CHAPTER 5

5. WRITER IDENTIFICATION MODEL THROUGH SVM WITH MODIFIED LINEAR KERNELS

Generally in most of the existing SVM implementations RBF kernel showed better performance than linear kernel and polynomial kernel. Also learning with linear kernel is faster and it has less computational complexity. The same is proved by the experimental results discussed in chapter 4. This chapter presents a modified linear kernel with parameter estimation and the implementation of SVM with modified linear kernel to enrich the performance of the model in recognizing the Tamil writer. Parameter estimation methods such as Weighted Least Square (WLS) Regression, Bayesian Linear Regression (BLR) and Principal Component Regression (PCR) used to determine the coefficients of linear kernel are also described in this chapter. The three kernels defined here are Weighted Least Kernel (WLK), Bayesian Linear Kernel (BLK), Principal Component Kernel (PCK). The models built using SVM with these three kernels are presented in sections 5.1 to 5.3. The classification results of SVM by training the same three datasets such as TWINC, TWINW and TWINP are presented in this chapter.

5.1. MODEL II - WRITER IDENTIFICATION MODEL THROUGH SVM WITH WEIGHTED LEAST KERNEL

It is proposed to employ the new form of linear kernel by considering the accuracy requirement, computational time, computational complexity and the nature of the problem. As described in chapter 2, the standard form of L2 norm SVM formulation in matrix format is,

$$\min_u L(u) = \frac{1}{2} u^T Qu - e^T u \tag{5.1}$$

Subject to

$$d^T u = 0,\ u \geq 0$$

Q can be computed as Q = ($A * A^T$ + I/C).* ($d * d^T$). It is noted that the algorithm SVM finally requires three pieces of data Q, d and C where C is the regularization parameter, d is the diagonal matrix of class labels. Q is the obtained from $A * A^T$ and $d * d^T$ [47].

$$AA^T = \begin{pmatrix} x_1^T x_1 & \cdots & x_l^T x_m \\ \vdots & \ddots & \vdots \\ x_m^T x_1 & \cdots & x_m^T x_m \end{pmatrix} = K \tag{5.2}$$

The i, jth element of AA^T is $x_i^T x_j$ i.e. a dot product of two feature vectors x_i and x_j. The matrix K is called the linear kernel matrix which implies that all information needed for training is captured in the form of dot products of the training vectors. K is positive definite matrix and the set of kernels satisfy closure property. Complex kernels can be defined using simple one and employed in SVM for better learning. Some of the forms of linear kernel K are stated below,

$$K_1 = AA^T \tag{5.3}$$

$$K_2 = \alpha AA^T \tag{5.4}$$

$$K_3 = AA^T + \beta \tag{5.5}$$

$$K_4 = \alpha AA^T + \beta \tag{5.6}$$

In this work the linear kernel K_2 is used and the parameter α is obtained using parameter estimation method. The constant vector α is of dimension equal to number of samples in the training dataset and each element is a parameter added to the sum of the squares of the features. Weighted Least Square parameter estimation method is used here to estimate the dot products of the linear kernel and the modified linear kernel is formed as WLK. This method of parameter estimation is described below.

Weighted Least Square Regression

In weighted least squares the distribution of the errors is unknown and permits general forms of unknown heteroscedasticity [81] [82]. This will reflect the behavior of the random errors in the model and it can be castoff with functions that are either linear or non-linear in the parameters. It works by adding extra weights to each data points, into the fitting criterion. The size of the weight identifies the exactness of the information contained in the associated observation.

The method of ordinary least squares assumes that there is constant variance in the errors. Weighted Least Squares is a method which can be used when the ordinary least squares assumption of constant variance in the errors is violated. The model under consideration is described as follows,

$$Y = X\beta + \epsilon \tag{5.7}$$

where now ϵ is assumed to be (multivariate) normally distributed with mean vector 0 and nonconstant variance-covariance matrix.

$$\begin{pmatrix} \sigma_1^2 & 0 & \dots & 0 \\ 0 & \sigma_1^2 & \dots & 0 \\ & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \end{pmatrix} \tag{5.8}$$

Define the reciprocal of each variance, σ_i^2, as the weight, $w_i=1/\sigma_i^2$, then let matrix **W** be a diagonal matrix containing these weights:

$$w = \begin{pmatrix} w_1 & 0 & \dots & 0 \\ 0 & w_2 & \dots & 0 \\ & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & w_n \end{pmatrix} \tag{5.9}$$

