x_i\|_0 \leqslant T0$$

对于

$\|Y-DX\|_F^2 = \|Y-\sum_1^K d_j x_T^j\|_F^2 =$
$\|(Y-\sum_{j\neq k} d_j x_T^j) - d_k x_T^k\|_F^2 = \|E_k - d_k x_T^k\|_F^2$

做一个变换:$\|E_k\Omega_k - d_k x_T^k \Omega_k\|_F^2 = \|E_R - d_k x_R^k\|_F^2$
使用奇异值分解SVD,得到$E_k^R = U\Delta V^T$
$d - U, x - V$

K-SVD应用

很容易理解,使用K-SVD可以得到图像的稀疏表示，只需存储码表以及字典就可以保存图像的信息。实现了图像压缩。

另外K-SVD可以应用于图像降噪.原理是降维是一种有效的去除噪声的方法。可以假设无噪声图像可以很好得使用稀疏表示，当对一个噪声图像进行稀疏表示时，噪声部分随着降维被滤掉。

去噪问题要优化的问题是

$$\tilde{\alpha} = argmin_{\alpha}\|\alpha_0\| \ s.t. \ \|D\alpha - y\|_2^2 \leqslant T$$

改写成罚函数的形式：$\tilde{\alpha} = argmin_{\alpha}\|D\alpha - y\|_2^2 + \mu\|\alpha_0\|$

当图像比较大的时候，K-SVD字典学习会很慢，有两种方法来处理，其中之一是将大图像分解成一块一块的小图像来处理如$\sqrt{n} \times \sqrt{n}$

K-SVD应用

罚函数变为：$\{\widetilde{\alpha_{ij}}, \widetilde{X}\} =$
$argmin_{\{\alpha_{ij}, X\}} \lambda \|X - Y\|_2^2 + \sum_{ij} \mu_{ij} \|\alpha_{ij}\|_0 + \sum_{ij} \|D\alpha_{ij} - R_{ij} X\|_2^2$
将D看作已知，然后Y=X,可以得到所有的优化解$\tilde{\alpha}$类似之前的K-SVD操作更新D,并解出X,优化目标
$\widetilde{X} = argmin_x \lambda \|X - Y\|_2^2 + \sum_{ij} \|D\alpha_{ij} - R_{ij}\|_2 X^2$