K–SVD: An Algorithm for Designing Overcomplete Dictionaries for Sparse Representation

Yao Zhang

1.1 Introduction

We assume that the vector $\boldsymbol{x}$ is sparse, i.e. there are only a few non-zeros.

Definition 1.1 (spare representation of signal)**.** *Let $\boldsymbol{y} \in \mathbb{R}^{n\times 1}$ be an observed signal. Let $D \in \mathbb{R}^{n\times K}$ be a dictionary. Let $\boldsymbol{x} \in \mathbb{R}^{K\times 1}$ be the representation coefficients. In the absence of noise. we assume*

$$\boldsymbol{y} = D\boldsymbol{x}. \tag{1.1}$$

More precisely, if $D = [\boldsymbol{d}_1, ..., \boldsymbol{d}_K]$, where $\boldsymbol{d}_k \in \mathbb{R}^{n\times 1}$, are the basis vectors, then

$$\boldsymbol{y} = \sum_{k=1}^{K} x_k \boldsymbol{d}_k \tag{1.2}$$

Suppose we have $\{\boldsymbol{y}_1, ..., \boldsymbol{y}_N\}$ observations share a common dictionary D and is known to have spare representations. How to design D ?

Our problem is:

$$\min_{X,D} \|\boldsymbol{x}_i\|_0 \quad s.t. \;\; DX = Y \tag{1.3}$$

where $Y = [\boldsymbol{y}_1, ..., \boldsymbol{y}_N](\boldsymbol{y}_i \in \mathbb{R}^{n\times 1})$ is the collection of N observations, and $X = [\boldsymbol{x}_1, ..., \boldsymbol{x}_N](\boldsymbol{x}_i \in \mathbb{R}^{K\times 1})$ is the collection of N representation coefficient vectors.

Remark. *Eq 1.3 is not convex, and in reality, there is always noise and so $DX \approx Y$.*

Therefore,

$$\min_{X,D} \|DX - Y\|_F^2 \quad s.t. \;\; \|\boldsymbol{x}_i\|_0 \leqslant T \tag{1.4}$$

Solve the problem 1.4 using alternating minimization:

1. update the sparse coding: $X^{k+1} = \min_{X} \left\|D^{(k)}X - Y\right\|_F^2 \;\; s.t. \; \|\boldsymbol{x}_i\|_0 \leqslant T$
2. update the dictionary: $D^{k+1} = \min_{D} \left\|DX^{(k+1)} - Y\right\|_F^2$

1.2 K-Means

Remark. *K is from $D \in \mathbb{R}^{n \times K}$, means is from average, respectively.*

Sparse coding

Suppose we have a dictionary D. For now let us assume that D is known and fixed. Suppose we want to fire one and only column(denote k). How should we do it?

1. For the i-th observation $\boldsymbol{y}_i (i = 1, ..., N)$, we should select column k if

$$\|\boldsymbol{y}_i - D\boldsymbol{e}_k\|_2^2 \leqslant \|\boldsymbol{y}_i - D\boldsymbol{e}_j\|_2^2 \tag{1.5}$$

   for $j \neq k$, where $\boldsymbol{e}_k = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$ is the standard basis.

2. Repeat the same process for all $i = 1, ..., N$. The each $\boldsymbol{y}_i$ will have its own closest column and we can partition the indices $\{1, ..., N\}$ into at most K groups $R_1, ..., R_k$:

$$R_k = \left\{ i : \ \|\boldsymbol{y}_i - D\boldsymbol{e}_k\|_2^2 \leqslant \ \|\boldsymbol{y}_i - D\boldsymbol{e}_j\|_2^2 \,, \ j \neq k \right\} \tag{1.6}$$

Dictionary Update

Now, once the observations $\boldsymbol{y}_1, ..., \boldsymbol{y}_N$ are grouped into K groups specified by $R_1, ..., R_K$, how can we update the dictionary D?

1. Replace the column by the mean of observations in the group:

$$\boldsymbol{d}_k = \frac{1}{|R_k|} \sum_{i \in R_k} \boldsymbol{y} y_i \tag{1.7}$$

   Why? See Theorem 1.1.

2. After that, go back to the sparse coding step. Stop until stopping criteria is met.

Theorem 1.1. *K-Means is equivalent to*

$$\min_{D,X} \|Y - DX\|_F^2 \quad s.t. \ \forall i, \exists k, \ \boldsymbol{x}_i = \boldsymbol{e}_k \tag{1.8}$$

where $Y = [\boldsymbol{y}_1, ..., \boldsymbol{y}_N] \in \mathbb{R}^{n \times N}, X = [\boldsymbol{x}_1, ..., \boldsymbol{x}_N] \in \mathbb{R}^{K \times N}$

Proof.

