Face Recognition system Using SVM classifier(LIBSVM)

H BALAJI(194102307)

1. ABSTRACT

Feature representation and classification are two key steps for face recognition. A novel method for face recognition was presented based on combination of PCA (Principal Component Analysis) and SVM (Support Vector Machine). PCA was used for dimensionality reduction and SVM were used for classification. The experiments were implemented on fetch lfw people data set in sklearn datasets.our approach improved face recognition rate.

2. INTRODUCTION

SVMs (Support Vector Machines) have been developed in the framework of statistical learning theory, and have been successfully applied to a number of applications, ranging from time series prediction, to face recognition, to biological data processing for medical diagnosis.SVM can well resolve some practical problems such as small sample size, nonlinear, highdimensional problems.

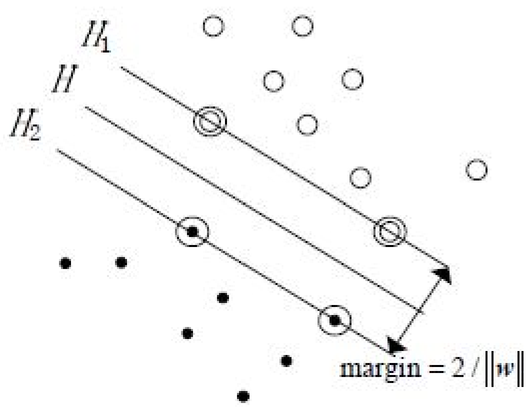
3. METHODOLOGY

3.1 PCA

PCA is a standard technique used to approximate the original data with lower dimensional feature vectors. The face image of PCA algorithm is seen as a random vector. By solving the scatter matrix eigenvalue problem of training samples, a group of new orthogonal bases are obtained to indicate the subspace spanned by training samples, and the features extracted are just project-vectors of the face images. When the subspace orthogonal bases are arranged as image array, they will show the shape of faces, so orthogonal bases are called eigenfaces.

3.2 SVM FOR BINARY CLASSIFICATION

SVM method was derived from optimal hyperlane in the linearly separable case. Consider the example of a two-class as shown in the fig



Here H is the correct classification line, H1 and H2 are parallel lines of H. H1 and H2 get through the nearest points from H. H is the only line which can both separate the data and maximize the margin (the distance between the hyperplane and the nearest data point of each class), namely optimal line. In high-dimension space, the optimal line becomes optimal hyperlane. Assumed

$$(x_i, y_i) \quad i = 1, 2, \cdots, n \quad x \epsilon R^d \tag{S1}$$

be the sample set

$$y \epsilon \{-1, +1\} \tag{S2}$$

are class labels,the linear discriminate function of the d-dimension space is

$$g(x) = w \cdot x + b \tag{S3}$$

The corresponding hyper plane equation is

$$w \cdot x + b = 0 \tag{S4}$$

And then the optimal hyperlane problem turns into the following constrained quadratic programming problem: Minimize

$$\phi(w) = \frac{1}{2}\|w\|^2 = \frac{1}{2}(w \cdot w) \tag{S5}$$

subjected to

$$y_i[w \cdot x_i + b] - 1 \geq 0 \quad i = 1, 2, \cdots, n \tag{S6}$$

The optimal classification function can be obtained by solving the above problem using legrangian multipliers function

$$L(w, b, \alpha) = \frac{1}{2}\|w\|^2 - \sum_{i=1}^{n} \alpha_i (y_i(w^t x_i + b) - 1) \tag{S7}$$

where α_i is the Lagrange Multiplier with respect to the i^{th}inequality We can get the following conditions of optimality, through differentiating L(w,b,α) with respect to w and b setting the results equal to 0:

$$\frac{\partial L(w, b, \alpha)}{\partial w} = 0 \quad \frac{\partial L(w, b, \alpha)}{\partial b} = 0 \tag{S8}$$

thus, we obtain

$$w = \sum_{i=1}^{n} \alpha_i y_i x_i \quad \sum_{i=1}^{n} \alpha_i y_i = 0 \tag{S9}$$

The corresponding Dual Problem can be inferred by means of substituting (S9) in (S7)

$$\max_{\alpha} \quad w(\alpha) = \sum_{i=1}^{n} \alpha_i - \sum_{i=1}^{n} \sum_{j=1}^{n} \alpha_i \alpha_j y_i y_j x_i^t x_j \tag{S10}$$

subjected to

$$\sum_{i=1}^{n} \alpha_i y_i = 0. \quad \alpha_i \geq 0 \quad i = 1, 2, \cdots, n \tag{S11}$$

