ARTIFICIAL NEURAL NETWORK

Chinmayananda A.
Department of Electronics and Communication Engineering

OUTLINE

- ***Principal Component Analysis (PCA)***
  - *Eigenstructure of PCA*
  - *Basic Data Representations*
  - *Dimensionality Reduction*
  - *Principle of Orthogonality*

Chinmayananda A.

Department of Electronics and Communication Engineering

PRINCPAL COMPONENT ANALYSIS

Chinmayananda A.
Department of Electronics and Communication Engineering

- Problem : Given the data set which are realizations of a random vector X, we want to find suitable orthogonal coordinates along which the variances of the projections are significant (large).
- Formalisation :
  - Let X be an $m-$dimensional data vector (column vector).
  - We assume that this vector has zero mean.
  - Else, cancel the mean, and continue the analysis.
  - If $X = [X_1\ X_2\ X_3 \dots X_m]^T$, then $E[X] = 0$.
    - That is , $E[X_i] = 0,\ i \in \{1,2,\dots,m\}$.

PRINCIPAL COMPONENT ANALYSIS (PCA)

Problem Formulation : Eigenstructure of PCA

PES UNIVERSITY ONLINE

- Let q be a unit vector, and A be the projection of X on q.
- That is , $||q|| = (q^T q)^{1/2} = 1$, $A = X^T q = q^T X$.
- A is a random variable with zero mean and variance σ^2.
- $E[A] = E[X^T q] = E[\sum_{i=1}^{m} X_i q_i] = \sum_{i=1}^{m} q_i E[X_i] = 0.$
- $\sigma^2 = E[(A - E[A])^2] = E[A^2] = E[(q^T X)(X^T q)].$
- $\Rightarrow \sigma^2 = q^T E[XX^T]q = q^T Rq.$
- The matrix R is called the correlation matrix of vector X.
- $R^T = (E[XX^T])^T = E[(X^T)^T X^T] = E[XX^T] = R$ – Symmetric.
- The variance is dependent on the vector q and denoted as $\psi(q)$.

PRINCIPAL COMPONENT ANALYSIS (PCA)

Problem Formulation (Continued)

- That is, $\psi(q) = \sigma^2 = q^T R q -$ also called the variance probe.
- <u>Problem</u> : Find unit vectors q along which the variance probe has extremal or stationary values (local maxima or minima).
- For any small perturbation δq of q, we can approximate the variance probe to first order in δq as $\psi(q + \delta q) = \psi(q)$.
- Use Taylor series to get the above approximation.
- $\psi(q + \delta q) = (q + \delta q)^T R(q + \delta q) = q^T Rq + 2(\delta q)^T Rq + (\delta q)^T R\delta q.$
- Neglecting the quadratic term in δq we get the following.
- $\psi(q + \delta q) = \psi(q) + 2(\delta q)^T R\, q.$

PRINCIPAL COMPONENT ANALYSIS (PCA)

Problem Formulation (Continued)

- $\psi(q+\delta q) = \psi(q) + 2(\delta q)^T R\, q.$
- Comparing with $\psi(q+\delta q) = \psi(q)$, we have $(\delta q)^T Rq = 0.$
- Not all perturbations are allowed, as $||q + \delta q|| = 1.$
- $\Rightarrow (q+\delta q)^T (q + \delta q) = 1 \Rightarrow (\delta q)^T q = 0.$
- $\Rightarrow$ Not all perturbations are allowed, as $\delta q \perp q$.
- Equations $(\delta q)^T Rq = 0$ and $(\delta q)^T q = 0$ can be combined as below.
- $(\delta q)^T Rq - \lambda(\delta q)^T q = (\delta q)^T (Rq - \lambda q) = 0.$
- It is necessary and sufficient to have $Rq = \lambda q.$
- This is same as the Eigenvalue Problem in Linear Algebra.

PRINCIPLE COMPONENT ANALYSIS (PCA)

Eigenvalue Decomposition of a Correlation matrix (Recap)

- For an $m \times m$ correlation matrix R, assume that $\lambda_1 > \lambda_2 > \ \ldots > \lambda_m$, with the corresponding eigenvectors $q_1, q_2, \ldots, q_m$.
- Let $Q = [q_1, q_2, \ldots, q_j, \ldots q_m]$, and $\Lambda = diag[\lambda_1, \lambda_2, \ldots, \lambda_j, \ldots \lambda_m]$.
- The m equations $Rq_j = \lambda_j q_j$ can be written as $RQ = Q\Lambda$.
- $\Rightarrow Q^T Q = I \Rightarrow Q^T = Q^{-1}$.
- $\Rightarrow RQ = Q\Lambda$ is same as $Q^{-1}RQ = \Lambda \Rightarrow Q^T RQ = \Lambda$.
- $\Rightarrow q_j^T \ Rq_k = \lambda_j$, if $k = j$, and $q_j^T Rq_k = 0$ otherwise.
- $\Rightarrow R = \sum_{i=1}^{m} \lambda_i \ q_i \ q_i^T$
- Note that $\psi(q_i) = q_i^T Rq_i = \lambda_i$ (from Spectral decomposition).

