Vol. 7
No. 3
May
2015

ISSN: 2075-4124
E-ISSN: 2075-7107

An international journal

BEYNƏLXALQ ELMİ ARAŞDIRMALAR JURNALI

INTERNATIONAL JOURNAL OF ACADEMIC RESEARCH

PART A

APPLIED AND NATURAL SCIENCES

PROQRES
BAKU, AZERBAIJAN

The nation's future success lies with science and education!

Heydar Aliyev
National Leader of Azerbaijan

INTERNATIONAL JOURNAL OF ACADEMIC RESEARCH

Vol. 7. No. 3. Iss.2
May, 2015

PART A. APPLIED AND NATURAL SCIENCES

Member of crossref

DOI: 10.7813/2075-4124.2015

Daxil edildiyi elmi bazalar:
Indexed by:

Master Journal List (ISI-Thomson Reuters, USA)
CAB Abstracts (ISI-Thomson Reuters, USA)
Zoological Records (ISI-Thomson Reuters, USA)
Norwegian Social Science Data Services (Norway)
ZentralBlatt MATH (Springer-Verlag, European Math. Society, Germany)
IndexCopernicus International (Poland)
EBSCO-Academic Search Complete (USA)
ULRICH's Web (USA)

"PROGRESS" IPS
Baku, Azerbaijan, 2015

INTERNATIONAL JOURNAL of ACADEMIC RESEARCH

Vol. 7, No. 3. Iss.2, May, 2015, Part A

DOI for issue: 10.7813/2075-4124.2015/7-3

Publishing bimonthly
Print **ISSN: 2075-4124**
Online **ISSN: 2075-7107**
National reg. No. **2996**

Editorial office:

97/2, I.Gutgashinli str.,
Baku, Azerbaijan
E-mail: **subijar@gmail.com**

Beynəlxalq Elmi Araşdırmalar Jurnalı (BEAJ)

2009-cu il, Milli Mətbuat Günündə Azərbaycan Respublikası Ədliyyə Nazirliyi tərəfindən rəsmi Dövlət Qeydiyyatına alınıb **(№ 2996)**. BEAJ Beynəlxalq ISSN Mərkəzində (Paris, Fransa) qeydiyyatdan keçərək mətbu orqan kimi **ISSN 2075-4124**, elektron jurnal kimi **E-ISSN 2075-7107** nömrələri ilə beynəlxalq nəşr statusu qazanıb.

Jurnal dünyanın 83 ölkəsinə (universitet və kitabxanalar) paylanır. Jurnalın təsisçisi "Proqres" İnternet və Poliqrafiya Xidmətləri MMC-dir. BEAJ ildə 6 dəfə - **Yanvar, Mart, May, İyul, Sentyabr və Noyabr** aylarında dərc olunur. Redaksiyanın yazılı icazəsi olmadan materialların təkrar nəşri, tərcümə edilərək yayılması qadağandır. Məqalələr bir qayda olaraq Beynəlxalq Redaksiya Heyətinin yekun qərarı ilə dərc olunur. Məqalələr unikal DOI ilə nömrələnir.

Format: 60x84 $^{1}/_{8}$. Şrift: Arial. Tiraj: 300
Jurnal "Proqres İPX" tərəfindən nəşrə hazırlanıb və çap olunub.

STUDY OF KERNEL PRINCIPAL COMPONENT REGRESSION FUNCTION ESTIMATORS PROPERTIES

I.Djakaria[1], **S.Guritno**[2], **S.K.Haryatmi**[3]

[1]Math. Depart., Gorontalo State Univ., Gorontalo, [2,3]Math. Depart., Gadjah Mada Univ., Yogyakarta **(INDONESIA)**
E-mails: iskar@ung.ac.id, suryoguritno@ug.ac.id, s_kartiko@yahoo.com

DOI: 10.7813/2075-4124.2015/7-3/A.69

Received: 25 Apr, 2015
Accepted: 17 May, 2015

ABSTRACT

The purpose of this paper is to examine the kernel principal component regression (KPCR), which is the development of the principal component regression (PCR) with the radial basis fungction (RBF) kernel or Gaussian kernel. This study started to present a standard principal component analysis (PCA) and kernel principal component analysis (KPCA), that includes PCA in feature space and KPCA in input space. The focus of this study describes the propertles of model KPCR published in several theorems, includes linear estimator, unbiased estimator, and the best estimator, that known the best linear unbiased estimator (BLUE).

