PCA Basic Meaning

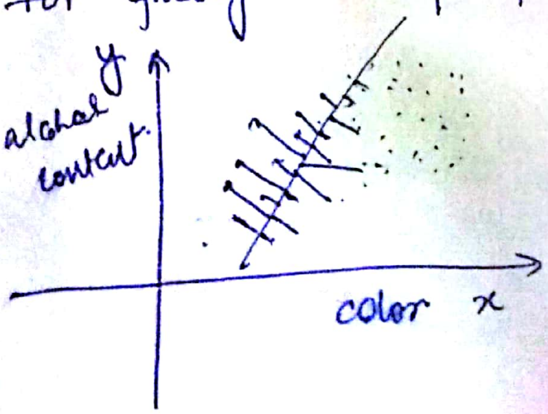
→ PCA is method of summarizing some data.
for. example some Perfume ~~bottle~~ bottles are there, It has many properties like color, strong smell, ~~age~~ ~~etc.~~ longetivity etc.
We can have a whole list of characterstics for each perfume But many of them will measure related properties & will be redundant. Thus PCA summarize each perfume with less characterstics.

→ It is not selecting some characterstics & discarding others. Instead it constructs new characterstics that turn out to summarize our list of perfumes well.
e.g. a new characterstics may be computed as perfume's longetivity minus its pH level or some other linear combination.

→ PCA will look for properties that show as much as variation across ~~items~~ perfumes as possible instead of constructing a property which make them look similar

→ New properties will be such that old characterstics can be reconstructed.

→ ~~For finding the principal components~~



We can construct a new property by drawing a line through the point cloud and projecting all points onto the line, This new property will be given by $w_1 x + w_2 y$ (linear combination)

PCA will find best line acc to diff criteria of what is the best. ~~fast~~
→ Variation of values along this line should be maximal i.e. Covariance

→ Second, If we reconstruct the original two characterstics from new one, the reconstruction error will be given by length of the connecting lines. i.e. minimum error.
Both these properties gives the same quantity ie. variance.

$$\begin{pmatrix} 1.07 & 0.63 \\ 0.63 & 0.64 \end{pmatrix}$$

As it is a square symmetric matrix, it can be diaganized by choosing a new orthogonal coordinate system given by its Eigen vectors

∴ In new coordinate system. covariance matrix is diagnol & looks like

$$\begin{pmatrix} 1.52 & 0 \\ 0 & 0.19 \end{pmatrix}$$

Thus there comes the role of eigen values & eigen vectors.

Thus we need to know what is projection, eigen vectors & variance.

Thus PCA creates new, uncorrelated features that are linear combination of the original features.

Principal Components Analysis (PCA)

Input data is m-dimensional vector. Sometimes dimension of input data is more than required so we go for dimension reduction by finding redundancy.

$$\vec{x} = (\underbrace{x_1, x_2, x_3 \dots x_{m_0}}_{\text{Significant}} \mid \overbrace{x_{m_0+1} \dots x_m}^{\text{insignificant}})$$

(m is very large, we want to reduce it to m_0)

Error on chopping $(m-m_0)$ dimensions (MSE, if dimension is reduced without considering the significance of these ~~dimension~~ features)

[MSE = Sum of variance of the eliminated features]

If MSE is small enough, we can afford dimension reduction

PCA is
→ dimension reduction technique
→ developed by Hotelling ~~H~~ in 1933.
→ - Lower dimension
- Orthogonality of transformed dimension X_1, X_2,

Correlation Matrix

$\vec{X}$ m-dimensional

$\vec{T}\,\vec{X} \rightarrow \vec{X}_{new}$

$m \times m \quad m \times 1 \quad [m \times 1]$

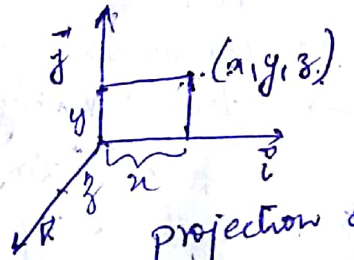
In the transformed space, Out of m points, m_0 points are significant & rest points are insignificant. Thus, we need to design this transformation.