The following equation (s12) gives the Karush-Kuhn-Tucker (KKT) complementary condition:

$$\alpha_i[y_i(w \cdot x_i + b) - 1] = 0 \quad i = 1, 2, \cdots, n \tag{S12}$$

As a result of it, only the support vectors (x_i, y_i) (the closest samples to the optimal separating hyperplane, which determine the maximal margin) correspond to the nonzero α_i s and α_i s are zero for the points away from the hyper plane.The optimal weight vector w^* after determining optimal the optimal Lagrange multipliers α_i^* s is

$$w^* = \sum_{i=1}^{n} \alpha_i^* y_i x_i \tag{S13}$$

therefore, the corresponding optimal biasb^* can be expressed as follows:

$$b^* = 1 - w^* \cdot x_s \quad for \quad y_s = +1 \tag{S14}$$

3.2.1 Solving Linearly Inseparable Cases by Kernel based trick SVM

For linearly inseparable cases, kernel trick is another commonly used technique. An appropriate kernel function, which is based on the inner product between the given samples, is to be defined as a nonlinear transformation of samples from the original space to a feature space with higher or even infinite dimension so as to make the problems linearly separable. That is, a complicated classification problem cast in a high-dimensional space nonlinearly is more likely to be linearly separable than in a low-dimensional space. Actually, we can adopt a nonlinear mapping function

$$\Phi : X \to \phi(X) \quad R^d \to F \tag{S15}$$

to map data X in original (or primal) space into a higher (ever infinite) dimension space F , such that the mapped data in new feature space is more likely linearly separable. Thus we can extend the separating hyperplane function into the following form.

$$\begin{aligned} \mathrm{w}\cdot\phi(x_i) + b_i \geq 1 \quad \forall y_i = +1, \\ w \cdot \phi(x_i) + b_i \leq -1 \quad \forall y_i = -1; \end{aligned}$$

(S16)

that is,

$$y_i[w \cdot \phi(x_i) + b] - 1 \geq 0 \quad i = 1, 2, \cdots, n \tag{S17}$$

The separating hyperplane is

$$w \cdot \phi(x) + b = 0 \tag{S18}$$

The optimization problem in such case is

$$\min_{w,b} \frac{1}{2}\|w\|^2 \tag{S19}$$

subjected to

$$y_i[w \cdot \phi(x_i) + b] \geq 1 \quad i = 1, 2, \cdots, n \tag{S20}$$

Using the same mathematical trick in linearly separable SVM, we get the corresponding Dual Function

$$\max_{\alpha} \quad w(\alpha) = \sum_{i=1}^{n} \alpha_i - \sum_{i=1}^{n}\sum_{j=1}^{n} \alpha_i\alpha_j y_i y_j \phi(x_i)^t \phi(x_j) \tag{S21}$$

subjected to

$$\sum_{i=1}^{n} \alpha_i y_i = 0. \quad \alpha_i \geq 0 \quad i = 1, 2, \cdots, n \tag{S22}$$

The KKT complementary condition is

$$\alpha_i[y_i(w \cdot \phi(x_i) + b) - 1] = 0 \quad i = 1, 2, \cdots, n \tag{S23}$$

The optimal weight vector w^* after determining optimal the optimal Lagrange multipliers is

$$w^* = \sum_{i=1}^{n} \alpha_i^* y_i \phi(x_i) \tag{S24}$$

$$b^* = 1 - \sum_{sv} \alpha_i^* y_i \phi(x_i)\phi(x_s) \quad for \quad y_s = +1 \tag{S25}$$

Fortunately, the inner product like the form of $\phi(x_i)\phi(x_s)$ can be instituted with , $K(x_i, x_j)$that is,

$$K(x_i, x_j) = \phi(x_i)\phi(x_j) \tag{S26}$$

The corresponding constrained optimization problem in kernel form can be now formulated as

$$\max_{\alpha} \quad w(\alpha) = \sum_{i=1}^{n} \alpha_i - \sum_{i=1}^{n}\sum_{j=1}^{n} \alpha_i\alpha_j y_i y_j K(x_i, x_j) \tag{S27}$$

subjected to

$$\sum_{i=1}^{n} \alpha_i y_i = 0. \quad \alpha_i \geq 0 \quad i = 1, 2, \cdots, n \tag{S28}$$

Accordingly, the optimal classifier is

$$f(x) = \sum_{i=1}^{n} \alpha_i^* y_i \phi(x_i) + b^* \tag{S29}$$

where

$$b^* = 1 - \sum_{i=1}^{n} \alpha_i^* y_i K(x_i, x_s) \tag{S30}$$

The most frequently used kernel functions are

$$\begin{aligned} K(x,y) =& x \cdot y \\ K(x,y) =& (x \cdot y + c)^h \\ K(x,y) =& exp - \{\frac{1}{2\sigma^2} \|x - y\|^2\} \end{aligned}$$

