2014 International Joint Conference on Neural Networks (IJCNN)
July 6-11, 2014, Beijing, China

Kernel Ridge Regression Classification

Jinrong He, Lixin Ding, Lei Jiang and Ling Ma

***Abstract*—We present a nearest nonlinear subspace classifier that extends ridge regression classification method to kernel version which is called Kernel Ridge Regression Classification (KRRC). Kernel method is usually considered effective in discovering the nonlinear structure of the data manifold. The basic idea of KRRC is to implicitly map the observed data into potentially much higher dimensional feature space by using kernel trick and perform ridge regression classification in feature space. In this new feature space, samples from a single-object class may lie on a linear subspace, such that a new test sample can be represented as a linear combination of class-specific galleries, then the minimum distance between the new test sample and class specific subspace is used for classification. Our experimental studies on synthetic data sets and some UCI benchmark datasets confirm the effectiveness of the proposed method.**

I. Introduction

MANY applications of machine learning, ranging from text categorization to computer vision, require the classification of large volumes of complex data sets. Among the algorithms, the K nearest neighbor (KNN) method is one of the most successful and robust methods for many classification problems at the same time being simple and intuitive [1]. In this naive approach, a test sample is assigned to the class which contains the nearest sample. Despite its advantages, the KNN algorithm suffers from poor generalization ability and becomes less effective when the samples having different class labels are comparable in the neighborhood of a test sample [2]. To overcome the drawbacks of KNN, various methods have been proposed in the literature [3][4][5][6].

Subspace learning [7][8] is a traditional method for pattern classification, which always assumes that data is sampled from a linear subspace. Other than KNN method, nearest subspace classification methods [9][10] classify a new test sample into the class whose subspace is the closest. CLASS featuring information compression (CLAFIC) is one of the earliest and well-known subspace methods [11] and its extension into the nonlinear subspace is the Kernel CLAFIC (KCLAFIC). This method employs the principal component analysis to compute the basis vectors spanning subspace of each class. Linear Regression Classification (LRC) [12] method is another popular subspace method. In LRC, classification is taken as a class specific linear regression problem and the regression coefficients are estimated by using the least square estimation method, and then the classification is made by the minimum distance between the original sample and the projected sample. However, the least square estimation is very sensitive to outliers [13]. Therefore, the performance of LRC decreases sharply as the sample contaminated by outliers. Due to L2,1-norm based loss function can reduce the effect of outliers, Ren Chuan-Xian [14] proposed rotational invariant norm based regression classification method. In 2012, Naseem et al. proposed a robust linear regression classification algorithm (RLRC) [15] to estimate regression parameters by using the robust Huber estimation. Compared with least square estimation, the Huber's M-estimator weighs the large residuals more lightly. As a result, the outliers have less significant affection on the estimated coefficients. Moreover, to overcome the problem of multicollinearity in LRC, Huang and Yang [16] proposed an improved principal component regression classification (IPCRC) method which removes the mean of each sample before performing principal component analysis and drops the first principal components. The projected coefficients are then executed by the linear regression classification algorithm. However, when the axes of linear regression of class-specific samples have an intersection, LRC could not well classify the samples that distribute around the intersection. To improve the performance of LRC in this situation, Yuwu Lu et al. proposed a kernel linear regression classification (KLRC) algorithm [17], by integrating the kernel trick and LRC.

Although many regression-based approaches have been proposed to achieve successful classification tasks, LRC method fails when the number of sample in the class specific training set is smaller than their dimension. The ridge regression method [18] is a regularized least square method for classification and regression. It is suitable only for datasets with few training examples. Kernel methods [19] are effective framework to enhance the modeling capability by nonlinearly mapping the data from the original space to a high dimensional feature space which is called reproducing kernel Hilbert space (RKHS).

Jinrong He is with the School of Computer, State Key Laboratory of Software Engineering, Wuhan University, Wuhan, 430072 China (corresponding author to provide phone: 086-15392822848; e-mail: hejinrong@whu.edu.cn).