The weighted least squares estimate is then

$$\hat{\beta}_{WLS} = \arg\min_{\beta} \sum_{i=1}^{n} \epsilon_i^{*2} \tag{5.10}$$

Improving weighted fitting criterion to determine parameter facilitates weights to calculate each observation influence on the corresponding final parameter estimates. Hence each observation weight is relative to remaining observations weights. It is proved that different sets of absolute weights are able to have equal effects. Weights are able to be calculated using the following methods,

1. Weighted least squares are used when the variance is non-constant.

2. If the variance of the ith observation is σ_i^2, then weights $w_i = \frac{1}{\sigma_i^2}$ are used to determine the standard errors of coefficients, with $w_i = \frac{1}{\sigma_i^2}, \sum w_i e_i^2 = \sum \frac{1}{\sigma_i^2} e_i^2 = \sum \left(\frac{e^i}{\sigma^i}\right)^2$
3. The sum of squares of the standardized errors is minimized to obtain the parameter estimates when weights = 1/variance.

Building the Model

The coefficients of the linear kernel are computed by taking the set of feature vectors as the input matrix for all the three cases character, word, paragraph text using the above regression method and the respective linear kernels are defined. The modified linear kernel WLK is built-in Support Vector Machine algorithm and employed here to build writer identification models. SVM with WLK kernel is implemented by tuning the regularization parameter C from 1 to 10. It is found that the regularization parameter reaches a stable state for C = 1 in case of TWINC dataset, whereas C=5 and C=10 in cases of TWINW, TWINP datasets respectively. Three independent writer identification models have been built based on TWINC, TWINW and TWINP datasets.

Performance Evaluation

The performance of writer identification models based on SVM with WLK kernel is evaluated using precision, recall, F-measure, accuracy and compared with the results of SVM linear kernel based Writer identification models. The predictive accuracies of classifiers for values of C=1, 5, 10 is shown in Table 5.1. With the stable values of C=1, 5, 10 for TWINC, TWINW, TWINP datasets respectively, the models are tested. The performance results of WLK kernel have been compared with linear kernel in terms of various measures and are given in Table 5.2 and the comparative accuracy is illustrated in Fig. 5.1.

Table 5.1 Results of SVM with WLK Kernel by Tuning C-Regularization Parameter

Dataset	TWINC			TWINW			TWINP		
Measures	**C=1**	**C=5**	**C=10**	**C=1**	**C=5**	**C=10**	**C=1**	**C=5**	**C=10**
Accuracy (%)	72.3	70.6	71.4	75	77.2	72	79.2	80.4	90.8
Time Taken (in secs)	0.03	0.03	0.03	0.02	0.02	0.01	0.532	0.493	0.521

Table 5.2 Performance Comparison of Linear SVM and WLK - SVM Based Models

Measures	Kernels	Datasets		
		TWINC	**TWINW**	**TWINP**
Accuracy (%)	**Linear SVM**	70.6	75	88
	WLK-SVM	72.3	77.2	90.8
Time Taken (in secs)	**Linear SVM**	0.02	0.03	0.942
	WLK-SVM	0.03	0.02	0.521
Precision	**Linear SVM**	0.689	0.706	0.942
	WLK-SVM	0.733	0.771	0.915
Recall	**Linear SVM**	0.827	0.748	0.989
	WLK-SVM	0.968	0.889	0.831
F-measure	**Linear SVM**	0.751	0.726	0.964
	WLK-SVM	0.834	0.826	0.871

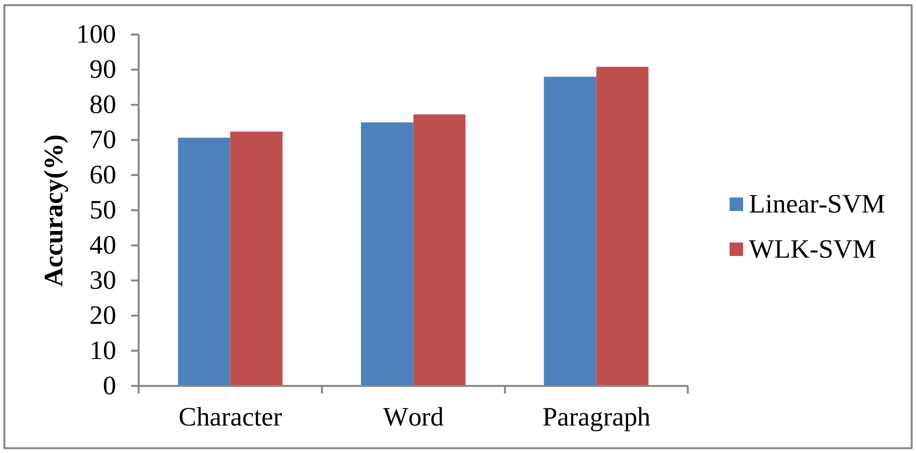

Fig. 5.1 Performance Comparison of Linear SVM and WLK - SVM Based Models

Ten output classes are taken to plot the ROC curves. The more each curve squeezes the left and top edges of the plot, the better the classification. The ROC charts derived for three datasets are illustrated in Fig. 5.2 – Fig. 5.4.

