~~29~~|3|2019

Dimensionality Reduction. and derivation of PCA

PCA : Principal Component Analysis. - Derivation

Initially I have F.V which is d, and want to reduce d'

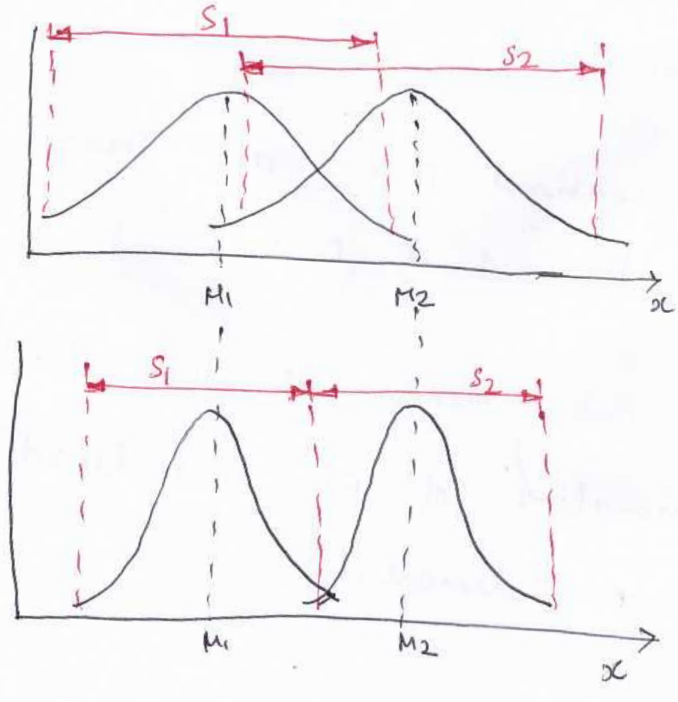
$$d \to d' \; ; \quad d' \le d$$

there are two popular approaches:

PCA: when we transform the dimension from d to d', after taking projection in the lower dimensional space (d') There should be minimum error between (d to d') original Feature vector (d) and the reduced feature vector (d')

The squared error between the original F.V (d) and the reduced F.V (d') that should be minimum.

MDA: Multiple | Linear Discriminant Analysis.



within class scatter is reduced (S_1 and S_2) whereas between Class scatter is increased

PCA:

Given

n - number of feature vector and Every feature vector is of dimension 'd'.

$X_1 \rightarrow [x_1, x_2, \ldots, x_d]$

X_2

$\vdots$

X_n

mean of these F.V $= m = \frac{1}{n} \sum_{K=1}^{n} X_K$ [sum all n F.V]

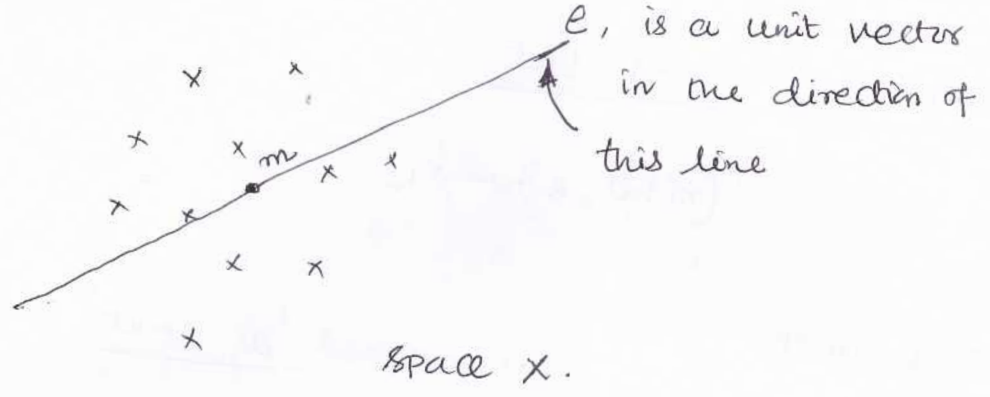
i) let me take an extreme case, that If we represent all d-dimensional F.V into 0-dimensional F.V (∵ i.e within F.V I don't have any variance) That is nothing but the mean vector. 'm'. which is a point. If

ii) I do that, I losses the variability within F.V. which is not good.

iii) Instead of zero-dimensional F.V; all these d-dimensional F.V is represented by 1-d F.V.

iv) I still have the number of F.V which is 'n' but the dimensionality of F.V is reduced from d to d' (here it is one-dimension).