Lixin Ding is with State Key Laboratory of Software Engineering, Wuhan University, Wuhan, 430072 China (e-mail: lxding@whu.edu.cn).

Lei Jiang is with Key Laboratory of Knowledge Processing and Networked Manufacture, Hunan University of Science and Technology, Xiangtan, 411201 China (e-mail: jlefe@126.com).

Ling Ma is with the Second Artillery Equipment Academy, Beijing 10085 China (e-mail: whjcc@163.com).

This work was supported in part by the Fundamental Research Funds for the Central Universities (No. 2012211020209), Special Project on the Integration of Industry, Education and Research of Ministry of Education and Guangdong Province (2011B090400477), Special Project on the Integration of Industry, Education and Research of Zhuhai City (2011A050101005, 2012D0501990016), Zhuhai Key Laboratory Program for Science and Technique (2012D0501990026).

978-1-4799-1484-5/14/$31.00 ©2014 IEEE

In order to detect nonlinear structure of samples and improve the robustness of LRC algorithm, we propose a Kernel Ridge Regression Classification (KRRC) method to boost the effectiveness of the LRC. KRRC is a nonlinear extension of ridge regression classification method based on kernel trick, which implicitly maps the data into a high-dimensional kernel space by using a nonlinear mapping determined by a kernel function. KRRC method falls in the category of nearest subspace classification that shares similar idea with LRC.

In the remainder of the paper, we will first describe the ridge regression classification method, and then the proposed KRRC method is presented in Section II. It is followed by extensive experiments using synthetic data sets and UCI data sets in Section III. The paper concludes in Section IV.

II. KERNEL RIDGE REGRESSION CLASSIFICATION

A. Ridge Regression Classification (RRC)

Ridge regression [20] is a classical data modeling method to solve multicollinearity problem of covariates in samples. Here, multicollinearity refers to a situation in which more than one predictor variables in a multiple regression model are highly correlated. If multicollinearity is perfect, the regression coefficients are indeterminate and their standard errors are infinite. If it is less than perfect, the regression coefficients although determinate but possess large standard errors, which means that the coefficients cannot be estimated with great accuracy [21].

Using a fundamental concept that samples from a specific class lie on a linear subspace, a new test sample from any class can be represented as a linear combination of class-specific training samples. This assumption can be formulated as a linear model in terms of ridge regression.

Assume that we have C classes and each class has n_i samples in the d-dimensional space. Let X_i be the training set of the ith class whose data matrix is

$$X_i = \left[x_1^i, x_2^i, \cdots, x_{n_i}^i \right] \in R^{d \times n_i}$$

According to subspace assumption, the new test sample x belongs to the ith class can be represented by the linear combinations of these samples with an error ε according to LRC method. Hence

$$x = X_i \alpha_i + \varepsilon \tag{1}$$

Where α_i is $n_i \times 1$ dimensional regression coefficients vector.

Similar to LRC, we formulated ridge regression classification method as follows.

For any new test sample x, the goal of the ridge regression is to find $\hat{\alpha}_i$ to minimize the residual error as:

$$\hat{\alpha}_{ii} = \arg\min_{\alpha_i} \left\| x - X_i \alpha_i \right\|_2^2 + \lambda \left\| \alpha_i \right\|_2^2 \tag{2}$$

Here, λ is regularization parameter. Ridge regression can reduce the variance by penalizing the norm of the linear transform and balance the bias and variance by adjusting the regularization parameter.