$\vec{X}$: m-dimensional random vector

$E[\vec{X}] = 0$ [Assumption] [Not always they so we make it such random vector by substracting the mean from the elements]

$\vec{q}$: m-dimensional unit vector

Similarly we are going to project m dimensional $\vec{X}$ into m dimensional unit vector space



projection of point in 3 d unit vector space

Projection of $\vec{X}$ onto $\vec{q}$.

$$A = \vec{X}^T \cdot \vec{q} = \vec{q}^T \vec{X}$$

subject to $\boxed{\|\vec{q}\| = (\vec{q}^T \vec{q})^{1/2} = 1}$

norm of $\vec{q}$

In the projected space, we need to determine variance & expectation

$$E[A] = \vec{q}^T E[\vec{X}] = 0 \qquad \sigma^2 = E[A^2] - E[A]^2$$

$$\sigma^2 = E[A^2] = E[(\vec{q}^T\vec{X})(\vec{X}^T\vec{q})]$$

$$= \vec{q}^T \underbrace{E[\vec{X}\vec{X}^T]}_{\text{correlation Matrix of } \vec{X}\ (m \times m)} \vec{q}$$

$\vec{R}$ is correlation or Variance

$$= \vec{q}^T \vec{R}\, \vec{q}$$

[$\vec{R}$ is dependent on feature, we need to find $\vec{q}$ intelligently]

Also, Correlation Matrix is symmetric

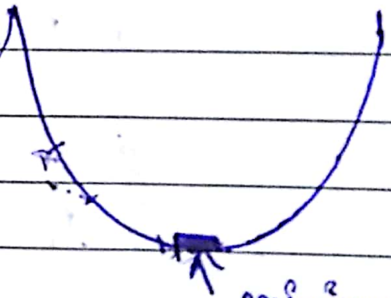
$$\sigma^2 = \psi(\vec{q}) = \vec{q}^T \vec{R} \vec{q} \quad \ldots\ldots (1)$$

we have to minimize $\psi(\vec{q})$ – variance probe
We need to find extreme position of Variance probe (minima).

For ~~minimal~~ extremal value of the variance

$$\psi(\vec{q} + \delta\vec{q}) = \psi(\vec{q}) \quad \ldots (2)$$

this means if we peturb the value of $\vec{q}$, value of $\psi(\vec{q})$ will not change much.



$$\psi(\vec{q} + \delta\vec{q}) = (\vec{q} + \delta\vec{q})^T \vec{R} (\vec{q} + \delta\vec{q})$$

$$= \cancel{\vec{q}^T \vec{R} \vec{q} + 2(\delta\vec{q})^T \vec{R} \vec{q} + (\delta\vec{q})^T \vec{R} \delta\vec{q}}$$

$$= \cancel{\vec{q}^T \delta\vec{q} \vec{R} (\vec{q} + \delta\vec{q})}, \quad \cancel{\vec{q}\, \delta\vec{q}^T \vec{R} (\vec{q})}$$

$$= \cancel{\vec{q}\, \delta\vec{q}^T \vec{R} \vec{q} + \vec{q}^T \delta\vec{q} \vec{R} \delta\vec{q}}$$

$$= (\vec{q} + \delta\vec{q})^T (\vec{R}\vec{q} + \vec{R}\delta\vec{q})$$

$$= \vec{q}^T \vec{R} \vec{q} + \delta\vec{q}^T \vec{R} \vec{q} + \vec{q} \vec{R} \delta\vec{q} + \delta\vec{q}^T \vec{R} \delta\vec{q}$$

$$= \vec{q}^T \vec{R} \vec{q} + 2\delta\vec{q} \vec{R} \vec{q} + \delta\vec{q} \vec{R}^T (\delta\vec{q}) \quad \ldots (3)$$

दिनांक Using Eq (1) & (2) in (3) अध्ययन केन्द्र, उत्तर

$$(\delta \vec{q})^T \vec{R}\, \vec{q} = 0 \quad \ldots\ldots (4)$$

Since $\vec{q}$ is unit vector $\|\vec{q} + \delta \vec{q}\| = 1$

or equivalently

$$(\vec{q} + \delta \vec{q})^T (q + \delta q) = 1$$

$$\vec{q}^T q + \delta \vec{q}^T q + \vec{q}^T \delta q + \delta q^T \delta q = 1$$

$$\vec{q}^T q + 2\, \delta \vec{q}^T \delta q + \underbrace{\delta q^T \delta q}_{\text{negligible}} = 1$$

$$\cancel{1} + 2\delta \vec{q}^T \delta q = \cancel{1}$$

$$(\delta \vec{q})^T \vec{q} = 0 \quad \ldots\ldots (5)$$

This means $(\delta \vec{q})$ is orthogonal to $(\vec{q})$ thus we can just change the direction of unit vector.