A vector is getting mapped to a scalar. [a point on the line].



$y = mx + c$

eqn of the line $\quad X = m + ae$

$X = m \quad$ If $a = 0$

$X = m + e \quad$ If $a = 1$

$X = m + 2e \quad$ If $a = 2$.

for different values of 'a' I'm moving along the line in the direction of 'e'.

$\boxed{a = \text{positions of different points on the line.}}$

Given this, ~~a point~~ I can represent a point x_k

$X = m + a e$

$x_k \approx m + a_k e$.

* a_k represents, a point x_k where it is mapped on to this line 'e'.
* while doing this, there may be some error introduced.

because x_K is d-dimension and I'm representing it 'a point' on the line.

The error is nothing but.

$$(m + a_K e) - x_K.$$

x_K is a vector $\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_d \end{bmatrix}$ is mapped 'a point' on a line.

The error is $\left[(m + a_K e) - x_K\right]$

PCA tries to minimize the sum of squared error in terms of 'e'. [e is fixed]

I can define the error function (or) criteria function.

$$J(a_1, a_2, \ldots, a_n, \underset{\substack{\uparrow \\ \text{direction of line}}}{e}) = \overbrace{\sum_{K=1}^{n} \underbrace{\left\| (m + a_K e) - x_K \right\|^2}_{\text{squared error}}}^{\text{Sum of squared error}}$$

$$J(a_1, a_2, \ldots, a_n, e) = \sum_{k=1}^{n} \| \underbrace{(m + a_k e)}_{\text{reduced F.V}} - \underbrace{x_k}_{\text{original F.V}} \|^2$$

rewritten the same exp

$$= \sum_{k=1}^{n} \| \underbrace{a_k e}_{a} - \underbrace{(x_k - m)}_{b} \|^2$$

$(a-b)^2 = a^2 - 2ab + b^2$

$$= \sum a_k^2 \|e\|^2 - 2 \sum a_k e^t (x_k - m) + \sum \|x_k - m\|^2 \quad - \text{①}$$

I want to reduce this sum of squared error by varying the value of a [what I have to do is that; take differentiation w.r.t 'a' equate that for = 0] ⇒ minimization problem.

Solve for a?

$$\frac{\partial J}{\partial a} = 0 \; ; \quad \|e\| = 1.$$

$$\frac{\partial J}{\partial a} = 2 \sum a_k - 2 \sum e^t (x_k - m) + 0$$

$$\Downarrow$$

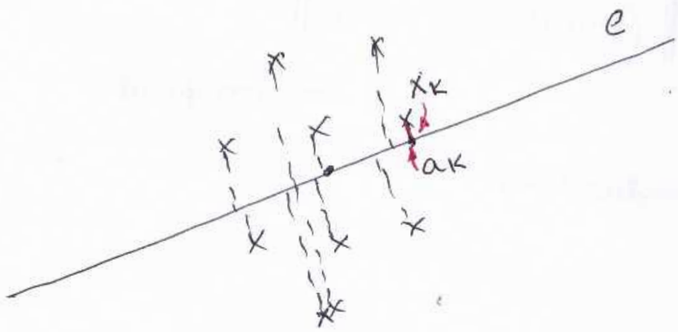
$$0 = 2 \sum a_k - 2 \sum e^t (x_k - m)$$

$$\cancel{2} \sum a_k = \cancel{2} \sum e^t (x_k - m)$$

$$\boxed{a_k = e^t (x_k - m)}$$

unit vector in the direction of line

a_k is nothing but orthogonal projection of x_k on to a line in the direction of 'e' passing through mean vector 'm'.



Recall: $\boxed{\text{Lagrange optimization}}$ ⇒ from Duda Harat book, at back side.

