FOREX Trend Classification using Machine Learning Techniques

AREEJ ABDULLAH BAASHER, MOHAMED WALEED FAKHR
Computer Science Department
Arab Academy for Science and Technology
Cairo, EGYPT
baasher_areej@yahoo.com, waleedf@aast.edu

Abstract :- Foreign Currency Exchange market (Forex) is a highly volatile complex time series for which predicting the daily trend is a challenging problem. In this paper, we investigate the prediction of the High exchange rate daily trend as a binary classification problem, with uptrend and downtrend outcomes. A large number of basic features driven from the time series data, including technical analysis features are generated using multiple history time windows. Various feature selection and feature extraction techniques are used to find best subsets for the classification problem. Machine learning systems are tested for each feature subset and results are analyzed. Four important Forex currency pairs are investigated and the results show consistent success in the daily prediction and in the expected profit.

Keywords: - Technical analysis, Feature selection, Feature extraction, Machine-learning techniques, Bagging Trees, SVM, Forex prediction.

1 Introduction

This paper is about predicting the Foreign Exchange (Forex) market trend using classification and machine learning techniques for the sake of gaining long-term profits. Our trading strategy is to take one action per day, where this action is either buy or sell based on the prediction we have. We view the prediction problem as a binary classification task, thus we are not trying to predict the actual exchange rate value between two currencies, but rather, if that exchange rate is going to rise or fall. Each day there are four observed rates, namely, the "Open", "Close", "Low" and "High". In this work, we focus on predicting the direction of the "High".

Forex daily exchange rate values can be seen as a time series data and all time series data forecasting and data mining techniques can be used to do the required classification task.

In time series analysis, it is always a challenge to determine the required history window used by the classification or forecasting system to do its prediction. In this paper, we have taken an approach of providing features from multiple time windows ranging from one day up to 30 days. This of course results in a number of features larger than using a single time window. Processing of the raw time domain daily values is done to produce the basic features used in the feature selection and extraction steps. This processing involves calculation of technical analysis and other time and frequency domain features over multiple time windows with a total of 81 basic features. Our approach is to let feature selection and feature extraction techniques find the best set of features for the classification task. Feature selection techniques choose a subset of the basic features while feature extraction techniques find features in new projected spaces. In this paper, two feature selection and six feature extraction techniques are used which all aim at finding feature subsets to enhance the classification performance.

For each feature subset, three supervised machine learning classifiers are used, namely, radial basis function neural network (RBF), multilayer perceptron neural network (MLP) and support vector machine (SVM). This gives a large array of different feature subsets and different classifiers. Comparison between these different systems is done based on two factors. The first is the percentage classification performance on the test data. The second is a novel function we call the percentage normalized profit (*PNP*), which represents the ratio between the accumulated profits using the predicted trends versus the accumulated profit using perfect predictions along the duration of the test period. We have selected four Forex pairs which are namely the USD/YEN, USD/EGP, EURO/EGP and EURO/SAR (where EGP is the Egyptian pound and SAR is the Saudi Riyal). The remaining of this paper is organized as follows. Section 2 discusses the nature of the Forex data, and the most prominent approaches used in the literature to represent this data. In particular, the Technical Analysis (TA) approach is explained and the relevant TA features used in this work are outlined. Section III focuses on the particular features used in this work, where we have combined the ideas of the TA and some signal analysis ideas, over multiple window sizes to get 81 basic features.

Section 4 outlines the feature selection and feature extraction techniques used in this work. For feature selection, an SVM-based, and a Bagging decision tree ensemble methods are used. For feature extraction five methods are used, namely, PCA per class method (PCAClass), linear discriminant analysis between clusters and class's method (LDA Cluster/Class), and between pairs of clusters (LDA Cluster/Cluster), maximally collapsing metric learning algorithm (MCML) and neighborhood components analysis (NCA). In section 5, the array of experiments and experimental results obtained are detailed and discussed. Finally, section 6 gives summary, conclusion and future work directions.

2 FOREX Data Processing

2.1 Nature of FOREX

Forex is trading of currencies in the international market. The Forex market operates 24 hours a day, 5 days a week where the different trading sessions of the European, Asian and American Forex markets take place. The nature of Forex trading has resulted in a high frequency market that has many changes of directions which produce advantageous entry and exit points. The Forex signal is an economic time series that is composed of two composite structures; a long-term trend and a short-term high frequency oscillation as shown in Figure 1.