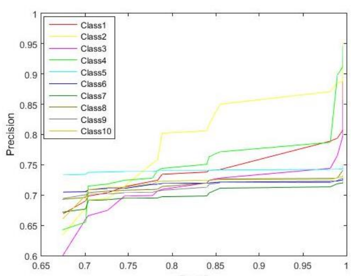

Fig. 5.2 ROC for TWINC Dataset

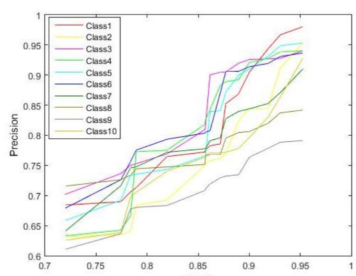

Fig. 5.3 ROC for TWINW Dataset

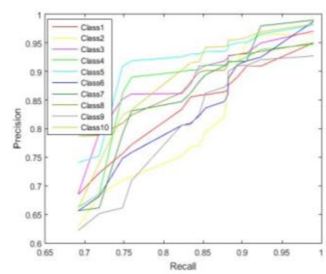

Fig. 5.4 ROC for TWINP Dataset

It is observed that from ROC curve for dataset TWINC, class 4 have high precision of 0.95 and class 3 is curved low at 0.60 whereas for dataset TWINW, class 1 have high precision of 0.97 and class 9 is curved low at 0.62. In case of dataset TWINP, class 7 have high precision of 0.99 and class 9 is curved low at 0.64.

Findings

Linear kernel is a dot product of features with less computational complexity. The system optimizes only C regularization parameter in linear kernel which makes it faster than other kernels. It is found that the linear kernel yields the accuracy of 88% in TWINP dataset, 75% in TWINW dataset and 70.6% in TWINC dataset where as SVM with modified linear kernel shows 90.8% in TWINP dataset, 77.2% in TWINW dataset and 72% in TWINC dataset which confirms that WLK offers high accuracy in writer prediction than linear kernel. The weighted linear kernel determines the optimum hyperplane with less computational complexity and achieved better performance. Due to less computational complexity, the time taken to train the model is very less which will help to overcome the situation even if the data is unstructured. The prominent accuracy of 90.8% is achieved when the modified weighted linear kernel based SVM is trained with vectors of paragraph text images.

5.2. MODEL III - WRITER IDENTIFICATION MODEL THROUGH SVM WITH BAYESIAN LINEAR KERNEL

In this case, linear kernel is extended to develop another kernel with Bayesian Linear Regression (BLR) and promoted the significance of linear kernel. BLR is a parameter estimation method based on Bayesian inference statistical analysis that is used to estimate the regression parameter ‘α’ given Equation (5.4). The parameter will act as the co-efficient of the modified linear kernel named as BLK linear kernel. The constant vector ‘α’ is of dimension equal to number of samples in the training dataset and each element is a parameter added to the sum of the squares of the features. This method of parameter estimation is described below.

Bayesian Linear Regression

Bayesian Linear Regression is a method for linear regression in which the statistical analysis is commenced within the outline of Bayesian inference. In the regression model errors with normal distribution, and if a specific prior distribution is anticipated results are seen that are characterized by model's parameters [83] posterior probability distributions. The Bayesian Linear Regression in statistical modeling normally consists of the following two steps:

The posterior distribution of model parameters w, which is proportional to the product of the BLR is a parameter estimation method based on Bayesian inference statistical analysis that is used

$$p(w|D) \propto p(w)p(D|w) \tag{5.11}$$

The strategy of the predictive distribution of y* given any new predictors x*. Now the prediction is obtained by integrating over the posterior distribution of w as follows:

$$p(y^*|x^*),D = \int p(y^*|x^*,w)p(w|D)dw \tag{5.12}$$

The predictive distribution obviously quantifies the uncertainty associated with the model parameters, and it delivers the expected way to construct the prediction intervals. The parameters are updated according to the following equations based on the interpretation made over Bayesian learning.