If we take the derivative of Equation (2) with respect to α_i and set it to zero, we get

$$X_i^T X_i \alpha_i - X_i^T x + \lambda \alpha_i = 0 \tag{3}$$

Then the estimate of the regression parameter vectors can be computed by

$$\hat{\alpha}_i = \left(X_i^T X_i + \lambda I \right)^{-1} X_i^T x \tag{4}$$

Thus the projection of x onto the subspace of the ith class can be computed as

$$\tilde{x}^i = X_i \hat{\alpha}_i = X_i \left(X_i^T X_i + \lambda I \right)^{-1} X_i^T x_i \triangleq H_i x \tag{5}$$

Where H_i is the class specific projection matrix which is defined as follows:

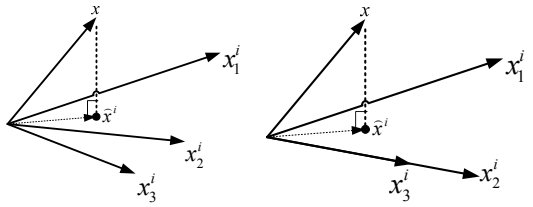
$$H_i = X_i \left(X_i^T X_i + \lambda I \right)^{-1} X_i^T \tag{6}$$

Note that the projection matrix is a symmetric matrix and also idempotent, i.e., $H_i^T = H_i$, $H_i^2 = H_i \cdot H_i = H_i$.

After projecting new test sample onto every class-specific subspace, the minimum distance between the new test sample and class specific subspace is used for classification. If the original sample belongs to the subspace of class i, the projected sample $\tilde{x}^i$ onto the class specific subspace X_i will be the closet sample to the original sample.

$$i^* = \arg\min_i \left\| \tilde{x}^i - x \right\|_2^2 \tag{7}$$

Figure 1 shows geometric interpretation of LRC and RRC method. Ridge regression classification is a regularized least square method to model the linear dependency between class specific samples and the new test sample which can deal with multicollinearity problem.



(a) LRC (b) RRC

Fig.1. Geometric Interpretation of LRC and RRC methods

B. Kernel Ridge Regression Classification

The main idea of Kernel Ridge Regression Classification (KRRC) is to map the original samples into a higher dimensional Hilbert space F and apply RRC method on this Hilbert space F. Kernel trick my increase the linearity of samples, i.e., a nonlinear curve can be taken as lying on a plane. The nonlinear mapping function can be denoted as $\phi : X \to F$. For any new test sample x, the goal of kernel ridge regression is to find $\hat{\alpha}_i$ to minimize the residual error as:

$$\hat{\alpha}_{ii} = \arg\min_{\alpha_i} \left\| \phi(x) - \phi(X_i) \alpha_i \right\|_2^2 + \lambda \left\| \alpha_i \right\|_2^2 \tag{8}$$

Similar to Equation (4), the estimate of the regression parameter vectors can be computed by

$$\hat{\alpha}_i = \left(\phi^T(X_i) \phi(X_i) + \lambda I \right)^{-1} \phi^T(X_i) \phi(x) \tag{9}$$

Then we can predict the response vector $\widehat{\phi^i(x)}$ for the ith class as

$$\begin{aligned}\widehat{\phi^i(x)} &= \phi(X_i)\hat{\alpha}_i \\ &= \phi(X_i)\left(\phi^T(X_i)\phi(X_i)+\lambda I\right)^{-1}\phi^T(X_i)\phi(x) \\ &\triangleq H_i^{\phi}\phi(x)\end{aligned} \tag{10}$$

Where $\widehat{\phi^i(x)}$ is the projection of $\phi(x)$ onto the subspace of the *i*th class by the class specific projection matrix which is defined as follows:

$$H_i^{\phi} = \phi(X_i)\left(\phi^T(X_i)\phi(X_i)+\lambda I\right)^{-1}\phi^T(X_i) \tag{11}$$

If the original sample belongs to the subspace of class *i*, the predicted sample $\widehat{\phi^i(x)}$ in kernel space *F* will be the closet sample to the original sample.

$$\begin{aligned}i^* &= \arg\min_i \left\|\widehat{\phi^i(x)}-\phi(x)\right\|_2^2 \\ &= \widehat{\phi^i(x)}^T\widehat{\phi^i(x)} - 2*\widehat{\phi^i(x)}^T\phi(x)+\phi^T(x)\phi(x)\end{aligned} \tag{12}$$