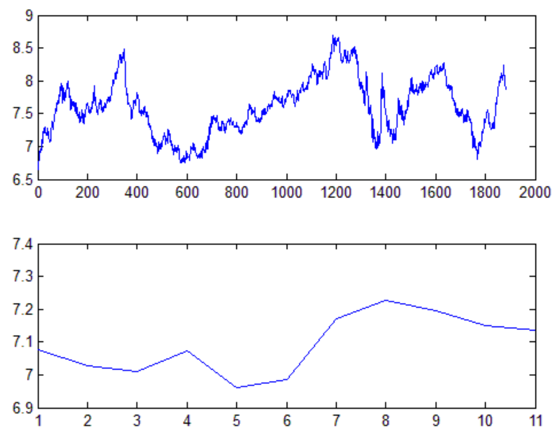

Fig. 1 Long-Term and Short-Term daily trend structures

2.2 FOREX Prediction Review

Economic time series prediction techniques are classified into two main categories; namely, techniques that try to predict the actual value of the rate exchange or the actual returns value and techniques that try to predict the trend direction (uptrend, downtrend and sideways).

Exchange rates prediction is one of the most challenging applications of modern time series forecasting. The rates are inherently noisy, non-stationary and deterministically chaotic [1]. One general assumption is made in such cases is that the historical data incorporate all of those behaviors. As a result, the historical data is the major player in the prediction process [1, 2].

For more than two decades, Box and Jenkins' Auto-Regressive Integrated Moving Average (ARIMA) technique [2] has been widely used for time series forecasting. Because of its popularity, the ARIMA model has been used as a benchmark to evaluate some new modeling approaches [3]. However, ARIMA is a general univariate model and it is developed based on the assumption that the time series being forecasted are linear and stationary [4]. Recently, the generalized autoregressive conditional Heteroskedasticity (GARCH) modeling has been used in Stock and Forex forecasting [5-7] with better results than the ARIMA models. Also, hidden Markov Models (HMMs) have been used recently for the same purpose [8, 9].

Neural networks and support vector machines (SVM), the well-known function approximators in prediction and classification, have also been used in Forex forecasting [10-13].

In a recent paper [6], the GARCH (5,1) model was compared to Random Walk, Historical mean, Simple Regression, moving average (MA), exponentially weighted moving average (EWMA) for Forex prediction. The GARCH was the best model in the prediction task. In this paper, GARCH prediction outcomes are taken as part of the basic feature set.

2.3 Technical Analysis Features

Technical analysis is defined as the use of numerical series generated by market activity, such as price and volume, to predict future trends in that market. The techniques can be applied to any market with a comprehensive price history [14]. Technical analysis features are based on statistical examination of price and volume information over a given period of time (different time windows).

The objective of this examination is to produce signals that predict where and in which direction the price may move in the near Future.

Technical analysis features are divided into leading indicators, lagging indicators, stochastic, oscillators and indexes [15].

In the literature, the total number of technical analysis features is more than 100 of which the most popular ones are used in this work as explained in the next section [16].

3 FOREX Feature Generation

3.1 Technical Analysis Features Used

In this work we have used 11 technical analysis feature methods. The TA features used are namely; Stochastic oscillator, Momentum, Williams %R, Price Rate of change (PROC), Weighted Closing Price (WPC), Williams Accumulation Distribution Line (WADL) and the Accumulation Distribution Oscillator (ADOSC), Moving Average Convergence, Divergence (MACD), Commodity

Channel Index (CCI), Bollinger Bands, and the Heiken-Ashi candles indicator. The details of the mathematical calculations for these features are found in [16, 17]. In each method we have employed different time windows to allow for a multi-scale feature generation. This has led to a total of 46 technical analysis features generated with the different window durations used for each feature as shown in Table 1.

Table 1: Technical Analysis Features Applied using Different Time Windows

Feature number	Type	Duration (days)	Num. of features	Type	Duration (days)
1- 12	Momentum	3,4,5,8,9,10	36-40	ADOSC	1,2,3,4,5
13-24	Stochastic	3,4,5,8,9,10	41	MACD	15,30
25 -29	Williams	6,7,8,9, 10	42	CCI	15
30-33	PROC	12,13,14,15	43,44	Bollinger Bands	15
34	WCP	15	45,46	Heiken Ashi	15
35	WADL	15			

3.2 GARCH Features

GARCH stands for generalized autoregressive conditional Heteroscedasticity [5-7]. It is a mechanism that models time series that has time-varying variance. It has shown promising results in both Stock and Forex prediction. In this paper we trained a GARCH (3, 1) model using a 15 day window to get two features; namely, mean and variance of the time series each day. This is done by moving the 15-day window by one day step. The training is done on 1600 days training data for each Forex currency.

