Robust Face Recognition Via Sparse Representation

→ General Idea

→ Pre Processing of Data

→ Significance of l^1 norm

→ Problem Formulation

→ SRC Algorithm → Steps

→ Data Validation → Test case.
  ↳ Sparsity concentration Index (SCI).

→ Role of Feature Extraction.

→ Robustness to Occlusion.

Robust Face Recognition
Via
Sparse Representation

PAMI · 2009

Idea

→ General classification algorithm for object recognition (image based) based on Sparse representation with l^1-minimization

Classification based on Sparse Representation

Overcomplete Dictionary: A

Number of equations < Number of unknown

Steps

(1) → 1207 images from Extended Yale B database were selected for training.

(2) → Subsample → images from original size 192 X 168 to 12 X 1

(3) → Convert 12 X 10 to 120 X 1 1D Vector

(4) → ADD each 120 X 1 Vector as a column of Matrix A
Size of A → $\boxed{120 \times 1207}$

Matrix A →

Concatenation of n training samples of K object classes

$$A = [A_1, A_2, \ldots, A_K]$$

$$= [v_{1,1}, v_{1,2} \ldots\ldots, v_{K,n_K}]$$

any new ~~training~~ test sample y can be written as

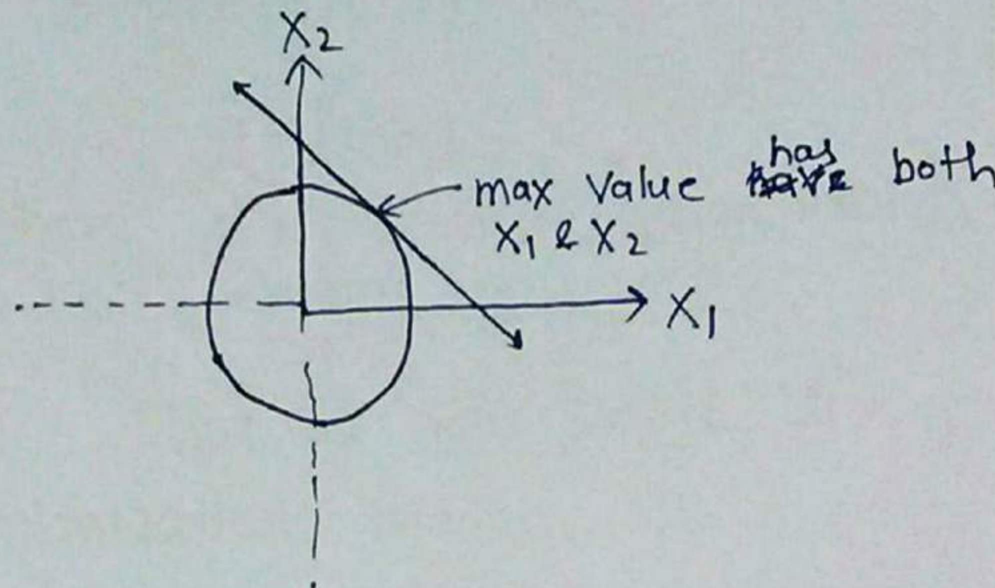
$$y = A x_0$$

x_0 → Coefficient vector whose entries are zero except those associated with ith class.

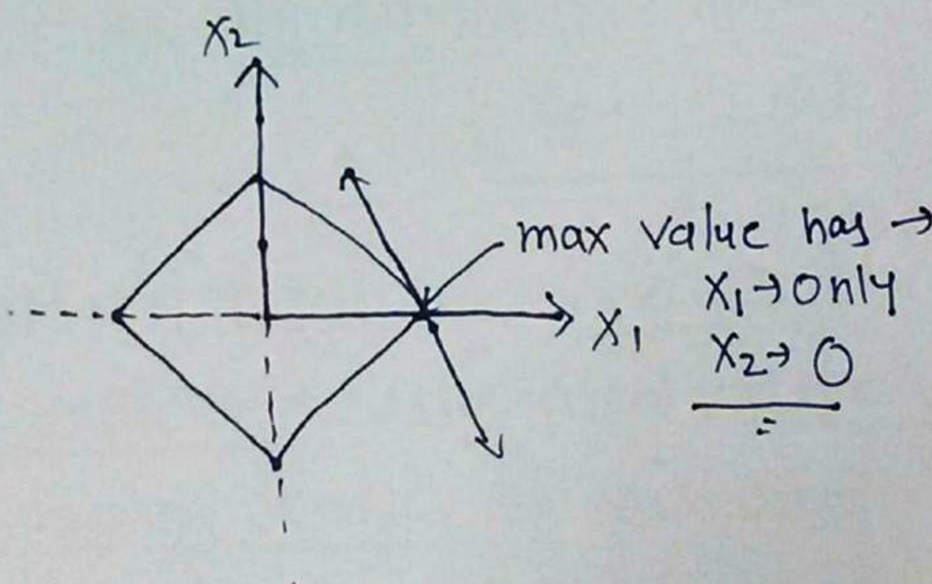
$$x_0 = [0, \ldots, 0, \alpha_{i,1}, \alpha_{i,2}, \ldots \alpha_{in}, 0 \ldots 0$$

→ Why l^1 norm is better compare to l^2 ? <u>To achieve</u> <u>sparsity</u>

Represention of l^2 in 2D



Representation of l^1 in 2D.