$$\mu_n = (X^T X + \Lambda_0)^{-1}\left(\Lambda_0\mu_0 + X^T X\hat{\beta}\right) = (X^T X + \Lambda_0)^{-1}(\Lambda_0\mu_0 + X^T y) \tag{5.13}$$

$$\Lambda_n = (X^T X + \Lambda_0) \tag{5.14}$$

$$a_n = a_0 + \frac{n}{2} \tag{5.15}$$

$$b_n = b_0 + \frac{1}{2}(y^T y + \mu_0^T\Lambda_0\mu_0 - \mu_n^T\Lambda_n\mu_n) \tag{5.16}$$

The Bayesian approach is one where the data is supplemented with extra information as prior probability distribution. Previously it was believed that parameters are clubbed along with data's likelihood function as per Bayes theorem. This is done to generate posterior belief about parameters β and σ. The prior can take different functional forms depending on the domain and the availability of information.

Building the Model

The coefficients of the linear kernel are computed by taking the set of feature vectors as the input matrix for all the three cases character, word, paragraph text using the above regression method and the respective linear kernels are determined. The modified linear kernel BLK is built-in Support Vector Machine algorithm and employed here to build writer identification models.

SVM with BLK kernel is implemented by tuning the regularization parameter C from 1 to 10. It is found that the regularization parameter reaches a stable state for C = 5 in case of TWINC dataset, whereas C=1 in case of both TWINW, TWINP datasets. Three independent writer identification models have been built based on TWINC, TWINW and TWINP datasets.

Performance Evaluation

The performance of writer identification models based on SVM with BLK is evaluated using precision, recall, F-measure, accuracy. The predictive accuracies of classifiers for C=1, 5, 10 is shown in Table 5.3. With the stable values of C= 1 for TWINC dataset and C = 1 for both TWINW, TWINP datasets, the models are tested. The performance results of BLK-SVM have been compared with linear SVM and WLK-SVM in terms of various measures and are given in Table 5.4 and the comparative accuracy is illustrated in Fig. 5.5.

Table 5.3 Results of SVM with BLK Kernel by Tuning C-Regularization Parameter

Datasets	**TWINC**			**TWINW**			**TWINP**		
Measures	**C=1**	**C=5**	**C=10**	**C=1**	**C=5**	**C=10**	**C=1**	**C=5**	**C=10**
Accuracy (%)	71	74.6	72.3	80.1	76.1	74	92.3	86.4	88.9
Time Taken (in secs)	0.02	0.03	0.02	0.02	0.02	0.02	0.660	0.591	0.0591

Table 5.4 Performance Comparison of Linear SVM,
WLK – SVM and BLK - SVM Based Models

Measures	**Kernels**	**Datasets**		
		TWINC	**TWINW**	**TWINP**
Accuracy (%)	Linear SVM	70.6	75	88
	WLK-SVM	72.3	77.2	90.8
	BLK-SVM	74.6	80.1	92.3
Time Taken (in secs)	Linear SVM	0.02	0.03	0.942
	WLK-SVM	0.03	0.02	0.521
	BLK-SVM	0.03	0.02	0.660
Precision	Linear SVM	0.689	0.706	0.942
	WLK-SVM	0.733	0.771	0.915
	BLK-SVM	0.812	0.723	0.708
Recall	Linear SVM	0.827	0.748	0.989
	WLK-SVM	0.968	0.889	0.831
	BLK-SVM	0.494	0.542	0.952
F-measure	Linear SVM	0.751	0.726	0.964
	WLK-SVM	0.834	0.826	0.871
	BLK-SVM	0.614	0.619	0.812