Key words: PCR model, KPCR model, RBF kernel, BLUE

1. INTORDUCITON

Principal component analysis (PCA), is a widely used method to reduce the number of dimensions, for example p dimensions, of a set of data (variables) observed, became the k new variables, with $k \leq p$. Each k new variables reduction result is a linear combination of p original variables. with variance which is owned by the p original variables, can largely be explained by a k new variables. PCA has been studied since the early 20th century, Pearson. year of 1901, and Hotelling. year of 1935, have learned with practical computing methods[1]. PCA is so rapid development, particularly its application in various fields, for example in health, biology, environment, image processing, etc., so that it becomes a very good method to reduce the number of variables and nonlinear higher dimension, using the kernel function. This method is known as kernel principal component analysis (KPCA).

KPCA many scientists discussed[2], which compares the benefits KPCA feature space for pattern recognition by using a linear classifier, and found two benefits KPCA. First, the kernel (nonlinear) principal component (KPC) gives the best recognition evaluation compared with the sum of the linear principal components (PCs) associated; and second, the formation of nonlinear components can be developed using components more than in the corresponds linear case. This discussion is still in tune with the discussion of scientists[3], which the KPCA is regarded as an extension of the PCA. PCA is a classic technique to decipher linear features, while KPCA is one of the methods that describe nonlinear features.

Kernel principal component regression (KPCR) is a technique used for regressing kernel principal component (KPC) with the response variable through the least squares method. The first stage of the main components of the kernel regression procedures that determine which is the main component of the kernel linear combination of several variables in the form of entries kernel matrix **K**, and the second stage is the response variable regressed on the main components of the kernel in a linear regression model.

2. LITERATURE REVIEW

Principal Component Analysis

Principal component analysis (PCA), which is one method of multivariate analysis, is one of the statistical techniques specifically developed to overcome data reduction. PCA not only allows for the reduction of data, but also can solve the problem of reducing data in a way that allows the results to be used in other applications of multivariate statistical methods (eg, analysis of variance, or regression analysis, cluster analysis, and so on).

In algebra, the PC is a linear combination of p random variables $X_1, \ldots, X_p$. While, in geometry, these linear combinations are a new coordinate system obtained by rotating the original system with $X_1, \ldots, X_p$ as the coordinate axes. The new axis is the direction of the maximum variability and provide a simple covariance structure.

The PCs depends on the covariance matrix $\boldsymbol{\Sigma}$ (or correlation matrix $\boldsymbol{\rho}$) of $X_1, \ldots, X_p$, and its development is not required assuming multinormal (multivariate normal). On the other hand, the principal components described for multivariate normal population has a useful interpretation in terms of constant density elipsoida.

Kernel Principal Component Analysis

PCA only use a linear combination between variables to represent the data, so it can only be overcome linear relationship between variables. However, in reality a lot of data which have non-linear and non-separable relationships between variables, so we need a method to shown the form of non-linear of the PCA, using kernel PCA, which in turn can calculate principal components more efficiently in dimension higher in feature space (abstract space which is often not known mapping results).

Traditionally, non-linear PCA based on the degree of non-linear optimization in the input space. In this study, a non-linear PCA based on non-linear mapping into the feature space where the original PCA algorithm[2]. The complexity of some of the process is reduced by using a kernel that makes the possibility of calculating the inner product in a feature space using a kernel function in the input space.

- *PCA in feature space.*

Considering that the set are centered of m observations $\mathbf{x}_k$, $k = 1, \ldots, m$, $\mathbf{x}_k \in {}^n$, $\sum_{k=1}^{m} \mathbf{x}_k = 0$, PCA diagonalizes the covariance matrix

$$\mathbf{A} = \frac{1}{m}\sum_{j=1}^{m} \mathbf{x}_j \mathbf{x}_j' \qquad (1)$$

Based on the PCA that principally, and matrix **A** as a positive definite, then **A** is diagonalizable with eigenalues-eigenvectors

$\mathbf{A}\mathbf{v} = \lambda\mathbf{v}$, with eigenvalues $\lambda \geq 0$ and eigenvectors $\mathbf{v} \in {}^n\backslash\{0\}$.

Considering to $\mathbf{A}\mathbf{v} = \frac{1}{m}\sum_{j=1}^{m} \mathbf{x}_j \mathbf{x}_j' \mathbf{v} = \lambda \mathbf{v}$ then $\mathbf{v} = \frac{1}{m\lambda}\sum_{j=1}^{m} \mathbf{x}_j \mathbf{x}_j' \mathbf{v} = \frac{1}{m\lambda}\sum_{j=1}^{m} \langle \mathbf{x}_j, \mathbf{v}\rangle \mathbf{x}_j$. Here, $\langle \mathbf{x}_j, \mathbf{v}\rangle$ is the scalar, then all solutions $\mathbf{v}$, with $\lambda \neq 0$, must lie in the span of $\mathbf{x}_1, \ldots, \mathbf{x}_m$, such that $\mathbf{v} = \sum_{j=1}^{m} \alpha_j \mathbf{x}_j$. Consequently, $\lambda\langle \mathbf{x}_k, \mathbf{v}\rangle = \langle \mathbf{x}_k, \mathbf{A}\mathbf{v}\rangle$ for all $k = 1, \ldots, m$.