Suppose we seek the position x_0 of an extremum of a scalar valued function $f(x)$, subject to some constraint. If a constraint can be expressed in the form $g(x) = 0$, then we can find the extremum of $f(x)$ as follows. First we form the lagrangian function

$$L(x, \lambda) = f(x) + \lambda \underbrace{g(x)}_{=0}$$

scalar lagrange undetermined multiplier.

We convert this constrained optimization problem into an unconstrained problem by taking the derivative,

$$\frac{\partial L(x, \lambda)}{\partial x} = \frac{\partial f(x)}{\partial x} + \lambda \frac{\partial g(x)}{\partial (x)} = 0$$

and using standard methods from calculus to solve the resulting equations for λ and extremizing value of x.

PCA - contd...

which given line is best? which particular direction of a line is considered to be best?

Substitute $\boxed{a_k = e^t (x_k - m)}$ in (1) (Page 3)

$$= \sum_{i=1}^{n} a_k^2 \underbrace{\|e\|^2}_{1} - 2 \sum a_k \underbrace{e^t (x_k - m)}_{a_k} + \sum \|x_k - m\|^2$$

$$J(e) = \underbrace{\sum a_k^2 - 2 \sum a_k^2} + \sum \| x_k - m \|^2$$

direction of the line

$$= - \sum a_k^2 + \sum \| x_k - m\|^2$$

$$= - \sum \left[e^t (x_k - m) \right]^2 + \sum \| x_k - m \|^2$$

$$= - \sum \left[e^t (x_k - m)(x_k - m)^t e \right] + \sum \| x_k - m \|^2$$

$$= - e^t \left[\underbrace{\sum (x_k - m)(x_k - m)^t} \right] e + \sum \| x_k - m \|^2$$

$$= - e^t S e + \sum \|x_k - m\|^2$$

$$\boxed{S = \sum_{k=1}^{n} (x_k - m)(x_k - m)^t}$$ $\Rightarrow$ scatter matrix, S

Scatter matrix S, it simply represents how the data is spread in the space.

Covariance matrix 'Σ':

$$\Sigma = \frac{1}{n} \sum_{K=1}^{n} (X_K - m)(X_K - m)^t$$

$\boxed{\text{scatter matrix is scaled version of } \Sigma \times n = S}$

Relation between Scatter matrix 'S' and Co-variance matrix Σ.

minimize $J(e)$ $\rightarrow$ $J(e) = \underbrace{-e^t S e}_{} + \underbrace{\sum \|x_K - m\|^2}_{\text{independent of 'e'}}$

unit vector in the direction of line 'e'

$\Downarrow$

maximization of $e^t S e$

This can be done

$\boxed{\text{Aim is to minimize } J(e) \text{ by varying 'e' by maximization of } e^t S e \text{, by negating } -e^t S e \; J(e) \text{ is minimized.}}$

This can be done by using Lagrangian optimization:

$u = e^t S e - \lambda \underbrace{(e^t e - 1)}_{}$ subject to the Constraint $\|e\| = 1$.

$\|e\| = 1$

$e \cdot e^t = 1$

to maximize $e^t S e$, we should take differentiation w.r.t 'e'

$\Downarrow$

$(e^t \cdot e - 1)$ $\qquad \|e\| = \sqrt{e_1^2 + e_2^2}$

$$u = e^t se - \lambda(e^t e - 1)$$

$$\frac{\partial u}{\partial e} = \underbrace{e^t se} - \lambda e^t e + \lambda \qquad e^t e = e^2$$

$$\frac{\partial u}{\partial e} = e^2 s - \lambda e^2 + \lambda \leftarrow \text{constant}$$

$$\frac{\partial u}{\partial e} = 2es - 2\lambda e + 0$$

$$0 = 2es - 2\lambda e$$

$$\cancel{2} es = \cancel{2} \lambda e$$

$$\boxed{Se = \lambda e} \Rightarrow \text{eigen value expression.} \quad -(2)$$

eigenvalue of 'S'   eigenvector of 'S'

But, our aim is to maximize $e^t \underline{se}$ from (2)

$$e^t \lambda \underline{e} \quad \text{maximize } \lambda e$$

$$\lambda \underline{e^t e} \qquad e^t e = \|e\| = 1.$$

If we want to maximize $e^t \underline{se}$, we should maximize $e^t \lambda e$. where 'e' is the eigen vector of matrix 'S'. so, simply choose the eigen vector whose eigen value is maximum,

(or) choose the eigen vector corresponding to the maximum eigen value.

by using this whatever value of a_K that is what represent the points is called pricipal component.