The third one is the RBF kernel function, which is frequently used

4. RESULTS

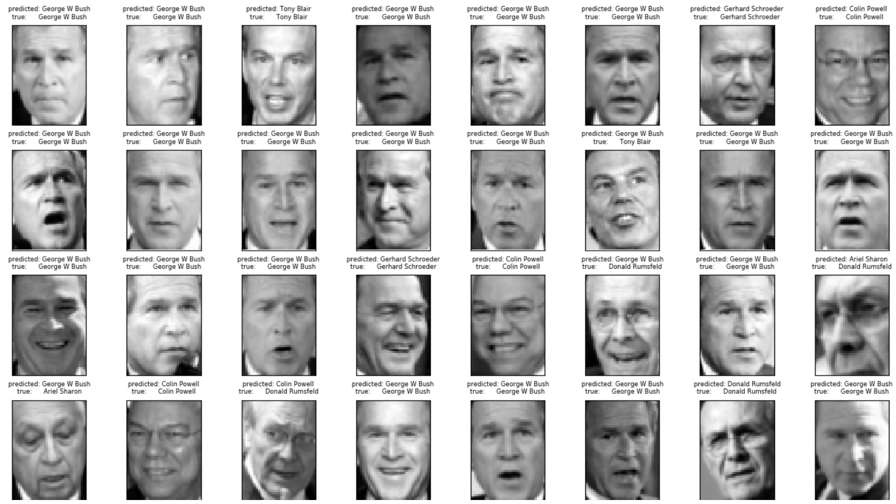
SVM was originally designed for binary classification. Face recognition is a multi-class classification problem. There are two basic methods for face recognition with SVMs: oneagainst- one and one-against-all. The one-against-one method is classification between each pair classes. The one-against-all is classification between each class and all the rest classes. In our experiments the one-against-all method was used for classification.

fetch lfw people face data set of 7 individuals is one of the sklearn data sets , 184 face images of each individual ,138 images among 184 images of each individual were taken for training and the rest of them for testing

Images of first individual was taken and marked as positive samples, the all images of other training samples as negative samples. Both positive samples and negative samples were taken as input samples to train a SVM classifier to get corresponding support vectors and optimal hyperplane. The SVM was labeled as SVM1. In turn we can get the SVM for every individual and labeled as SVM2, . . . , SVM7 respectively. The maximum accuracy attained through this is 83%

Table S1. CLASSIFICATION REPORT

Person	precision	recall	f1 score	support
Ariel Sharon	0.70	0.54	0.61	13
Colin Powell	0.77	0.85	0.81	60
Donald Rumsfeld	0.79	0.70	0.75	27
George W Bush	0.85	0.92	0.88	146
Gerhard Schroeder	0.79	0.76	0.78	25
Hugo Chavez	0.80	0.53	0.64	15
Tony Blair	0.90	0.78	0.84	36



predicted: George W Bush
true: George W Bush
predicted: George W Bush
true: George W Bush
predicted: Tony Blair
true: Tony Blair
predicted: George W Bush
true: George W Bush
predicted: George W Bush
true: George W Bush
predicted: George W Bush
true: George W Bush
predicted: Gerhard Schroeder
true: Gerhard Schroeder
predicted: Colin Powell
true: Colin Powell
predicted: George W Bush
true: George W Bush
predicted: George W Bush
true: George W Bush
predicted: George W Bush
true: George W Bush
predicted: George W Bush
true: George W Bush
predicted: George W Bush
true: George W Bush
predicted: George W Bush
true: Tony Blair
predicted: George W Bush
true: George W Bush
predicted: George W Bush
true: George W Bush
predicted: George W Bush
true: George W Bush
predicted: George W Bush
true: George W Bush
predicted: George W Bush
true: George W Bush
predicted: Gerhard Schroeder
true: Gerhard Schroeder
predicted: Colin Powell
true: Colin Powell
predicted: George W Bush
true: Donald Rumsfeld
predicted: George W Bush
true: George W Bush
predicted: Ariel Sharon
true: Donald Rumsfeld
predicted: George W Bush
true: Ariel Sharon
predicted: Colin Powell
true: Colin Powell
predicted: Colin Powell
true: Donald Rumsfeld
predicted: George W Bush
true: George W Bush
predicted: George W Bush
true: George W Bush
predicted: George W Bush
true: George W Bush
predicted: Donald Rumsfeld
true: Donald Rumsfeld
predicted: George W Bush
true: George W Bush