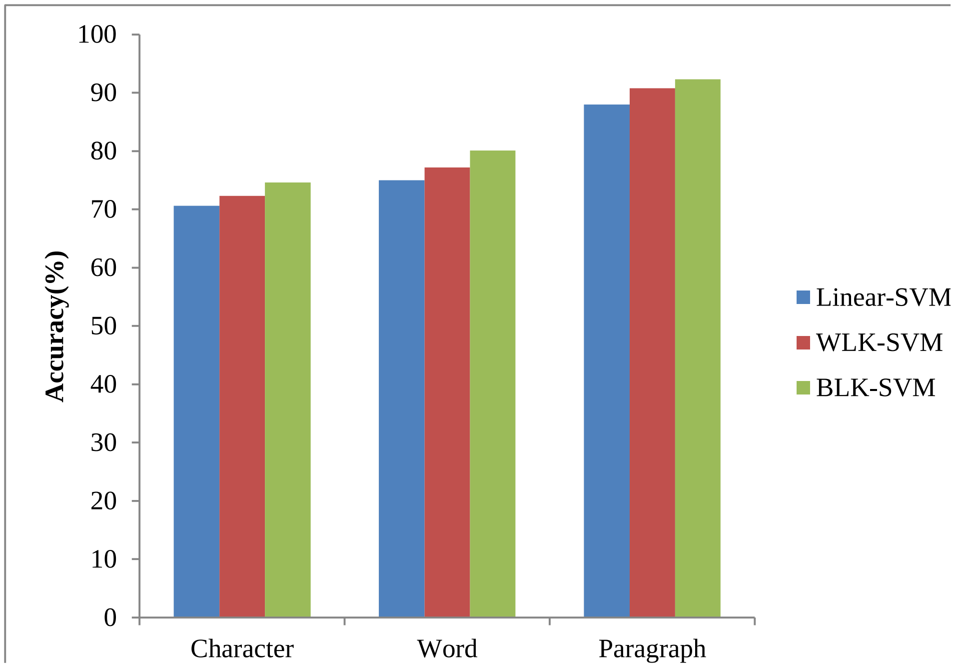

Fig. 5.5 Performance Comparison of Linear SVM, WLK – SVM and BLK - SVM Based Models

Receiver Operating Characteristic (ROC) curves for a sample of ten output classes is plotted. The more each curve squeezes the left and top edges of the plot, the better the classification. The ROC charts derived for three datasets are illustrated in Fig. 5.6 to Fig. 5.8.

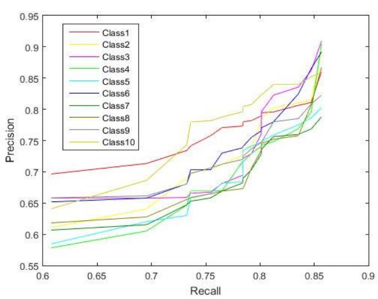

Fig. 5.6 ROC for TWINC Dataset

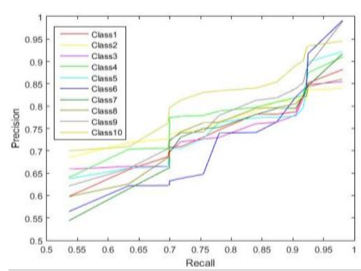

Fig. 5.7 ROC for TWINW Dataset

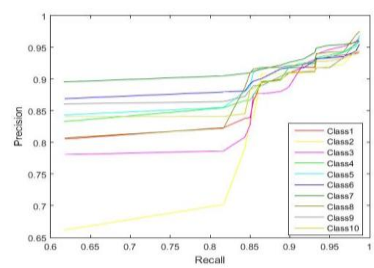

Fig. 5.8 ROC for TWINP Dataset

It is observed that from ROC curve, for dataset TWINC, class 3 have high precision of 0.90 and class 4 is curved low at 0.58 whereas for dataset TWINW, class 6 have high precision of 0.98 and class 7 is curved low at 0.54. In case of dataset TWINP, class 8 has high precision of 0.96 and class 2 is curved low at 0.66.

Findings

Training and evaluating a Bayesian linear SVM on writer's dataset is easily linearly separable, the SVM is able to find a margin that perfectly separates the training data, which also generalizes very well to the test set. It is found that the linear kernel yields the accuracy of 88% in TWINP dataset, 75% in TWINW dataset and 70.6% in TWINC dataset where as WLK – SVM shows 90.8% in TWINP dataset, 77.2% in TWINW dataset and 72% in TWINC dataset. In this case BLK – SVM shows 92.3% in TWINP dataset, 80.1% in TWINW dataset and 74.6% in TWINC dataset which confirms that BLK offers high accuracy in writer prediction than linear and WLK – SVM kernels. Also Bayesian Linear Kernel determines the optimum hyperplane with less computational complexity and achieved better performance. The prominent accuracy of 92.3% is achieved when the Bayesian Linear Kernel based SVM is trained with vectors of paragraph text images.

5.3. MODEL IV - WRITER IDENTIFICATION MODEL THROUGH SVM WITH PRINCIPAL COMPONENT KERNEL

In this experiment another linear kernel is proposed using Principal Component Regression which determines principal component that are used to estimate the coefficients of the linear kernel. Principal components regression reduces the standard errors by adding a degree of bias to the regression estimates. The coefficient 'α' given in Equation 5.4 is determined using the following method and the linear kernel with this coefficient is referred as PCK.