According to Mercer's theorem [19], the form of nonlinear function $\phi(x)$ is not necessarily known explicitly and could be determined by a kernel function $k: X\times X\to R$ which has the following property:

$$k(x_i,x_j) = \phi^T(x_i)\phi(x_j) \tag{13}$$

There are numerous types of kernel functions [19]. In our experiments, we adopt most popular Gaussian kernel that is given by

$$k(x_i,x_j) = \exp\left(-\frac{\left\|x_i-x_j\right\|_2^2}{t}\right) \tag{14}$$

The parameter *t* is empirically set as the average Euclidean distance of all training samples.

Obviously, the classification process (12) can be expressed in terms of inner products between mapped training samples in *F*. Let us define kernel matrix *K* whose elements is

$$K(X_i,X_j) = \begin{pmatrix} k(x_1^i,x_1^j) & k(x_1^i,x_2^j) & \cdots & k(x_1^i,x_{n_j}^j) \\ k(x_2^i,x_1^j) & k(x_2^i,x_2^j) & \cdots & k(x_2^i,x_{n_j}^j) \\ \vdots & \vdots & \cdots & \vdots \\ k(x_{n_i}^i,x_1^j) & k(x_{n_i}^i,x_2^j) & \cdots & k(x_{n_i}^i,x_{n_j}^j) \end{pmatrix} \tag{15}$$

Following some simple algebraic steps, we see that the first term in Equation (12) can be reformulated as

$$\begin{aligned}&\widehat{\phi^i(x)}^T\widehat{\phi^i(x)} \\ &= \phi^T(x)\phi(X_i)\left(\phi^T(X_i)\phi(X_i)+\lambda I\right)^{-1}\phi^T(X_i) \\ &\quad \phi(X_i)\left(\phi^T(X_i)\phi(X_i)+\lambda I\right)^{-1}\phi^T(X_i)\phi(x) \\ &= K(x,X_i)\left(K(X_i,X_i)+\lambda I\right)^{-1} \\ &\quad K(X_i,X_i)\left(K(X_i,X_i)+\lambda I\right)^{-1}K(X_i,x) \\ &= A^T K(X_i,X_i)A\end{aligned} \tag{16}$$

Similarly, the second term in Equation (12) can be reformulated as

$$\begin{aligned}&\widehat{\phi^i(x)}^T\phi(x) \\ &= \phi^T(x)\phi(X_i)\left(\phi^T(X_i)\phi(X_i)+\lambda I\right)^{-1}\phi^T(X_i)\phi(x) \\ &= K(x,X_i)\left(K(X_i,X_i)+\lambda I\right)^{-1}K(X_i,x) \\ &= K(x,X_i)A\end{aligned} \tag{17}$$

Here

$$A = \left(K(X_i,X_i)+\lambda I\right)^{-1}K(X_i,x) \tag{18}$$

Since $K(X_i,X_i)+\lambda I$ is positive definite, and its Cholsky decomposition can be written as

$$K(X_i,X_i)+\lambda I = L^T L \tag{19}$$

Then the matrix *A* in Equation (18) can be efficiently computed by solving the following linear equation

$$L^T LA = K(X_i,x) \tag{20}$$

Note that the third term in Equation (12) has no effect on classification results, since it has nothing to do with the class information. Therefore, after neglecting the third term, we have

$$\begin{aligned}&\widehat{\phi^i(x)}^T\widehat{\phi^i(x)} - 2*\widehat{\phi^i(x)}^T\phi(x) \\ &= A^T K(X_i,X_i)A - 2*K(x,X_i)A \\ &= A^T\left(K(X_i,X_i) - 2*\left(K(X_i,X_i)+\lambda I\right)\right)A \\ &= -A^T\left(K(X_i,X_i)+2\lambda I\right)A\end{aligned} \tag{21}$$

Equivalently, the classification process (12) can be reformulated as

$$i^* = \arg\max_i \left\{A^T\left(K(X_i,X_i)+2\lambda I\right)A\right\} \tag{22}$$

The KRRC algorithm is given in Algorithm 1.