The PCA in feature space rewritten in the form of inner product[2]. Furthermore, the same calculations outlined in dot space (dot product) F.

Now, let Φ function that maps all data input $\mathbf{x} \in {}^n$, so

$$\Phi : {}^n \rightarrow F, \mathbf{x} \mapsto \mathbf{X} := \Phi(\mathbf{x}),$$

with $\Phi(\mathbf{x}) \in F$, where F defined as the feature space, which has a dimension larger and even infinite. Based on this transformation, it appears that feature space constructed by the vectors $\{\Phi(\mathbf{x}_1), \Phi(\mathbf{x}_2), \ldots, \Phi(\mathbf{x}_m)\}$. Then all vectors in feature space F can be expressed as a linear combination of the vectors $\{\Phi(\mathbf{x}_1), \Phi(\mathbf{x}_2), \ldots, \Phi(\mathbf{x}_m)\}$. By the same argument as above, data in feature space F, $\Phi(\mathbf{x}_1), \ldots, \Phi(\mathbf{x}_m)$, centered, i.e. $\sum_{k=1}^{m} \Phi(\mathbf{x}_k) = 0$, then the covariance matrix can be written by

$$\mathbf{C} = \frac{1}{m}\sum_{j=1}^{m} \Phi(\mathbf{x}_j)\Phi(\mathbf{x}_j)'. \qquad (2)$$

In another sense, the issue of PCA in high-dimensional feature space F can be constructed by diagonalization estimate of m-sample covariance matrix (2).[4] Now, we remainder found eigenvalues $\lambda \geq 0$ and eigenvectors $\mathbf{V} \in F\backslash\{0\}$ that satisfies

$$\lambda\mathbf{V} = \mathbf{C}\mathbf{V} \text{ dengan } \|\mathbf{V}\| = \langle \mathbf{V}, \mathbf{V}\rangle = 1 \qquad (3)$$

Based the describing above, that all vectors in feature space F can be expressed as a linear combination of the vectors $\{\Phi(\mathbf{x}_1), \Phi(\mathbf{x}_2), \ldots, \Phi(\mathbf{x}_m)\}$, then eigenvectors, which is the solution of eigenvalues-eigenvectors $\lambda\mathbf{V} = \mathbf{C}\mathbf{V}$, can also be expressed as a linear combination of the vectors $\{\Phi(\mathbf{x}_1), \Phi(\mathbf{x}_2), \ldots, \Phi(\mathbf{x}_m)\}$. So, there is a constant α_i where $i = 1, 2, \ldots, m$, such that

$$\mathbf{V} = \sum_{i=1}^{m} \alpha_i \Phi(\mathbf{x}_i) \qquad (4)$$

Let $\mathbf{K} = [k_{ij}]$ the $m \times m$ matrix with, and

$$k(\mathbf{x}_i, \mathbf{x}_j) = k_{ij} = \Phi(\mathbf{x}_i)\Phi(\mathbf{x}_j)', \qquad (5)$$

we will be found

$$\lambda m\mathbf{K}\alpha = \mathbf{K}^2\alpha \qquad (6)$$

where α is a column vector with elements $\alpha_1, \alpha_2, \ldots, \alpha_m$. Because **K** is a symmetric matrix, it would have a number of eigenvectors that are on the whole space, so that the solution of (6), can be found by solving

$$\lambda m\alpha = \mathbf{K}\alpha, \text{ untuk } \lambda \neq 0. \qquad (7)$$

Also,

$$\mathbf{V} = \textstyle\sum_{i=1}^{m} \alpha_i \Phi(\mathbf{x}_i) \qquad (8)$$

Besides symmetry, **K** semidefinit positive also, so that **K** has eigenvalues, $\lambda \geq 0$, and then yield solution $m\lambda$ from (6). In this case we need diagonalizes **K**. Let eigenvalues $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_m$ in **K**, and eigenvectors $\alpha^1, \ldots, \alpha^m$ corresponds, with λ_p first nonzero eigenvalues. Since normalized vectors in **K**, which satisfies $\|\mathbf{V}^k, \mathbf{V}^k\| = (\mathbf{V}^k)'(\mathbf{V}^k) = 1$ for all $k = p, \ldots, m$, then $\alpha^p, \ldots, \alpha^m$ have to be normalized.

