Manifold Learning and Sparse Representation

Tao Hu
Department of Computer Science
Peking University
taohu@pku.edu.cn

1 Solution List

please click to jump to the corresponding solution.

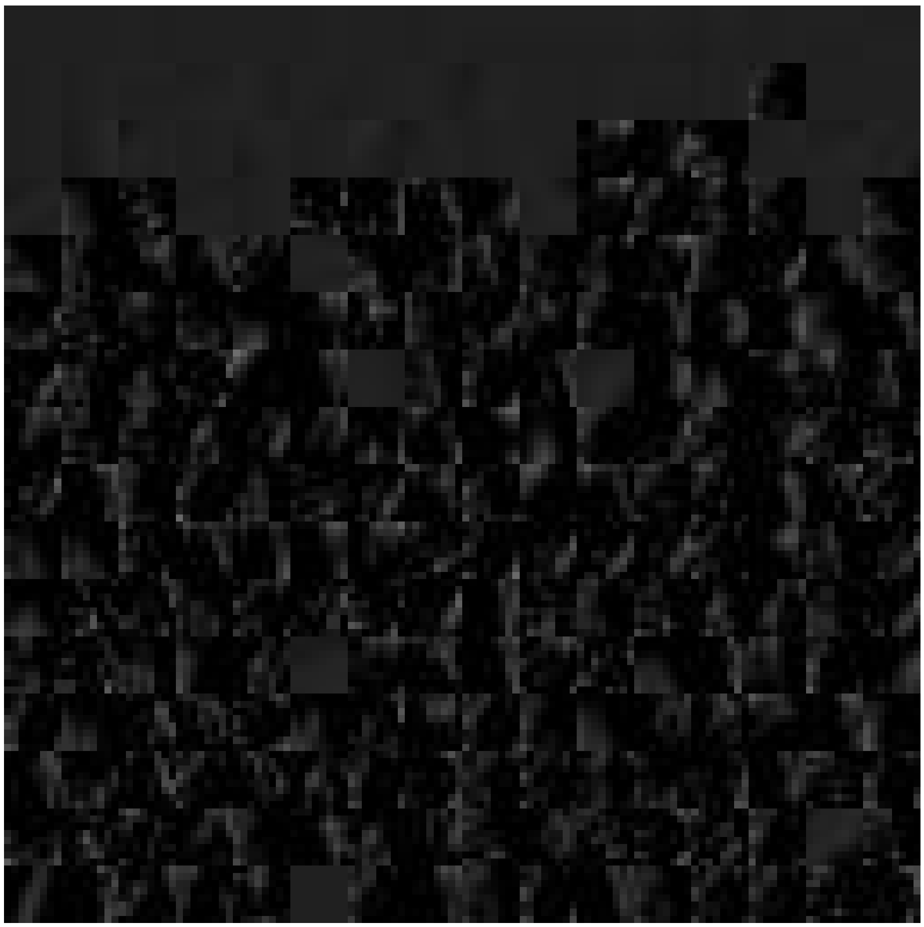
2 First-Order Sparse Representation

课后作业190. *Use image Lena and K-SVD to learn a 256-atom dictionary for* 8×8 *patches. Order the dictionary from "low frequency" to "high frequency" and rearrange in 2D in the zigzag manner.*

---

atom: 256
patch size: 8*8
patch number: 1000

KSVD iteration times: 1000
the code is in `https://github.com/dongzhuoyao/sparserepresentation.git`
the final dictionary rearrange in 2D in the zigzag manner is :



---

for Basis Sorting Algorithm[2], n is data number, m is base amount.(we need to sort the m bases.) S is 1-m vector just responsible for the number recording. I^i is 1-m vector recording the indice. and $U^{(x_i)}$ is a set of base vector.

KSVD procedure:

Algorithm 7 The K-SVD dictionary-learning algorithm.