Algorithm 1 Kernel Ridge Regression Classification

Input: training data matrix *X_train*, Label vector for training data *L_train* and testing data matrix *X_test*.

Procedure:

For each testing data sample *x*, predict its label as follows:

Step 1. Compute the kernel matrix $K(X_i,X_i)$ and $K(X_i,x)$ with Gaussian kernel (14).

Step 2. Compute matrix *A* with (20).

Step 3. Decision is made in favor of the class with the minimum distance in (22).

Output: Label vector for testing data *L_test*.

III. Experimental Results

In this section, we conduct experiments on synthetic data sets and UCI data sets to evaluate results of our proposed method for classification task and compare its results with those of the related classification methods.

A. Experimental Setup

The proposed KRRC method is compared with the related algorithms, such as KNN, LRC and KLRC. We use Gaussian kernel in Equation (14) for KRRC and KLRC. In our experiments, the regularization parameter λ was set as 0.005. For each data set, we use 5-fold cross-validation to evaluate the performance of proposed method, i.e., 4 folds are used for

training and the last fold is used for testing. This process is repeated 5 times, leaving one different fold for testing each time. The average accuracy and corresponding standard deviation over the five runs of cross validation is reported for evaluation.

B. Experiments on Synthetic Data Sets

We first conduct experiments on two synthetic data sets displayed in Fig. 2 and Fig.3. In these figure, the data points that belong to the same class are shown with the same color and style. Obviously, they can't be classified linearly. The performance is shown in Tables I – II.

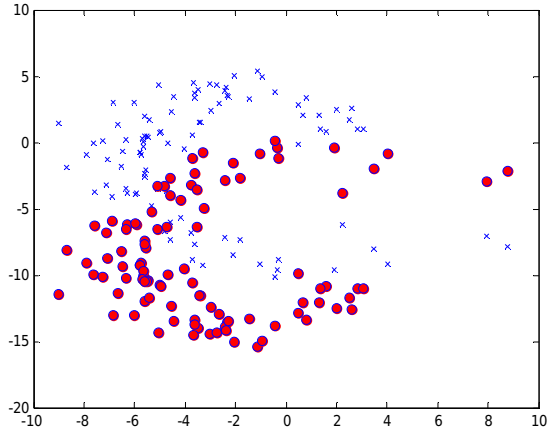


Fig.2. The Synthetic Data Set 1

TABLE I
CLASSIFICATION RESULTS (%) COMPARISONS ON SYNTHETIC DATA SET 1

method	KNN	LRC	KLRC	KRRC
Accuracy	88.50	74.00	87.00	89.00
Standard deviation	5.61	8.46	2.92	3.74

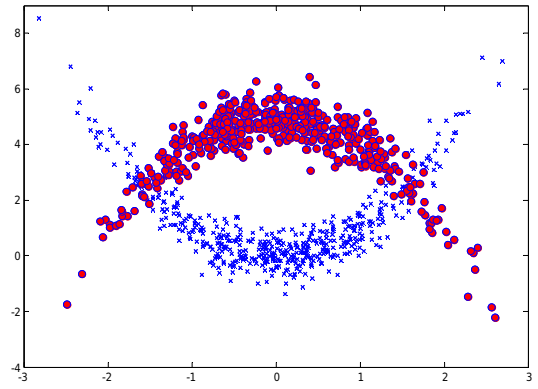


Fig.3. The Synthetic Data Set 2

TABLE II
CLASSIFICATION RESULTS (%) COMPARISONS ON SYNTHETIC DATA SET 2

method	KNN	LRC	KLRC	KRRC
Accuracy	95.50	75.70	96.80	97.00
Standard deviation	0.84	8.90	0.87	1.05

According to Table I and II, LRC performed not well on these two synthetic data sets, while KNN, KLRC and KRRC give better results. Since these synthetic data sets has nonlinear structure and the assumption underlying LRC method is not satisfy.