To combine (4) & (5), we need some quantity which has dimension of that $\vec{R}$.
λ has dimension $m \times m$.

$$(\delta \vec{q})^T \vec{R} \vec{q} - \lambda (\delta q)^T \vec{q} = 0$$

$$(\delta q)^T [\vec{R} \vec{q} - \lambda \vec{q}] = 0$$

$$\therefore \boxed{\vec{R} \vec{q} = \lambda \vec{q}} \quad \ldots\ldots (6)$$

given a matrix $\vec{R}$ we have to find solution
λ is eigen value of the matrix $\vec{R}$ $(m \times m)$
Thus we have m eigen value.

Eigen Values: $\lambda_1, \lambda_2 \ldots\ldots \lambda_m$ & the associated values of $\vec{q}$
Eigen Vectors: $\vec{q}_1, \vec{q}_2, \ldots\ldots \vec{q}_m$
of $\vec{R}$

$$\vec{R} \vec{q}_j = \lambda_j \vec{q}_j \quad \text{for } j = 1, 2, 3 \ldots m \quad \ldots\ldots (7)$$

Let us assume $\lambda_1 > \lambda_2 > \ldots\ldots > \lambda_j > \ldots \lambda_m$
$\lambda_1 = \lambda_{max}$

Let $\vec{Q} = [\vec{q}_1 \, \vec{q}_2 \ldots, \vec{q}_j \ldots \vec{q}_m]$

$$\vec{R} \vec{Q} = \vec{Q} \vec{\Lambda} \quad \ldots\ldots (8)$$

where $\vec{\Lambda} = \text{diag}[\lambda_1 \cdot \lambda_2 \ldots \lambda_m]$

The matrix $\vec{Q}$ is an orthogonal matrix, satisfying

$$\vec{q_i^T} \cdot q_j = \begin{cases} 1 & j = i \\ 0 & j \neq i \end{cases}$$

Thus. $\vec{Q^T} \vec{Q} = \vec{I}$

$\Rightarrow \quad \vec{Q^T} = \vec{Q}^{-1}$

Multiplying $\vec{Q^T}$ on both sides of eqn (8)

$$\boxed{\vec{Q^T} \vec{R} \vec{Q} = \vec{\Delta}} \quad \ldots\ldots (9)$$

In Expanded form,

$$\vec{q_j^T} \vec{R} \vec{q_k} = \begin{cases} \lambda_j & k = j \\ 0 & k \neq j \end{cases}$$

we can also write $\vec{R}$ as

$$\vec{R} = \vec{Q} \vec{\Delta} \vec{Q^T}$$

Thus

$$\vec{R} = \sum_{i=1}^{m} \lambda_i \vec{q_i} \vec{q_i^T} \quad \ldots\ldots (10)$$

Scanned by CamScanner

from Eq (1),

$$\boxed{\psi(q_j) = \lambda_j} \quad \dots\dots (11) \quad j = 1, 2 \dots m$$

If we arrange the eigen values in decreasing orders, ~~so~~ ~~th~~ we will be eliminating the elements with low variance.

Actually we are going to project $\vec{X}$ onto the unit ~~q~~ vector $\vec{q}_j$.

$$a_j = \vec{q}_j^T \vec{X} = \vec{X}^T \vec{q}_j \qquad j = 1 \dots m \quad -(12)$$

As. a_j : projection of $\vec{x}$ on-to principal directions as represented by $\vec{q}_j$.

a_j : principal Component

we have m principal components

Eq^n (12) represents ~~analysis~~ analysis Equation

In synthesis, from a_j, we should ~~be able to~~ recover $\vec{X}$.