Principal Component Regression

Principal Component Regression is a method for evaluating multiple regression data that suffer from multicollinearity. This is a form of data disturbance and if inherent corresponding statistical inferences drawn on the basis of data are not considered as being reliable. Once multicollinearity takes place, least squares are computed as impartial; however corresponding variances are large and may vastly differ from the true value. It causes multiple errors in the estimation of parameter. The PCR method consists of the following three major steps:

1. PCA must be carried out on the data matrix for whom observation has been already carried out for explanatory variables to compute the main components, and generally a subset is chosen, based on certain criteria, of main components so computed for further purposes.

2. Observed vector must be regressed from selected principal components as covariates outcomes, by deploying ordinary least squares regression in order to procure the vector of estimated regression coefficients.

3. Based on the selected PCA loadings now transform this vector back to the scale of the actual covariates, the eigenvectors corresponding to the selected principal components to obtain final PCR estimator having dimension equal to total covariates in order to estimate regression coefficients that are specific to the original model.

PCR is well known technique as elucidated upon before is based on the classical PCA. It takes in to consideration the linear regression model in order to compute or predict outcome that is drawn on the basis of covariates. This though cannot be generalized so easily to the kernel machine setting as regression function will not be linear in existing covariates. However this can be considered as a part of the Reproducing Kernel Hilbert Space that is correlated to any arbitrary, symmetric positive-definite kernel.

Principal Component Kernel

Principal Component Kernel is a linear kernel [84] defined using the coefficients derived from PCR, thus facilitating linear kernel based SVM model. PCK is useful when the variance of the feature matrix cannot be well explained with a linear hyperplane. Instead of directly calculating nonlinear principal components, the feature matrix is implicitly mapped into a higher dimensional kernel space where a higher dimensional hyperplane can better fit the direction of highest variance. Therefore, given a p-dimensional random vector $x = (x_1,..x_p)^t$ with covariance matrix $\sum$ and assume that $\sum$ is positive definite. Let $V = (v_1, \dots v_p)$ be a $(p \times p)$ matrix with orthogonal column vectors that is $v_i^t, v_i = 1$ where $i = 1, \dots p$ and$V^t = V^{-1}$. The linear transformation

$$z = V^t x \tag{5.17}$$

$$z_i = v_i^t x \tag{5.18}$$

The variance of the random variable z_i is

$$Var(Z_i) = E[v_i^t x x^t v_i] = v_i^t \sum v_i \tag{5.19}$$

Maximizing the variance Var (Z_i) under the conditions $v_i^t v_i = 1$ with Lagrange gives

$$\phi_i = v_i^t \sum v_i - a_i(v_i^t v_i - 1) \tag{5.20}$$

Setting the partial derivation to zero, get

$$\frac{\partial \phi_i}{\partial v_i} = 2 \sum v_i - 2a_i v_i = 0 \tag{5.21}$$

Which is?

$$\left(\sum - a_i I\right) v_i = 0 \tag{5.22}$$

In matrix form

$$\left(\sum V\right) = VA \tag{5.23}$$

of

$$\left(\sum V\right) = VAV^T \tag{5.24}$$

where $A = diag(a_1, a_2, \cdots, a_p)$. The principal components are orthogonal to all the other principal components since A is a diagonal matrix. Performance of PCR will help to achieve more reliable estimates.

The overall working principle of the proposed system consists of four major steps: In the first step, Principal Components Analysis (PCA) on X, using the PCA function are carried out and two principal components are retained. PCR is then linear regression of the response variable in these two components. Then each variable is normalized by its standard deviation with the variables have very different amounts of variability. After which a PCR model with two principal components are fitted in the second step. In order to make the PCR results easier need to interpret in terms of the original spectral data, the regression coefficients are transformed into uncentered variables. In the last step, final PCK estimator is obtained by performing Ordinary Least Squares Regression while estimating the regression coefficients.