Based (4) and (7), then the normalizing treatment $\alpha^p, \ldots, \alpha^m$ yield

$$1 = (\mathbf{V}^k)'(\mathbf{V}^k) = \sum_{i,j=1}^{m} \alpha_i^k \alpha_j^k \langle \Phi(\mathbf{x}_i), \Phi(\mathbf{x}_j)\rangle = \lambda_k(\alpha^k)'\alpha^k. \qquad (9)$$

To extracting PCs, the calculation projections on the eigenvectors $\mathbf{V}^k$ in F, for all $k = p, \ldots, m$. Let $\mathbf{x}$ be an input vector, the mapping $\Phi(\mathbf{x}) \in F$, it must be projected onto the vector **V** which has been normalized, and the projection into the k-th PCs used

$$\Phi^k(\mathbf{x}) = \langle \mathbf{V}^k, \Phi(\mathbf{x}) \rangle = \sum_{i=1}^{m} \alpha_i^k k(\mathbf{x}_i, \mathbf{x}) . \tag{10}$$

So, to extract the t-th (kernel) PC can use $\Phi^t(\mathbf{x}) = \sum_{i=1}^{m} \alpha_i^t k(\mathbf{x}_i, \mathbf{x})$.

To clarify the properties of the PCA in F with data in input space, should be considered the selection of specific kernel.

- ***KPCA in input space.***

As has been described, that, PCA in feature space is diagonalizes symmetric matrix **K**, which consists of the entire inner product that may be from a number of samples. This type of matrix is said Gram matrix (matrix Gramian).[2][5][6]

Considering to the properties of the kernel and Mercer's theorem,[6][7] that if k is a continuous kernel of a positive integral operator, so that mapping can be created into a space where k is a dot product, then the inner product in feature space $\langle \Phi(\mathbf{x}_i), \Phi(\mathbf{x}_j) \rangle$ can be applied to the input space using a kernel function $k(x_i, x_j)$, and matrix, **K**, that formed the kernel matrix is said:

$$\mathbf{K} = \begin{pmatrix} k(\mathbf{x}^1,\mathbf{x}^1) & k(\mathbf{x}^1,\mathbf{x}_2) & \cdots & k(\mathbf{x}^1,\mathbf{x}^m) \\ k(\mathbf{x}^2,\mathbf{x}^1) & k(\mathbf{x}^2,\mathbf{x}^2) & \cdots & k(\mathbf{x}^2,\mathbf{x}^m) \\ \vdots & \vdots & \ddots & \vdots \\ k(\mathbf{x}^m,\mathbf{x}^1) & k(\mathbf{x}^m,\mathbf{x}^2) & \cdots & k(\mathbf{x}^m,\mathbf{x}^m) \end{pmatrix}. \tag{11}$$

It has been considered that the data is centered in feature space, so that the projection of KPC can be done in space features:

$$\Phi_c(\mathbf{x}_i) = \Phi(\mathbf{x}_i) - \frac{1}{m}\sum_{k=1}^{m} \Phi(\mathbf{x}^k) \tag{12}$$

and applied directly to input space just to be done in dot product[7]:

$$k_c(\mathbf{x}, \mathbf{y}) = \langle \Phi_c(\mathbf{x}), \Phi_c(\mathbf{y}) \rangle = k(\mathbf{x}, \mathbf{y}) - \frac{1}{m}\sum_{k=1}^{m} k(\mathbf{x}, \mathbf{y}^k) - \frac{1}{m}\sum_{k=1}^{m} k(\mathbf{y}, \mathbf{x}^k) + \frac{1}{m^2}\sum_{k=1}^{m} k(\mathbf{x}^k, \mathbf{y}^k)$$

This expression is equivalent to a matrix[8]:

$$\mathbf{K}_c = \mathbf{K} - \mathbf{1}_m\mathbf{K} - \mathbf{K}\mathbf{1}_m + \mathbf{1}_m\mathbf{K}\mathbf{1}_m \tag{13}$$

where $\mathbf{1}_m$ is a square matrix with entries $(\mathbf{1}_m)_{ij} = 1/m$.

3. METHODOLOGY

The method used in this study is a discussion of literature, and other theoretical studies, as well as constructing a display of the software. The study of literature includes collecting basic literature, namely texbook and journals, both physical and electronic, and performed sporadically when needed.

Steps of research made to achieve the objective, i.e.:

i. To study and reviewing the basic concepts of PCA, especially in the decide of the PCs that contain information that is large compared to the other PCs, as well as the study of KPCA.

ii. To study of the properties estimator kernel principal component regression model (KPCR).