C. Experiments on UCI Data Sets

In the experiments, we choose 14 real world data sets with varying dimensions and number of data points from UCI data repository to test our algorithm. The data sets are named as wine, Soybean2, Soybean1, liver, heart, glass, breast, yeast, vowel, diabetes, seeds, dermatology, hepatitis and balance [22]. The detail of the data description is shown in Table III. Table IV shows the classification results of different methods. The numbers in the brackets are the corresponding standard deviation. According to Table IV, our method generally shows higher performance than the other methods.

TABLE III
UCI DATA DESCRIPTIONS AND EXPERIMENTAL SETTINGS

Data Set	#Dimension	#Number	#Class
wine	13	178	3
Soybean2	35	136	4
Soybean1	35	266	15
liver	6	345	2
heart	13	297	2
glass	9	214	4
breast	9	683	2
yeast	8	1484	10
vowel	10	528	11
diabetes	8	768	2
seeds	7	210	3
dermatology	34	366	6
hepatitis	19	155	2
balance	4	625	3

TABLE IV
ACCURACY (%) COMPARISONS ON UCI DATA SET

Data Set/Method	KNN	LRC	KLRC	KRRC
wine	78.05(2.96)	52.92(7.67)	84.87(6.79)	**87.14(3.53)**
Soybean2	86.75(6.05)	80.16(6.83)	86.75(6.05)	**87.49(6.04)**
Soybean1	86.46(2.51)	85.72(4.36)	89.11(3.63)	**89.48(3.49)**
liver	62.03(5.14)	60.29(5.23)	58.55(4.90)	**68.12(3.78)**
heart	57.93(2.90)	67.65(4.38)	71.06(2.53)	**73.40(4.04)**
glass	71.99(6.96)	50.96(5.25)	63.62(7.75)	**72.43(7.24)**
breast	96.19(1.41)	35.43(0.67)	96.93(1.81)	**97.36(1.83)**
yeast	51.75(1.00)	35.92(2.33)	53.10(3.85)	**59.84(1.68)**
vowel	98.67(0.97)	59.47(2.74)	98.86(0.71)	**99.24(0.71)**
diabetes	67.71(2.69)	62.10(3.90)	69.01(1.98)	**73.18(4.34)**
seeds	90.48(2.61)	62.38(4.86)	91.43(2.43)	**93.81(3.23)**
dermatology	88.80(1.57)	91.25(2.55)	93.99(2.22)	**94.26(1.03)**
hepatitis	54.19(8.75)	60.65(7.74)	55.48(3.76)	**61.29(5.40)**
balance	80.15(3.70)	90.72(0.83)	89.92(1.71)	**92.79(1.70)**

However, we cannot conclude which classification method will certainly beat the others. In this experiment, we see that KRRC performs little better in more of the selected data sets.

IV. CONCLUSION

In this paper, we presented a kernel ridge regression classification (KRRC) algorithm based on ridge regression for classification. KRRC algorithm firstly makes a nonlinear mapping of the data to a feature space, and then perform ridge regression classification method on this feature space, so KRRC is good at enhancing the linearity of distribution structure underlying samples and able to obtain higher accuracy than LRC. We showed the effective performance of our method by comparing its results on the synthetic and UCI data sets with related subspace based classification methods. However, KRRC require matrix inversion computation which can be computationally intensive for high dimensional and large datasets, including text documents, face images, and gene expression data. Therefore, developing efficient algorithms yet with theoretical guarantees will be interesting in future research.

ACKNOWLEDGMENT

The authors would like to thank the anonymous reviewers for their constructive comments and suggestions.

REFERENCES

[1] Cover, Thomas, and Peter Hart, "Nearest neighbor pattern classification," *IEEE Transactions on Information Theory*, vol.13, pp. 21-27, January 1967.