Hence it is called PCA. Also, Known as KL Transformation

This result can be readily extended from a one-dimensional to d' dimensional

$$X_K = m + \sum_{K=1}^{d'} a_K e_K$$

Point X_K in d' dimension

Choose the eigenvectors corresponding to first 'K' eigenvalues. largest

Lagrangian optimization (from wikipedia)

Gradient alignment between the target function and the constraint function

- $\nabla f(x,y) = \lambda \nabla g(x,y)$
- The constraint itself $g(x,y) = c$.

$$L(x,y,\dots,\lambda) = f(x,y\dots) - \lambda[g(x,y,\dots) - c]$$

L ↓ Lagrangian

λ ↳ lagrange multiplier

when we want to maximize a multivariable function $f(x,y,\dots)$ subject to the constraint that another multivariable function equals a constant $g(x,y,\dots) = c$ follow these steps, take gradient of $L(x,y,\dots,\lambda)$ and equate it to zero.

$$\nabla L(x,y,\dots,\lambda) = 0$$

Here

$f(x,y) = \underline{e^t s e}$

$g(x,y) = e = \underline{e^t e}$ and the constraint is

$g(x,y)$ is constant $\|e\| = 1$.

$$\nabla L(x,y,\dots,\lambda) = e^t s e - [\lambda(e^t e) - 1]$$

$$\boxed{L(x,y,\dots,\lambda) = f(x,y\dots) - \lambda[g(x,y,\dots) - c]}$$

$e^t s e - \lambda \|e\|$ $\qquad \|e\| = e^t$

from this,

$$\nabla \left[L(x, y, \dots, \lambda) \right] = e^t S e - \left[\lambda \left(e^t e - 1 \right) \right]$$

Now take gradient

$$\nabla \left[L(x, y, \dots, \lambda) \right] = e^t S e - \lambda e^t e + \lambda$$

$$= e^t S e - \lambda e^t e + \lambda \quad \left[e^t e = e^2 \right]$$

$$\frac{\partial u}{\partial e} = S e^2 - \lambda e^2 + \lambda$$

$$\frac{\partial u}{\partial e} = 2 S e - 2 \lambda e + 0$$

$$\Downarrow$$

$$0 = 2 S e - 2 \lambda e$$

$$\cancel{2} S e = \cancel{2} \lambda e$$

$$\boxed{S e = \lambda e}$$ $\Rightarrow$ eigenvalue expression.

LAGRANGE MULTIPLIERS

- To find the optimal solution for a given constraint.

Step 1 :-

Form the Lagrange function $L(x)$,

for the function to be optimized $f(x)$ &

the constraint function : $g(x) = c$

$$\therefore L(x) = f(x) - \lambda[g(x) - c]$$

Step 2 :-

Apply the partial / derivative. &

equate to zero.

Step 3 :-

Solve for λ.

Choose the value of x that optimizes λ.

Let the value be x_{opt}

Put x_{opt} in $f(x)$.

$f(x_{opt})$ will be the optimal solution.

We have formulated the cost / error function

$$J_1(e) = -e^t S e + \sum_{k=1}^{n} ||x_k - m||^2 .$$

Using Lagrange's Optimization,

Here, $f(x) = -e^t S e$ (function to be optimized)

$g(x) = ||e|| = 1$ (constraint function)

$(e^t . e - 1) = 0$

$\therefore \quad f(x) - \lambda(g(x)$

$u = e^t S e - \lambda(e^t . e - 1)$

Taking the partial of derivative :

$\frac{\partial u}{\partial e} \Rightarrow u = e^2 S - \lambda(e^2 - 1)$

$$\frac{\partial u}{\partial e} = 2eS - 2\lambda e + 0 .$$

Equate to zero;

$$\frac{\partial u}{\partial e} = 0;$$

$$\cancel{2} e S = \cancel{2} \lambda e .$$

$$\boxed{Se = \lambda e}$$

//