4. RESULTS AND DISCUSSION

Principal Component Regression Modeling

The discussion in this section begins with the principal component regression (PCR). The PCR are techniques to analyze the data is that regression multiple with multikolinearitas, namely least squares estimation (LSE) into unbiased estimator and consistent but inefficient because of a large variance so very different from the actual values. This condition often occurs when the regression model used, there is an independent variable or data predictors had a very high correlation with other predictors of data. Extremely, multicollinearity between the independent variables can be impacted to difficulties in distinguishing of each independent variable on the dependent variable or response data. One method to overcome the problem of multicollinearity is done with the PCR, by adding a degree of bias in estimation of regression, so the PCR can reduce the standard error.[9]

Let $\underset{(n\times 1)}{\mathbf{Y}} = (\mathbf{Y}_1, \mathbf{Y}_2, \ldots, \mathbf{Y}_n)'$ is a vector of observations result (dependent variable), also called the response vector, and $\underset{(n\times p)}{\mathbf{X}} = [\mathbf{X}_1, \mathbf{X}_2, \cdots, \mathbf{X}_n]'$ matrix of variable data observations corresponds, where n and p are the sample size and the number of (dimension) variables observations, respectively, with $n \geq p$.

Therefore, the PCR analysis involves the use of PCA, for example, in **X**, then the linear regression model for **Y** on **X** can be expressed as

$$\mathbf{Y} = \mathbf{X}\beta + \varepsilon \tag{14}$$

with, $\beta \in \mathbb{R}^p$ is the vector of unknown parameters of the regression coefficient, ε is a random error vector with $E(\varepsilon) = 0$, and $Var(\varepsilon) = \sigma^2 \mathbf{I}_{(n \times n)}$ for any unknwon variance parameters $\sigma^2 > 0$.

To found of the parameter unbias estimator β on data set, then used ordinary least squared (OLS) regression approach, assuming that **X** is a matrix of full rank column, that is, **X'X** invertible, as the following theorem.

Theorem 1

Let the PCR model following the linear regression model (ordinary), that can be rewriten into

$$\boldsymbol{\varepsilon} = \mathbf{Y} - \mathbf{X}\beta, \tag{15}$$

then the unbiased estimator to β is $\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$.

Proof:

Here, OLS method approach used, assuming **X** is a data full rank column matrix, so **X'X** invertible. Furthermore, to minimize the sum of squared errors, i.e. $\mathbf{J} = \sum_{i=1}^{n} \boldsymbol{\varepsilon}_i^2 = \boldsymbol{\varepsilon}_i'\boldsymbol{\varepsilon}_i$. So that

$$\mathbf{J} = (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}) = \sum_{i=1}^{n} (Y_i - \beta_0 - \beta_1 X_{i1} - \cdots - \beta_p X_{ip})^2.$$

This equation is also called normal equation, and can be written as (simple form), $(\mathbf{X}'\mathbf{X})\hat{\boldsymbol{\beta}} = \mathbf{X}'\mathbf{Y}$, as a result

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y} \quad \blacksquare \tag{16}$$

Kernel Principal Component Regression

With the same of argument above, the PCR can be developed into a KPCR. (See[8][12][13]). KPCR is a technique used to regression of the response variable through the least squares method. The first of the KPCR procedures that determine which is the PC of the kernel linear combination of several variables in the form of entries kernel matrix **K**, and the second is the response variable regressed on the KPCs in a linear regression model.

Common forms of KPCR model is

$$\mathbf{Y} = \gamma_{1i}\boldsymbol{\Phi}_{KPC}^1(\mathbf{x}) + \gamma_{2i}\boldsymbol{\Phi}_{KPC}^2(\mathbf{x}) + \cdots + \gamma_{li}\boldsymbol{\Phi}_{KPC}^l(\mathbf{x}) + \boldsymbol{\varepsilon}, \tag{17}$$

where **Y** : variable response, Φ : KPC as a predictor, γ_i : parameter KPCR model parameters, ε : error vector, with $i = 1, \ldots, m$, and $m \le p$. This model can be considered as the standard regression models feature space *F*

$$\mathbf{Y} = \beta_0 + \beta_1 k(\mathbf{x}_1, \mathbf{x}) + \ldots + \beta_m k(\mathbf{x}_m, \mathbf{x}) + \varepsilon. \tag{18}$$

and the simplest form is

$$\mathbf{Y} = \Phi\gamma + \varepsilon, \tag{19}$$

where **Y** is a vector of m observations depencent variable, Φ is a predictor $(m \times \ell)$ matrix with the i-th row is a vector $\Phi(\mathbf{x}_i)$ of observations mapping $\mathbf{x}_i$ into a feature space *F* with dimensions $\ell \le \infty$; γ is the vector of regression coefficients; ε is the error vector whose elements are the same as the variance σ^2 and independent form each other.