Initialize: $k = 0$, build $\mathbf{A}_{(0)} \in \mathbb{R}^{n\times m}$, either by using random entries, or using m randomly chosen examples, normalize the columns of $\mathbf{A}_{(0)}$.

while If the change in $\|\mathbf{Y} - \mathbf{A}_{(k)}\mathbf{X}_{(k)}\|_F^2$ is not small enough **do**

Sparse Coding: Use a pursuit algorithm to approximate the solution of

$$\hat{\mathbf{x}}_i = \underset{\mathbf{x}}{\operatorname{argmin}} \|\mathbf{y}_i - \mathbf{A}_{(k-1)}\mathbf{x}\|^2, \quad s.t. \quad \|\mathbf{x}\|_0 \leq k_0,$$

obtaining sparse representations $\hat{\mathbf{x}}_i$ for $1 \leq i \leq M$. These form the matrix $\mathbf{X}_{(k)}$.

K-SVD Dictionary-Update: Use the following procedure to update the columns of the dictionary and obtain $\mathbf{A}_{(k)}$:

Repeat for $j_0 = 1, 2, \cdots, m$,

1. Define the group of examples that use the atom $\mathbf{a}_{j_0}$,

$$\Omega_{j_0} = \{i | 1 \leq i \leq M, \mathbf{X}_{(k)}[j_0, j] \neq 0\}.$$

2. Compute the residual matrix

$$\mathbf{E}_{j_0} = \mathbf{Y} - \sum_{j \neq j_0} \mathbf{a}_j \mathbf{x}_j^T,$$

where $\mathbf{x}_j^T$ are the j'th rows in the matrix $\mathbf{X}_{(k)}$.

3. Restrict $\mathbf{E}_{j_0}$ by choosing only the columns corresponding to Ω_{j_0}, and obtain $\mathbf{E}_{j_0}^R$.

4. Apply SVD on $\mathbf{E}_{j_0}^R = \mathbf{U}\Sigma\mathbf{V}^T$. Update the dictionary atom $\mathbf{a}_{j_0} = \mathbf{u}_1$, and the representations by $\mathbf{x}_{j_0}^R = \Sigma(1,1) \cdot \mathbf{v}_1$.

5. **Update** k**:** Increase k by 1.

end while

Ensure: Output $\mathbf{A}_{(k)}$.

Bases Sorting procedure:

Algorithm 8 Bases Sorting (BS) algorithm

1: **Input:** training data matrix $X = (\mathbf{x}_1, \cdots, \mathbf{x}_n)$, dictionary $U = (\mathbf{u}_1, \cdots, \mathbf{u}_m)$, and parameter $\epsilon > 0$.
2: Initialize the accumulation buffer: $\mathbf{s} = (0, \cdots, 0)$.
3: **for** $i = 1$ to n **do**
4: Calculate the sparsest representation of $\mathbf{x}_i$ by solving

$$\mathbf{v}_i = \arg\min_{\mathbf{v}} \|\mathbf{v}\|_1, \quad s.t. \quad \|\mathbf{x}_i - U\mathbf{v}\|_2^2 < \epsilon. \tag{7.86}$$

5: Identify the support of sparse representation vector $\mathbf{v}_i$ and the set of active bases:

$$\begin{aligned} I^{(i)} &= \{j \mid v_{ij} \neq 0, \ j = 1, \cdots, m\}, \\ U^{(\mathbf{x}_i)} &= \{\mathbf{u}_j | j \in I^{(i)}\}. \end{aligned} \tag{7.87}$$

6: Compute the magnitude vector $\mathbf{a}_i$ of active coefficients, i.e., $a_{ij} = |v_{ij}|$.
7: Sort the activated bases in $U^{(\mathbf{x}_i)}$ according to the descending order of the magnitude vector $\mathbf{a}_i$:

$$\pi(U^{(\mathbf{x}_i)}) = \text{sort}(I^{(i)}, \mathbf{a}_i). \tag{7.88}$$

8: Update the accumulation buffer: $s_j = s_j + 1$ for all $j \in I^{(i)}$.
9: **end for**
10: Compute the averaged order by

$$\pi(U) = \ \sum_{i=1}^{n} \pi(U^{(\mathbf{x}_i)}) \ ./\mathbf{s}, \tag{7.89}$$

where ./ is element-wise division.
11: Sort U in an ascending order of $\pi(U)$.