Form of l^1- minimization Problem

$$\hat{x}_1 = \arg\min \|x\|_1 \quad \text{subject to } Ax = y$$

To take care of noise

$$y = Ax_0 + z$$

where z is a noise term with bounded energy $\|z\|_2 < \varepsilon$

Thus, a new stable l' - minimization Problem is

$\cancel{\hat{x}_1 = \arg\max}$

$$\hat{x}_1 = \arg\min \|x\|_1 \quad \text{subject to } \|Ax - y\|_2 < \varepsilon$$

Classify y based on how well the coefficients associated with all training Samples of each object reproduce y

For each class i, let

$\delta_i : R^n \to R^n$ be a characteristic function

For $x \in R^n$, $\delta_i(x) \in R^n$

→ Vector → non zero entries are the entries in X → that are associated with class i.

→ Test sample y can be approximated as $\hat{y}_i = A\delta_i(\hat{x}_1)$.

→ Classify Y based on

$$\min_i r_i(y) = \|y - A\delta_i(\hat{x}_1)\|_2.$$

Classify y based on above residuals $r_i(y)$.

Sparse-Representation based Classification (SRC).

1 → Input :- a matrix of training samples $A = [A_1, A_2, \ldots A_k] \in R^{m \times n}$ for k classes, a test-sample $y \in R^m$, (and optional error tolerance $\varepsilon > 0$)

2 → Normalize the columns of A to have unit ℓ^2-norm.

3 → Solve the ℓ^1-minimization problem

$$\hat{x}_1 = \arg\min_x \|x\|_1 \quad \text{subject to } Ax = y.$$

(or)

$$\hat{x} = \arg\min_x \|x\|_1 \quad \text{subject to } \|Ax - y\|_2 \leq \varepsilon$$

4 → Compute the residuals $r_i(y) = \|y - A\delta_i(\hat{x}_1)\|_2$ for $i = 1, 2, \ldots k$

5 → Output → $y = \arg\min_i r_i(y)$.

Validation Based on Sparse Representation

Situation

A face recognition system, could be given a face image of subject that is not in the database or an image that is not a face at all.

Idea → or Solution

Concentration of Sparse Coefficients

Validity of test image

A valid test image should have a sparse representation whose non-zero entries ~~constr~~ concentrates mostly on one subject

Sparsity Concentration Index (SCI)

$$SCI(x) = \frac{\frac{k \cdot \max_i \|\delta_i(x)\|_1}{\|x\|_1} - 1}{k-1}$$

$\in [0,1]$

→ if $SCI(\hat{x}) = 1$, the test image is represented using images ~~of~~ from a single object.

→ if $SCI(\hat{x}) = 0$, sparse coefficient are spread evenly over all classes

choose Threshold τ and accept a test image as valid if $SCI(\hat{x}) \geqslant \tau$.

* → Residual measures how well the representation approximates the test image, and, SCI measures how good the representation itself

Feature Extraction

Advantage: Data dimension and computational cost can be reduced.

For, image size $640 \times 480 \rightarrow$ the size of m is in the order of 10^5.

$\rightarrow$ ASRC algorithm $\rightarrow$ not work in regular computers with such a high-resolution image.

Solution $\rightarrow$ Apply Feature Transformation

$$\tilde{y} = Ry = RAx_0$$

$R \rightarrow$ Matrix with size $d \times m$ with $d << m$, It represents Projection from image space to feature space.

Here, $d << n$, & $\bar{y} \in R^d$, & $x \in R^n$

equation

$$\hat{x}_1 = \arg\min \|x\|_1 \text{ subject to } \|RAx - \bar{y}\|_2 \leq \varepsilon$$

$RA \in \mathbb{R}^{d \times n}$ and,

y is replaced by its features $\bar{y}$

How to choose R?

Randomfaces → advantages → efficient to generate
→ R → independent of training set

Matrix R :- d → raws → each raw has size $1 \times m$

↳ Generate randomface of total size m

* ↳ entries are independently sampled from zero mean normal distribution

* → Normalize each raw with unit length.

Robustness to Occlusion or Corruption

In Practical Scenarios

y could be partially corrupted or occluded.

$$y = y_0 + e_0 = Ax_0 + e_0$$

$e_0 \in R^m$, → Vector of errors with fraction ρ, of its entries are nonzero, & unknown location information.

Fundamental Principle of Coding Theory

Redundancy in the measurement is essential to detecting & correcting Gross error

m is to large.

thus if fraction of pixels are completely corrupted by occlusion → recognition can be possible based on remaining pixels.

Note :-

Original images are more robust, redundant or informative than any of its representation.

So,

It would be better to work with heighest possible resolution when dealing with occlusion and corruption.

A New Representation

$$y = [A, I] \begin{bmatrix} x_0 \\ e_0 \end{bmatrix} = B w_0$$

$$B = [A, I] \in R^{m \times (n+m)}$$

$$\text{As } A \in R^{m \times n}$$

$$\& \; I \in R^{m \times m}$$

Thus, $y = Bw$ is always underdetermined and does not have unique solution.