It is assumed that predictors $\{\Phi_1(\mathbf{x}), \Phi_2(\mathbf{x}), \ldots, \Phi_\ell(\mathbf{x})\}$ with a mean of zero. $\Phi'\Phi$ a proportional sample matrix and KPCA can be used to extract ℓ eigenvalues $(\lambda_1, \lambda_2, \ldots, \lambda_\ell)$ corresponds with the eigenvectors $\{\mathbf{V}^1, \mathbf{V}^2, \ldots, \mathbf{V}^\ell\}$. To projection $\Phi(\mathbf{x})$ into k-th non-linear PC is given by Equation (10). Furthermore, suppose matrix **V** consists of columns that formed with eigenvectors $\{\mathbf{V}^1, \mathbf{V}^2, \ldots, \mathbf{V}^\ell\}$ of covariance matrix **C** in Equation (2), and matrix α be constructed by extracting eigenvectors $\{\alpha^1, \alpha^2, \ldots, \alpha^\ell\}$ of **K**. Using this notation for the data $\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_m\}$, then Equation (10) can be written in a matrix form

$$\mathbf{P} = \Phi\mathbf{V} = \Phi\Phi'\alpha = \mathbf{K}\alpha \text{ (from Equation (5))}, \tag{20}$$

where $\mathbf{V} = = \Phi'\alpha$. Then, projecting all predictors into PCs, (19) can be written as

$$\mathbf{Y} = \mathbf{P}\mathbf{w} + \varepsilon, \tag{21}$$

where $\mathbf{P} = \Phi\mathbf{V}$ is a transformation predictor matrix $(m \times \ell)$ (20) and **V** is a $(\ell \times \ell)$ matrix that k-th column are eigenvectors $\mathbf{V}_k$. The columns of the matrix P is orthogonal and coefficient LSE, **w**, becomes (see Theorem 1)

$$\hat{\mathbf{w}} = (\mathbf{P}'\mathbf{P})^{-1}\mathbf{P}'\mathbf{Y} = \Lambda^{-1}\mathbf{P}'\mathbf{Y}, \tag{22}$$

$\Lambda = \mathrm{diag}(\lambda_1, \ldots, \lambda_\ell)$

Estimation $\hat{Y}$ of the model (19) (associated to (21)) can he expressed as

$$\hat{Y} = \mathbf{V}\hat{\mathbf{w}} = \mathbf{V}(\mathbf{P}'\mathbf{P})^{-1}\mathbf{P}'\mathbf{Y} = \sum_{i=1}^{\ell} \lambda_i^{-1}\mathbf{V}^i(\mathbf{V}^i)'\boldsymbol{\Phi}'\mathbf{Y} \tag{23}$$

and variance-covariance matrix that corresponds,[13] is

$$\text{Cov}(\hat{\gamma}) = \sigma^2\mathbf{V}(\mathbf{P}'\mathbf{P})^{-1}\mathbf{V}' = \sigma^2\mathbf{V}\mathbf{\Lambda}^{-1}\mathbf{V}' = \sigma^2\sum_{i=1}^{t}\lambda_i^{-1}\mathbf{V}'(\mathbf{V}')' . \quad (24)$$

Theorem 2

Let $\mathbf{V}_t$ is a $p \times t$ matrix for any $t = \{1, \ldots, p\}$, which consists of the first t-columns $\mathbf{V}$ with orthonormal columns, and

$$\mathbf{W}_t = \mathbf{X}\mathbf{V}_t = [\mathbf{X}\mathbf{v}_1, \ldots, \mathbf{X}\mathbf{v}_t] \quad (25)$$

are the $(n \times t)$ matrices with the first t-PCs as its columns. Let also

$$\hat{\gamma}_t = (\mathbf{W}_t'\mathbf{W}_t)^{-1}\mathbf{W}_t'\mathbf{Y} \in \mathbb{R}^t \quad (26)$$

are the regression coefficient estimator vector found by OLS regression of the response vector $\mathbf{Y}$ on data matrices $\mathbf{W}_t$. Then, for any $t = 1, \ldots, p$, KPCR estimator of β using the first t-PCs are given by $\hat{\boldsymbol{\beta}}_t = (\mathbf{V}_t\hat{\gamma}_t) \in \mathbb{R}^t$.