Step 1: PCA function

```
[PCALoadings, PCAScores, PCAVar] = pca (X,'Economy',false);
```

Step 2: Perform PCR function

```
betaPCR = regress(y-mean(y), PCAScores(:,1:2));
```

Step 3: Perform PCK kernel function

```
betaPCR = PCALoadings(:,1:2)*betaPCR;
betaPCR = [mean(b1) - mean(K)*betaPCR; betaPCR];
yfitPCR = [ones(length(b1),1) K]*betaPCR;
```

Step 4: Final OLS regression

```
b = mvregress(K,b1)
disp('b size');size(b)
for i = 1:size(K,2)
K1(:,i) = (b.*K(:,i));
End
disp('Kval'); size(K1)
```

End Process

Building the Model

The coefficients of the linear kernel are computed by taking the set of feature vectors as the input matrix for all the three cases character, word, paragraph text using the above regression method and the respective linear kernels are defined. The new linear kernel PCK is used in Support Vector Machine algorithm and PCK - SVM is implemented using MATLAB to build writer identification models. SVM with PCK kernel is experimented by tuning the regularization parameter C from 1 to 10. It is found that the regularization parameter reaches a stable state for C = 10 in case of TWINC dataset, C=5 for both TWINW, TWINP datasets. Three independent writer identification models have been built based on three training datasets.

Performance Evaluation

The performance of writer identification models based on SVM with PCK kernel is evaluated using precision, recall, F-measure, accuracy. The predictive accuracies of classifiers for C=1, 5, 10 is shown in Table 5.5. With the stable values of C= 10 for TWINC dataset and C = 5 for both TWINW, TWINP datasets, the models are tested. The performance results of PCK – SVM have been compared with linear SVM, WLK - SVM and BLK - SVM in terms of various measures and are tabulated in Table 5.6 and the comparative accuracy is illustrated in Fig. 5.9.

Table 5.5 Results of SVM with PCK Kernel by Tuning C-Regularization Parameter

Dataset	**TWINC**			**TWINW**			**TWINP**		
Measures	**C=1**	**C=5**	**C=10**	**C=1**	**C=5**	**C=10**	**C=1**	**C=5**	**C=10**
Accuracy (%)	86.6	84.2	86.6	85.5	90.4	89.3	91.6	94.9	93.4
Time Taken (in secs)	0.02	0.03	0.02	0.03	0.03	0.03	0.590	0.658	0.652

Table 5.6 Performance Comparison of Linear SVM,
WLK – SVM, BLK – SVM and PCK – SVM Based Models

Measures	Kernels	Datasets		
		TWINC	TWINW	TWINP
Accuracy (%)	Linear SVM	70.6	75	88
	WLK-SVM	72.3	77.2	90.8
	BLK-SVM	74.6	80.1	92.3
	PCK-SVM	86.6	90.4	94.9
Time Taken (in secs)	Linear SVM	0.02	0.03	0.942
	WLK-SVM	0.03	0.02	0.521
	BLK-SVM	0.03	0.02	0.660
	PCK-SVM	0.02	0.03	0.658
Precision	Linear SVM	0.689	0.706	0.942
	WLK-SVM	0.733	0.771	0.915
	BLK-SVM	0.812	0.723	0.708
	PCK-SVM	0.872	0.748	0.989
Recall	Linear SVM	0.827	0.748	0.989
	WLK-SVM	0.968	0.889	0.831
	BLK-SVM	0.494	0.542	0.952
	PCK-SVM	0.963	0.958	0.976
F-measure	Linear SVM	0.751	0.726	0.964
	WLK-SVM	0.834	0.826	0.871
	BLK-SVM	0.614	0.619	0.812
	PCK-SVM	0.915	0.935	0.959