1. Given D, if we can only fire one column(denote k), then the solution has to satisfy

$$\|\boldsymbol{y}_i - D\boldsymbol{e}_k\|_2^2 \leqslant \|\boldsymbol{y}_i - D\boldsymbol{e}_j\|_2^2 \tag{1.9}$$

2. Given X, we can separate the sum-square into K groups of individual terms. Each will take the form

$$\min_{\boldsymbol{d}_k} \sum_{i \in R_k} \|\boldsymbol{y}_i - \boldsymbol{d}_k\|_2^2 \tag{1.10}$$

   The optimal solution of 1.10 is the average of $\{\boldsymbol{y}_i\}_{i \in R_k}$

□

1.3 K-SVD

Remark. *K is from $D \in \mathbb{R}^{n \times K}$, SVD is from rank 1 decomposition, respectively.*

In fact, K-Means

$$\min_{D,X} \|Y - DX\|_F^2 \;\; s.t. \;\forall i, \;\|\boldsymbol{x}_i\|_0 = 1 \tag{1.11}$$

Now, K-SVD:

$$\min_{D,X} \|Y - DX\|_F^2 \;\; s.t. \;\forall i, \;\|\boldsymbol{x}_i\|_0 \leq T \tag{1.12}$$

1. Sparse Coding

   Fix D, solve X in

$$\min_{X} \|Y - DX\|_F^2 \;\; s.t. \;\forall i, \;\|\boldsymbol{x}_i\|_0 \leq T \tag{1.13}$$

   Note that

$$\|Y - DX\|_F^2 = \sum_{i=1}^{N} \|\boldsymbol{y}_i - D\boldsymbol{x}_i\|_2^2 \tag{1.14}$$

   Why do this?

   Thus, we only need to solve

$$\min_{\boldsymbol{x}_i} \|\boldsymbol{y}_i - D\boldsymbol{x}_i\|_2^2 \, s.t. \;\forall i, \;\|\boldsymbol{x}_i\|_0 \leqslant T \tag{1.15}$$

   This can be done using OMP, or any other algorithm along the same vein.

2. Dictionary Update

   Now, assume X is fixed. Suppose we want to update D.

   Can we solve this?

$$\min_{D} \|DX - Y\|_F^2 \tag{1.16}$$

   If we solve for D in this way, i.e.

$$D = YX^T \left(XX^T\right)^{-1} \tag{1.17}$$

   But, there are drawback in this method

   (a) $X \in \mathbb{R}^{K\times N}$, so $XX^T \in \mathbb{R}^{K\times K}$, Inversion is hard for large K.
   (b) There is no way of preserving sparsity inherent from X.

   Can we update the k-th column of D while fixing the others?

   We know that (let $\boldsymbol{x}^i$ be the j-th row, and $\boldsymbol{x}_j$ be the j-th column)

$$Y \approx \begin{bmatrix} \underbrace{\boldsymbol{d}_1}_{\in\mathbb{R}^{n\times 1}}, ..., \boldsymbol{d}_K \end{bmatrix} \begin{bmatrix} \underbrace{\boldsymbol{x}^1}_{\in\mathbb{R}^{1\times N}} \\ \vdots \\ \boldsymbol{x}^K \end{bmatrix} \in \mathbb{R}^{n\times N} \tag{1.18}$$

   then

$$\|Y - DX\|_F^2$$
$$= \left\| Y - \sum_{j=1}^{k} \boldsymbol{d}_j \boldsymbol{x}^j \right\|_F^2 = \left\| \left(Y - \sum_{j=1,j\neq k}^{k} \boldsymbol{d}_j \boldsymbol{x}^j \right) - \boldsymbol{d}_k \boldsymbol{x}^k \right\|_F^2 \triangleq \left\| E_k - \boldsymbol{d}_k \boldsymbol{x}^k \right\|_F^2 \tag{1.19}$$

   Therefore, since E_k is fixed, finding $\left(\boldsymbol{d}_k, \boldsymbol{x}^k\right)$ is the same as finding the best rank-1 update of E_k.

   To find $\left(\boldsymbol{d}_k, \boldsymbol{x}^k\right)$ such that

$$\min_{\boldsymbol{d}_k, \boldsymbol{x}^k} \left\| E_k - \boldsymbol{d}_k \boldsymbol{x}^k \right\|_F^2 \tag{1.20}$$

   Not quite! We also need to preserve sparsity of $\boldsymbol{x}^k$.

   Then, let us restrict ourselves to the existing non-zeros of $\boldsymbol{x}^k$. Define

$$\omega_k = \left\{ i : \boldsymbol{x}^k [i] \neq 0 \right\} \tag{1.21}$$

   and ω_k be an $n \times |\omega_k|$ matrix representing the sampling operator. Find rank-1 approximation for

$$\min_{\boldsymbol{d}_k, \boldsymbol{x}^k} \left\| E_k \Omega_k - \boldsymbol{d}_k \boldsymbol{x}^k \Omega_k \right\|_F^2 \tag{1.22}$$

Let $E_k^R \triangleq E_k \Omega_k$, then 1.22 can be solved by doing SVD on E_k^R:

$$E_k^R = U \Sigma V^T \tag{1.23}$$

Then

(a) $\boldsymbol{d}_k =$ the first column of U

(b) $\boldsymbol{x}^k \Omega_k =$ the first row of $V \times \Sigma(1,1)$. Fill $\boldsymbol{x}^k[i]$ with zero for $i \notin \omega_k$

(c) Repeat the process for $k = 1, ..., K$

Suppose the sparse coding is prefect. Then the dictionary update:

1. guarantees reduction or no charge of $\|Y - DX\|_F^2$
2. guarantees sparsity of X is unchanged

However,

1. Since the sparse coding step may not be perfect, convergence is not guaranteed i general
2. When T is small, OMP has worst case guarantee. Even for moderate T, OMP can still work under high probability (with some assumptions on the signal)
3. Practically, K-SVD works reasonably well (but slow)

1.4 Acknowledge

We would like to thank Dr. Stanley Chan for help with the description of the big picture of K-SVD.

1.5 References

[1] M. Aharon, M. Elad and A. Bruckstein, "K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation," *IEEE Transactions on Signal Processing* , 2006, Vol. 54(11), pp. 4311–4322.