Proof:

From (25) substitute to (26), we found

$$\hat{\gamma}_t = (\mathbf{W}_t'\mathbf{W}_t)^{-1}\mathbf{W}_t'\mathbf{Y} = (\mathbf{V}_t'\mathbf{X}'\mathbf{X}\mathbf{V}_t)^{-1}\mathbf{V}_t'\mathbf{X}'\mathbf{Y}.$$

So that

$$\mathbf{V}_t'\mathbf{X}'\mathbf{X}\mathbf{V}_t)\hat{\gamma}_t = \mathbf{V}_t'\mathbf{X}'\mathbf{Y}$$

$$\mathbf{V}_t\hat{\gamma}_t = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$$

Considering (16), then express on above can be

$$\mathbf{V}_t\hat{\gamma}_t = \hat{\boldsymbol{\beta}} .$$

since $t = 1, \ldots, p$, then $\hat{\boldsymbol{\beta}}_t = (\mathbf{V}_t\hat{\gamma}_t) \in \mathbb{R}^t$. ■ (27)

The KPCR coefficient estimates with this OLS are the best linear unbiased estimates (BLUE). It is given in the following theorem.

Theorem 3

Estimator $\hat{\boldsymbol{\beta}}_t = (\mathbf{V}_t\hat{\gamma}_t) \in \mathbb{R}^t$ in Theorem 2 above is the KPCR coefficient estimator which OLS that is the best linear unnbiased estimator (BLUE) for $\boldsymbol{\beta}_t = (\mathbf{V}_t\gamma_t) \in \mathbb{R}^t$.

Proof:

Suppose $\mathbf{W}_t$ in Equation (25), and $\mathbf{Y} = \mathbf{W}_t\gamma_t + \boldsymbol{\varepsilon}$, so that (26). $\hat{\gamma}_t = (\mathbf{W}_t'\mathbf{W}_t)^{-1}\mathbf{W}_t'\mathbf{Y}$. An estimator is said to BLUE, if it satisfies the following properties.

1. *Linear*; hence $\hat{\gamma}_t$ a linear function of Y (26) and $\mathbf{V}_t$ matrix with constant elements, then the $\hat{\boldsymbol{\beta}}_t = \mathbf{V}_t\hat{\gamma}_t$ are linear;
2. *Unbiased*; since $E(\hat{\boldsymbol{\beta}}_t) = E(\mathbf{V}_t\hat{\gamma}_t) = \mathbf{V}_tE(\hat{\gamma}_t) = \mathbf{V}_t[\gamma_t + 0] = \mathbf{V}_t\gamma_t = \boldsymbol{\beta}_t$, then $\hat{\boldsymbol{\beta}}_t = (\mathbf{V}_t\hat{\gamma}_t)$ unbiased for $\boldsymbol{\beta}_t = \mathbf{V}_t\gamma_t$;
3. The *best estimator*; to prove the best estimator, it have a minimum variance (smallest) between the other linear unbiased estimator variance. Then

$$\text{Var}(\hat{\boldsymbol{\beta}}_t) = E(\hat{\boldsymbol{\beta}}_t - \boldsymbol{\beta}_t)^2 = E[(\hat{\boldsymbol{\beta}}_t - \boldsymbol{\beta}_t)(\hat{\boldsymbol{\beta}}_t - \boldsymbol{\beta}_t)'] = \sigma^2\mathbf{V}_t(\mathbf{W}_t'\mathbf{W}_t)^{-1}\mathbf{V}_t' . \quad (28)$$

Now, we will show that $\text{Var}(\hat{\boldsymbol{\beta}}_t) < \text{Var}(\hat{\boldsymbol{\beta}}_t^*)$.

Suppose $\hat{\boldsymbol{\beta}}_t^*$ is another of the linear estimator $\boldsymbol{\beta}_t$, that we can be written as

$$\hat{\boldsymbol{\beta}}_t^* = \mathbf{V}_t\gamma_t^* = \mathbf{V}_t[(\mathbf{W}_t'\mathbf{W}_t)^{-1}\mathbf{W}_t' + \mathbf{c}]\mathbf{Y}$$

by c matrix constants, so

$$\hat{\boldsymbol{\beta}}_t^* = \mathbf{V}_t[(\mathbf{W}_t'\mathbf{W}_t)^{-1}\mathbf{W}_t' + \mathbf{c}]\mathbf{Y} = \mathbf{V}_t[\gamma_t + \mathbf{c}\mathbf{W}_t\gamma_t + (\mathbf{W}_t'\mathbf{W}_t)^{-1}\mathbf{W}_t'\boldsymbol{\varepsilon} + \mathbf{c}\boldsymbol{\varepsilon}] \quad (29)$$

Hence $\hat{\boldsymbol{\beta}}_t^*$ is unbiased linear estimator of $\boldsymbol{\beta}_t$, then $E(\hat{\boldsymbol{\beta}}_t^*) = \boldsymbol{\beta}_t$. It means $\mathbf{V}_t\mathbf{c}\mathbf{W}_t\gamma_t = \mathbf{0}$, or, $\mathbf{V}_t\mathbf{c}\mathbf{W}_t = \mathbf{0}$ so

$$E(\hat{\boldsymbol{\beta}}_t^*) = \mathbf{V}_tE(\gamma_t) + E[\{(\mathbf{W}_t'\mathbf{W}_t)^{-1}\mathbf{W}_t' + \mathbf{c}\}\boldsymbol{\varepsilon}] = \mathbf{V}_t\gamma_t + \mathbf{0} = \boldsymbol{\beta}_t . \quad (30)$$