PRINCIPLE COMPONENT ANALYSIS (PCA)

Eigenstructure of PCA and Dimensionality reduction

- Eigenvectors of the correlation matrix R of the mean-cancelled data vector X define the directions along which the variance probes have their extremal values, and the associated eigenvalues give the extreme values of the variance probes $\psi(q_i) = q_i^T R q_i = \lambda_i$.
- Dimensionality reduction procedure for a given data set :
  - Compute eigenvectors and eigenvalues of the correlation matrix.
  - Project the data onto the subspace spanned by the eigenvectors corresponding to the dominant eigenvalues.
- Also called subspace decomposition.

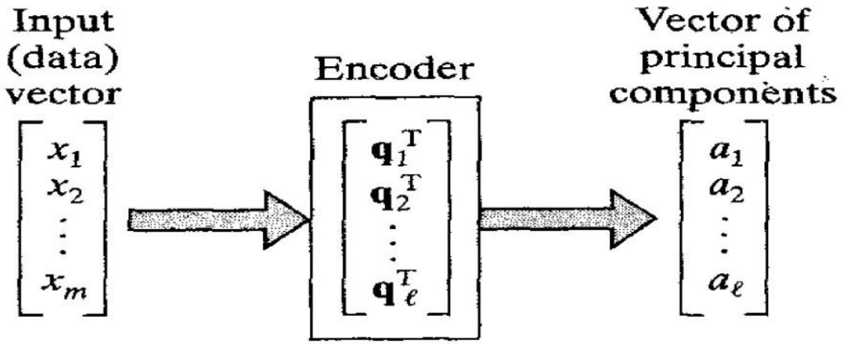
PRINCIPLE COMPONENT ANALYSIS (PCA)

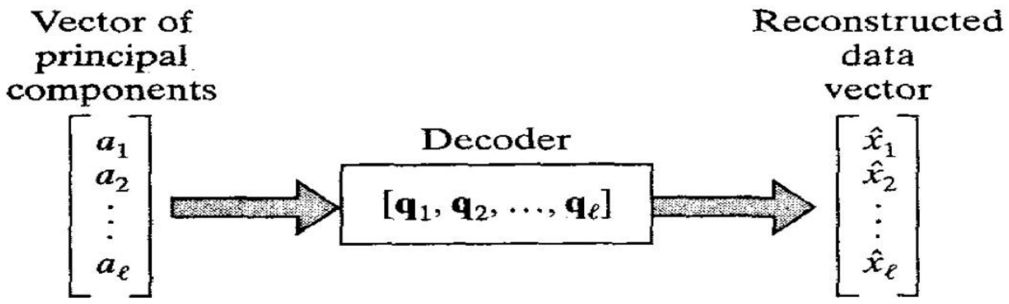
Basic Data Representation

- Let the data vector x denote a realization of the random vector X.
- The m projections of x given by $a_j = q_j^T x$, with $j = 1,2,\dots,m$ are called principal components – Analysis Equation.
- Reconstruction of x from the projections $a = [a_1, a_2, \dots, a_m]^T$.
- Note $[a_1, a_2, \dots, a_m]^T = [x^T q_1, x^T q_2, \dots, x^T q_m]^T = Q^T x$.
- $\Rightarrow x = Qa$ - Synthesis Equation.
- Unit vectors q_j represent a new basis of the data space.
- Data vector x is transformed into a point a in the feature space.

Dimensionality Reduction

- PCA gives a technique for dimensionality reduction of the data.
- Let $\lambda_1, \lambda_2, \dots, \lambda_l$ denote the l largest eigenvalues of the correlation matrix R.
- By truncating the synthesis equation to l terms, we get $\hat{x} = \sum_{j=1}^{l} a_j q_j$.
- Given the original data vector, we get $[a_1, a_2, \dots, a_l]^T = Q^T x$.
- Hence, these two operations can be seen as encoding and decoding the data.





PRINCIPLE COMPONENT ANALYSIS (PCA)

Principle of Orthogonality

- Approximation error vector - $e = x - \hat{x} = \sum_{j=l+1}^{m} a_j q_j$.
- Note that $e^T \hat{x} = \sum_{i=l+1}^{m} a_i q_i^T \sum_{j=1}^{l} a_j q_j = \sum_{i=l+1}^{m} \sum_{j=1}^{l} a_i a_j q_i^T q_j = 0$.
- Principle of Orthogonality : Error vector is orthogonal to $\hat{x}$.
- Note that variance of the j^{th} component of x is given by $\psi(q_j) = q_j^T R q_j$.
- Also, from eigenvalue decomposition $q_j^T R q_j = \lambda_j$.
- Total variance of components of $x = \sum_{j=1}^{m} \sigma_j^2 = \sum_{j=1}^{m} \lambda_j$.

THANK YOU

Chinmayananda A.
Department of Electronics and Communication Engineering

chinmay@pes.edu

+91 8197254535