Define projection vector $\vec{A} = \{a_1, a_2 \dots\dots a_m\}^T$

$$\{\vec{X}^T \vec{q}_1 \quad \vec{X}^T \vec{q}_2 \quad \dots\dots \quad \vec{X}^T \vec{q}_m\}$$

$$\vec{A} = \vec{Q}^T \vec{X} \quad \ldots \; 12(a)$$

pre-multiply both sides by $\vec{Q}$

$$\boxed{\vec{Q}\vec{A} = \vec{X}}$$

$$\vec{X} = \sum_{j=1}^{m} a_j \vec{q_j} \quad \ldots \; (13)$$

Eqn (13) represents synthesis Equation
q_j's are the basis vectors for synthesis.

PCA is also called Karhunen-Love Transformation
or K-L Transformation.

For Images, it is called Hotelling Transfo

$\lambda_1, \lambda_2 \ldots \lambda_l$ are the largest l eigen values of the correlation matrix $\vec{R}$.

$$\hat{\vec{x}} = \sum_{j=1}^{l} a_j \vec{q}_j$$

$$= \begin{bmatrix} \vec{q}_1 & \vec{q}_2 & \ldots & \vec{q}_l \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_l \end{bmatrix} \qquad l \leq m \qquad \ldots (14)$$

$$\begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_l \end{bmatrix} = \begin{bmatrix} \vec{q}_1^{\,T} \\ \vec{q}_2^{\,T} \\ \vdots \\ \vec{q}_l^{\,T} \end{bmatrix} \vec{x} \qquad l \leq m \qquad \ldots (15)$$

$$\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_m \end{bmatrix} \Rightarrow \begin{bmatrix} \vec{q}_1^{\,T} \\ \vec{q}_2^{\,T} \\ \vdots \\ \vec{q}_l \end{bmatrix} \Rightarrow \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_l \end{bmatrix}$$

// Encoder maps $\mathbb{R}^m \rightarrow \mathbb{R}^l$

दिनांक:

$$\begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_l \end{bmatrix} \Rightarrow \begin{bmatrix} \vec{q}_1 & \vec{q}_2 & \cdots & \vec{q}_l \end{bmatrix} \Rightarrow \begin{bmatrix} \hat{x}_1 \\ \hat{x}_2 \\ \vdots \\ \hat{x}_m \end{bmatrix}$$

approximated solⁿ.

<u>Decoder</u>

Error vector

$$\vec{e} = \vec{X} - \hat{X} \quad \ldots \quad (16)$$

Substituting Eq(13) & Eq.(14)

$$= \sum_{j=1}^{m} a_j \vec{q}_j - \sum_{j=1}^{l} a_j \vec{q}_j \qquad l \leq m$$

$$= \sum_{j=l+1}^{m} a_j \vec{q}_j$$

Lets take

$$\vec{e} \cdot \vec{\hat{X}} = \sum_{i=l+1}^{m} a_i \vec{q}_i \cdot \sum_{j=1}^{l} a_j \vec{q}_j$$

Scanned by CamScanner

दिनांक:

$$\sum_{i=l+1}^{m}\sum_{j=1}^{l} a_i \cdot a_j \; \vec{q_i} \cdot \vec{q_j}$$

Since $i \neq j$.

$\therefore$

$= \vec{0}$

Total Variance of the m-component of data vector $\vec{X}$

$$\sum_{j=1}^{m} \sigma_j^2 = \sum_{j=1}^{m} \lambda_j$$

Total Variance of l-elements of $\vec{X}$.

$$\sum_{j=1}^{l} \sigma_j^2 = \sum_{j=1}^{l} \lambda_j$$

Total Variance of the $(m-l)$ elements

$$\sum_{j=l+1}^{m} \sigma_j^2 = \sum_{j=l+1}^{m} \lambda_j$$

We are eliminating smallest variance strokes.

भारत पेट्रोलियम

PCA considers a set of N sample images $\{x_1, x_2, \ldots, x_N\}$ taking values in n-dimensional image space

Each image belongs to one of c classes. $\{X_1, X_2, X_3 \ldots X_c\}$

~~N~~ samples of m individuals.