classmate

Date 23.4.2020

Page 1

(3) OPTIMIZATION FORM 4

$$\text{minimize } \|X - \phi\phi^T X\|_F^2 \qquad X \in \mathbb{R}^{d\times n}, \quad \phi \in \mathbb{R}^{d\times d}$$

$\|\cdot\|_F$ → frobenius norm

$$\text{s.t.} \quad \phi^T \phi = I$$

here X is the matrix from the optimization problem
ϕ is the matrix consisting of colums of eigenvectors

the objective fn. can be simplified as:

$$\|X - \phi\phi^T X\|_F^2 = tr(X^T X - X X^T \phi\phi^T)$$

tr → trace of matrix.

its lagrangian is $(f(x) - \lambda g(x))$

$$\mathcal{L} = tr(X^T X) - tr(X X^T \phi\phi^T) - tr(\Lambda^T(\phi^T\phi - I))$$

Λ → eigen value matrix $\mathbb{R}^{d\times d}$

equating derivative of $\mathcal{L}$ to 0 gives:

$$\frac{\partial \mathcal{L}}{\partial \phi} = 2XX^T\phi - 2\phi\Lambda = 0$$

$$\Rightarrow XX^T\phi = \phi\Lambda \quad \Rightarrow \quad \underline{A\phi = \phi\Lambda}$$

→ eigen value problem in vector form.

this can be useful in machine learning, when we want to consider several PCA directions of dimensionality reduction

(4) FISHER DISCRIMINANT ANALYSIS

this is a dimensionality reduction technique which attempts to classify a set of points in reduced dimensions. it tries to keep the mean of both classes as separate as possible, while trying to minimize variance within a class.

objective: max. $\dfrac{w^T S_B w}{w^T S_w w}$ ← max. mean sep. / ← min. variance within.

w = projection surface

$$S_B = \sum_{j=1}^{c} (\mu_i - \mu_t)(\mu_i - \mu_t)^T$$

$$S_w = \sum_{j=1}^{c} \sum_{i=1}^{n_j} (x_{j,i} - \mu_i)(x_{j,i} - \mu_i)^T$$

this objective can be restated as:

$$\max.\ w^T S_B w$$
$$\text{st. } w^T S_w w = 1.$$

finding its lagrangian, and equating its derivative to 0 yields:

$$S_B w = \lambda S_w w$$

λ = lagrange multiplier

since this is now in generalized eigenvalue problem form, we can now one of the solution methods to obtain $\underline{\underline{w}}$. (projection plane).

(5) LLE PSEUDOCODE

A. select K nearest neighbours for each point
for each point X_i $i = 1 - N$

B. find reconstruction weights for each point
Z = matrix of all neighbours of X_i
Subtract X_i from every column of Z.
local covariance $C = Z^T Z$
solve $C w = 1$ for w
set $W_{ij} = 0$ if j is not neighbour of i.
remaining elements in row W_i normalized = $w/\text{sum}(w)$

C. map to target dimension (Y using weights W)
$M = (I - W)^T (I - W)$
find bottom $d+1$ eigenvectors of M.
(corresponding to $d+1$ smallest eigenvalues)
Set the i^{th} row of Y to be $i+1^{th}$ ~~sth~~ eigenvector
(discard eigenvector $[1, 1, 1, \ldots]$ with eigenvalue 0)

classmate

Date 23.4.2020
Page 4

(6) line : $x - y = 1$

100 points : (1,0) , (2,1) , (3,2) , ...

the output is computed in python.

covariance matrix :

$$\begin{bmatrix} 842 & 842 \\ 842 & 842 \end{bmatrix}$$

eigenvectors:

1. [0.707   0.707]   1683

2. [-0.707   0.707]   ~ 0

the main eigen vector is the first one which is parallel with respect to the slope line.

(eigen vectors are computed from the covariance matrix)

(7) (A) initialization : 626.1
(B) it. 1 : 16.5
(C) it. 2 : 16.5
3
conv.

(8) (A) if we increase K ⇒ we allow more classes so error can decrease further. agree.

(B) the max. value of K can be N (no. of points). then objective (error) would be 0.

(C) normalized objective (error) would decrease too. (but objective * K might not change).