Now, from (29) and (30), we found

$$\hat{\boldsymbol{\beta}}_t^* - \boldsymbol{\beta}_t = \mathbf{V}_t[\gamma_t + \{(\mathbf{W}_t'\mathbf{W}_t)^{-1}\mathbf{W}_t' + \mathbf{c}\}\boldsymbol{\varepsilon}] - \mathbf{V}_t\gamma_t = \mathbf{V}_t\{(\mathbf{W}_t'\mathbf{W}_t)^{-1}\mathbf{W}_t' + \mathbf{c}\}\boldsymbol{\varepsilon} ,$$

and

$$\begin{aligned}\mathrm{Var}(\hat{\boldsymbol{\beta}}_i^*) &= E[(\hat{\boldsymbol{\beta}}_i^* - \boldsymbol{\beta}_i)(\hat{\boldsymbol{\beta}}_i^* - \boldsymbol{\beta}_i)'] = \sigma^2 \mathbf{V}_i\{(\mathbf{W}_i'\mathbf{W}_i)^{-1}\mathbf{V}_i' + \sigma^2 \mathbf{V}_i\mathbf{cc}'\mathbf{V}_i' \\ &= \mathrm{Var}(\hat{\boldsymbol{\beta}}_i) + \sigma^2 \mathbf{V}_i\mathbf{cc}'\mathbf{V}_i', \text{ (from Equation (28))}\end{aligned} \quad (31)$$

Equation (31) shows that the linear unbiased estimator of variance matrices $\hat{\boldsymbol{\beta}}_i^*$ are the sum of OLS estimator variance matrices $\sigma^2 \mathbf{V}_i\mathbf{cc}'\mathbf{V}_i'$, thus $\mathrm{Var}(\hat{\boldsymbol{\beta}}) \leq \mathrm{Var}(\hat{\boldsymbol{\beta}}_i^*)$. ■

REFERENCES

1. Tipping. M. E. dan Bishop, C. M. 1999. Probabilistic principal component analysis. *Journal of the Royal Statistical Society*, Series B, 61, Part 3. 611-622
2. Schölkopf. B., Smola. A., dan Müller, K-R. 1998, Nonlinear component analysis as a kernel eigenvalue problem. *Neural Computation*, 10: 1299-1319
3. Yang. J., Frangi, A. F., dan Yang. J-y. 2004. A new kernel fisher discriminant algorilm with application to face recognition. *Neurocomputing* 56: 415-421
4. Rosipal, R., Trejo, L. J. Andrzej, C. 2001a, Principal Component Regression with EM Approach to Nonlinear Principal Component Extraction. *Technical Report*. Scotland, UK: School of Information and Communication Technologies, University of Paisley
5. Schölkopf. B., Smola, A., dan Müller, K-R. 2001. Kernel PCA: Nonlinear component analysis as a kernel eigenvalue problem. *Vision and Learning*. 1-21
6. Schölkopf, B. dan Smola, A. J. 2002. *Learning With Kernels*. MIT Press, Cambridge, MA
7. Min, R., Bonner, A., dan Zhang, Z. 2007. Modifying Kernels Using Label Information Improves SVM Classification Performance, Canada: Univ. of Toronto
8. Fauvel, M. 2007, Spectral and spatial methods for the classification of urban remote sensing data. *Doctoral Thesis*. Iceland: Faculty of Engineering, Iceland University
9. http://www.ncss.com/software/pass/pass-documentation/. 2015. Principal Component Regression, *NCSS Statistical Software*
10. Olliffe, I. T. 2002. *Principal Component Analisys, 2nd*. New York: Springer
11. Anderson. J. R., Hardy, E. E., Roach, J. T., dan Wittmer, R. E. 2001. A Land Use And Land Cover Classification System For Use With Remote Sensor Data. *Geological Survey Professional Paper 964*. Washington: United States Government Printing Office.
12. Rosipal, R., Trejo, L. J., dan Andrzej, C. 2001b, Principal Component Regression with Covariance Inflation Criterion for Model Selection. *Technical Report*. Scotland. UK: School of Information and Communication Technologies, University of Paisley
13. Rosipal, R., Trejo, L. J., dan Andrzej. C. 2002, Kernel PCA for Feature Extraction and De-Noising in Non-linear Regression. *Technical Report*. Scotland, UK: School of Information and Communication Technologies, University of Paisley