3 Structured Sparsity

Group Lasso[1]

课后作业194. *Prove that the overlapped group Lasso $\Omega^{\mathcal{G}}_{overlap}(\mathbf{w})$ is a norm.*

课后作业195. *Prove that*

$$\Omega^{\mathcal{G}}_{overlap}(\mathbf{w}) = \sup\{\boldsymbol{\alpha}^T\mathbf{w} | \boldsymbol{\alpha} \in \mathbb{R}^k, \|\boldsymbol{\alpha}_g\| \leq 1, \forall g \in \mathcal{G}\}.$$

课后作业196. *Consider 2D DCT transform. We know that for image patches, their high frequencies are more likely to be zeros than low frequencies are. So if we want to recover an image patch, we want the high frequencies to be zeros* **before** *the low frequencies. Then how to design a group sparsity regularizer on such a prior? Please refer to Figure 7.8(b). The entries on the i-th anti-diagonal are called of frequency i. Then this prior means that if $i < j$ and entries of frequency i are zeros, then entries of frequency j are also zeros.*

课后作业197. *Explain why the unit balls of $\Omega^{\mathcal{G}}_{group}(\cdot)$ and $\Omega^{\mathcal{G}}_{overlap}(\cdot)$ for the groups $\mathcal{G} = \{\{1,2\},\{2,3\}\}$ are as the left and the right figures in Figure 8.10, respectively.*

4 Second Order Sparsity: Linear Models

命题201. *The function $\|\mathbf{D}\,\mathrm{Diag}(\mathbf{w})\|_*$ is a norm w.r.t. the vector $\mathbf{w}$, provided that none of the columns of $\mathbf{D}$ is equal to $\mathbf{0}$.*

课后作业202. *Prove Proposition 201.*

定理228 ([84]). *Let $\mathbf{D} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T$ be the SVD of data matrix $\mathbf{D}$. The optimal solution to*

$$\min_{\mathbf{A},\mathbf{Z}} \|\mathbf{Z}\|_* + \frac{\alpha}{2}\|\mathbf{D}-\mathbf{A}\|_F^2, \quad s.t. \quad \mathbf{A} = \mathbf{A}\mathbf{Z}, \tag{10.68}$$

is given by $\mathbf{A}^* = \mathbf{U}_1\mathbf{\Sigma}_1\mathbf{V}_1^T$ and $\mathbf{Z}^* = \mathbf{V}_1\mathbf{V}_1^T$, where $\mathbf{\Sigma}_1$, $\mathbf{U}_1$, and $\mathbf{V}_1$ correspond to the top $r = \mathrm{argmin}_k \left(k + \frac{\alpha}{2}\sum_{i>k}\sigma_k^2\right)$ singular values and singular vectors of $\mathbf{D}$, respectively.

课后作业229. *Prove Theorem 228.*

命题238. *If f satisfies the EBD conditions (1), (2), and (3) on Ω, then also on $\Omega_1 \subseteq \Omega$, where $\Omega_1 \neq \emptyset$.*

命题239. *Let $\{f_i\}_{i=1}^m$ be a set of functions. If f_i satisfies the EBD conditions (1), (2), and (3) on Ω_i, $\forall i$, and $\cap_{i=1}^m \Omega_i \neq \emptyset$, then $\sum_{i=1}^m \lambda_i f_i$ also satisfies the EBD conditions on $\cap_{i=1}^m \Omega_i$, where $\lambda_i > 0$, $\forall i$.*

课后作业240. *Prove Propositions 238 and 239.*

5 Second Order Sparsity: Nonlinear Models

References

[1] Laurent Jacob, Guillaume Obozinski, and Jean-Philippe Vert. Group lasso with overlap and graph lasso. In *Proceedings of the 26th annual international conference on machine learning*, pages 433–440. ACM, 2009.

[2] Chun-Guang Li, Zhouchen Lin, and Jun Guo. Bases sorting: Generalizing the concept of frequency for over-complete dictionaries. *Neurocomputing*, 115:192–200, 2013.