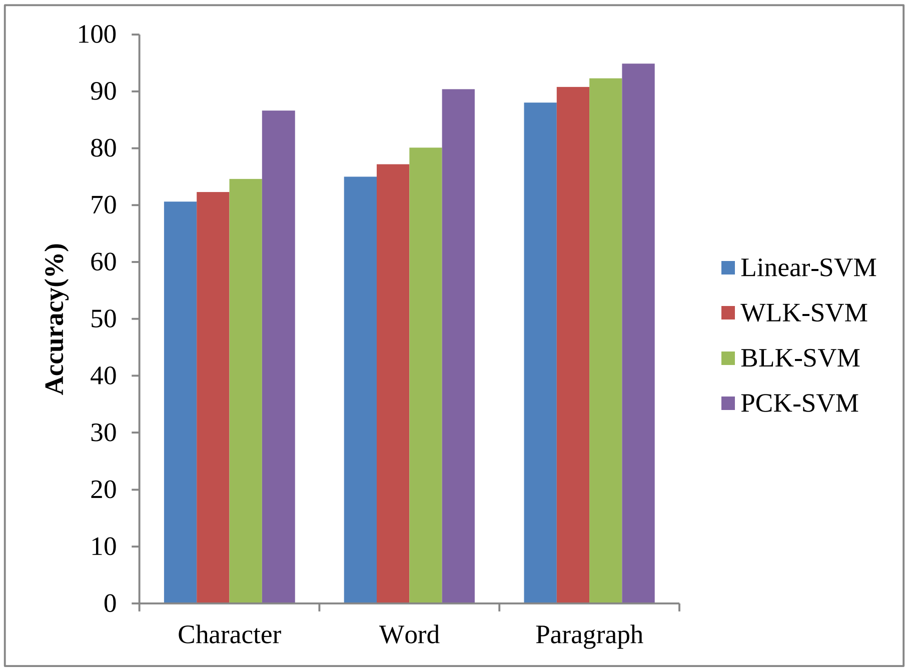

Fig. 5.9 Performance Comparison of Linear SVM,
WLK – SVM, BLK – SVM and PCK – SVM Based Models

Receiver Operating Characteristic curves for ten output classes are plotted. The more each curve squeezes the left and top edges of the plot, the better the classification. The ROC charts derived for three datasets are illustrated in Fig. 5.10 to Fig. 5.12.

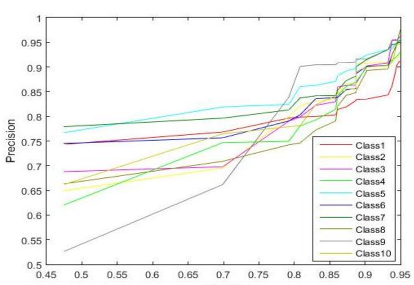

Fig. 5.10 ROC for TWINC Dataset

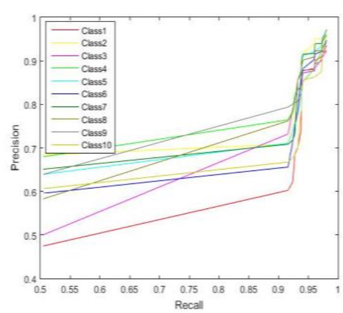

Fig. 5.11 ROC for TWINW Dataset

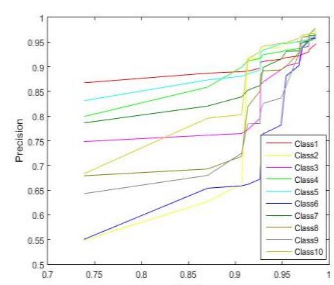

Fig. 5.12 ROC for TWINP Dataset

It is observed that ROC curve for TWINC dataset, class 8 have high precision of 0.98 and class 9 is curved low at 0.52 whereas for TWINW dataset, class 7 have high precision of 0.98 and class 1 is curved low at 0.48. In case of TWINP dataset, class 8 have high precision of 0.99 and class 2 is curved low at 0.52.

Findings

It is found from the results of experiments described in section 5.1, 5.2, Bayesian Linear Kernels shows better performance than linear SVM and WLK - SVM. This experiment of PCK – SVM kernel based prediction model produces 94.9% accuracy in TWINP dataset, 90.4% in TWINW dataset and 86.6% in TWINC dataset which confirms that PCK – SVM offers better performance than other three forms of linear kernels in writer prediction. The prominent accuracy of 94.9% is achieved when the Principal Component Kernel based SVM is trained with vectors of paragraph text images.

From the observations made through the sections 5.1, 5.2, 5.3 it is clearly proved that the descriptive features captured from paragraph text images are more contributive in developing the writer identification model. Utilizing parameter estimation method in redefining the linear kernel is recognized as the best strategy in SVM implementation. Eventually the linear kernel formulation through Principal Component Regression accomplished effectiveness in identifying the writer with minimum time taken and less computational complexity.

5.4. SUMMARY

This chapter discussed about the formulation of three different kernels such as WLK, BLK, PCK and its usage with Support Vector Machine for building writer identification models. The process of building models and the performance evaluation of SVM implementations with all three kernels are described. The results of experiments with new kernels are presented with tables and charts. The comparison of linear with all three new kernels is made and the corresponding findings are also discussed clearly in this chapter.

Remarks

1. A paper titled “Bayesian Linear Regression Co-efficients for SVM linear kernel to identify writers” has been published in Journal of Advanced Research in Dynamical and Control Systems (JARDCS) **(Scopus Indexed)**, Vol. 10, 01-Special Issue, 2018

2. A paper titled “Hybrid Linear Kernel with PCA in SVM Prediction Model of Tamil Writing Pattern” has been accepted for publication in International Journal of Simulation Systems, Science & Technology **(Scopus Indexed)**.