[2] Fayed, Hatem A., and Amir F. Atiya, "A novel template reduction approach for the k-nearest neighbor method," *IEEE Transactions on Neural Networks*, vol. 20, pp. 890-896, May 2009.

[3] T. Hastie and R. Tibshirani, "Discriminant adaptive nearest neighbor classification," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 18, no. 6, pp. 607–616, Jun. 1996.

[4] J. Peng, D. R. Heisterkamp, and H. K. Dai, "LDA/SVM driven nearest neighbor classification," *IEEE Trans. Neural Netw.*, vol. 14, no. 4, pp. 940–942, Jul. 2003.

[5] H. A. Fayed and A. F. Atiya, "A novel template reduction approach for the K-nearest neighbor method," *IEEE Trans. Neural Netw.*, vol. 20, no. 5, pp. 890–896, May 2009.

[6] Y. G. Liu , S. Z. S. Ge , C. G. Li and Z. S. You, "K-NS: A classifier by the distance to the nearest subspace", *IEEE Trans. Neural Netw.*, vol. 22, no. 8, pp.1256 -1268, 2011.

[7] Oja, Erkki. *Subspace methods of pattern recognition*. England: Research Studies Press, 1983, Vol. 4.

[8] Cappelli, Raffaele, Dario Maio, and Davide Maltoni,"Subspace classification for face recognition," *Biometric Authentication*. 2002.

[9] Li, Stan Z, "Face recognition based on nearest linear combinations," *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, IEEE, 1998.

[10] Li, Stan Z., and Juwei Lu, "Face recognition using the nearest feature line method," *IEEE Transactions on Neural Networks*, Vol. 10, pp. 439-443, February, 1999.

[11] S. Watanabe, P. F. Lambert, C. A. Kulikowski, J. L. Buxton, and R. Walker, *Evaluation and selection of variables in pattern recognition*, *In Computer and Information Sciences II*. New York: Academic, 1967.

[12] Naseem, Imran, Roberto Togneri, and Mohammed Bennamoun, "Linear regression for face recognition," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol.32, pp.2106-2112, 2010.

[13] Huber, Peter J. *Robust statistics*. Springer Berlin Heidelberg, 2011.

[14] Ren, Chuan-Xian, Dao-Qing Dai, and Hong Yan, "L2,1-norm based Regression for Classification," *2011 First Asian Conference on Pattern Recognition (ACPR)*, IEEE, 2011.

[15] Naseem, Imran, Roberto Togneri, and Mohammed Bennamoun, "Robust regression for face recognition," *Pattern Recognition*, vol.45, pp. 104-118, January 2012.

[16] Huang, Shih-Ming, and Jar-Ferr Yang. "Improved principal component regression for face recognition under illumination variations." *Signal Processing Letters*, IEEE. Vol.19, pp. 179-182. April, 2012.

[17] Lu, Yuwu, Xiaozhao Fang, and Binglei Xie. (June, 2013) Kernel linear regression for face recognition. Neural Computing and Applications [Online]. pp. 1-7. Available: http://dx.doi.org/10.1007/s00521-013-1435-6.

[18] Hastie, Trevor, et al. *The elements of statistical learning: data mining, inference and prediction*. The Mathematical Intelligencer 27.2 (2005): 83-85.

[19] Shawe-Taylor, John, and Nello Cristianini. *Kernel methods for pattern analysis*. Cambridge university press, 2004.

[20] Hoerl, Arthur E., and Robert W. Kennard, "Ridge regression: Biased estimation for nonorthogonal problems," *Technometrics*, Vol. 12, pp. 55-67, January, 1970.

[21] Gujarati, Damodar N., and J. B. Madsen, "Basic econometrics," *Journal of Applied Econometrics.* Vol. 13, pp. 209-212, February 1998.

[22] Bache, K. & Lichman, M. (2013). UCI Machine Learning Repository. Irvine, CA: University of California, School of Information and Computer Science. Available: http://archive.ics.uci.edu/ml