3.3 Signal Processing Features:

Signal processing inspired features were added to the TA features. 5 different averages are calculated as shown in Table 2 based on 1 and 2 days. 10 different slopes are calculated based on the High value over durations ranging from 3 to 30 days as shown in Table 2. Fourier series fitting using the form:
(a0 + a1*cos(x*w) + b1*sin(x*w)) is calculated over 10, 20 and 30 days to give 12 features. Finally, Sine wave fitting using the form
(a0 + b1*sin(x*w)) is done for 3 and 5 days windows. Both the Fourier and the Sine fitting were done after removing the trend assuming it was linear.

Table 2: Signal Processing Features Applied using Different Time Windows

Feature number	Type	Duration (days)
49	High price average	2
50	Low price average	2
51,52	High, low average	1,2
53	Trading day price average	1
54-63	Slope	3,4,5,10,20,30
64-75	Fourier	10,20,30
76-81	Sine	3,5

4 Feature Selection and Feature Extraction Techniques

4.1 Feature Selection

Variable and feature selection have become the focus of much research in areas of application for which datasets with tens or hundreds of thousands of variables are available. Feature selection is used to improve the prediction performance of the predictors and to provide faster and more robust estimation of model parameters. In this paper, two discriminative feature selection techniques are used, namely and SVM-based technique, and a bagging decision tree technique.

4.1.1 SVM-based Selection Technique

In this paper we used a feature selection technique which is an extension of the SVM-RFE algorithm [18-20]. It finds the best feature subset which minimizes SVM classification error. In this paper we used different subsets starting from 5 up to 20 features.

4.1.2 Bagging Trees Selection Technique:

A decision tree forest is an ensemble (collection) of decision trees whose predictions are combined to make the overall prediction for the forest [21] forming a bootstrap aggregation or bagging trees. A decision tree forest grows a number of independent trees in parallel, and they do not interact until after all of them have been built.
Bagging trees are excellent tools for feature selection. For each feature, the out-of-bag mean squared error after removing that feature is averaged over all trees. This is repeated for each feature to come up with the best features list [22]. In this paper we used a 100 trees forest and a threshold of 0.3. The top 20 features were used.

4.2 Feature Extraction

Feature extraction is transforming the basic set of features to a new domain such that the information in the basic features is compressed in a smaller number of features. Some feature extraction techniques are generative (e.g., PCA) and some are discriminative (e.g., LDA) [23].

4.2.1 MCML

Maximally Collapsing Metric Learning algorithm (MCML) is a quadratic transformation, which relies on the simple geometric intuition that if all points in the same class could be mapped into a single location in feature space and all points in other classes mapped to other locations then this would result in a good approximation of the equivalence relation [24]. In this technique 5, 10, 15 and 20 extracted features were tried.

4.2.2 NCA

Neighborhood components analysis is a supervised learning method for clustering multivariate data into distinct classes according to a given distance metric over the data [25] which is similar in concept to the MCML. 5, 10, 15 and 20 features were tried.

4.2.3 Class-based PCA

In this novel technique, we have a PCA transform per class. The data is mapped using both transforms and the resulting features are concatenated. In this paper we used 4, 5, and 6 features per transform giving 8, 10 and 12 total features. The motivation of this technique is to have a class-specific PCA transform, which would act as a filter to the opposite class.

4.2.4 Cluster-Class-based and Cluster-Cluster based LDA

The linear discriminant analysis (LDA) [23] maximizes the inter-class distances while minimizing the intra-class distances by doing a linear projection. The LDA produces number of features less than number of classes – 1. In order to use it in this 2-class problem, we have used two novel ideas. A cluster-class-based LDA and a cluster-cluster based LDA. In both cases, the data of each class is clustered using the K-means algorithm. In the former technique, the LDA is applied between each cluster and the opposite class, producing 2*K features. In the latter, the LDA is applied between each pair of opposite clusters, producing K^2 features. In this paper, K is 3 and the number of features was 6 and 9 respectively.

5 Experimental Setup and Results

5.1 Data preparation

In this paper we experimented with 4 Forex datasets; (USD/YEN), (USD/ EGP), (EURO/EGP) and (EURO/EGP). Each set contains daily High, Low, Open, Close and Volume. The datasets represent a time period of 1852 days from April 2003 to August 2010 excluding the weekends. The available data was divided into training (1600 days) and test (252 days) sets.

5.2 Classification Process

In this paper, the feature subsets resulting from the selection and extraction stage are used to train 2-class classifiers. The two classes are either uptrend or downtrend. The first 1600 days are used in training. The targets are calculated as: **sign** ($\Delta(t+1)$) where the $\Delta(t+1)$ is the difference between the rates at

ISBN: 978-1-61804-039-8

day (t+1) and *day (t)*.

By nature of Forex, this Δ is a variable depending on the volatility and variance changes in the time series.

Our approach is to allow the classifier to predict the sign of the $\Delta(t+1)$ regardless of its value. The motivation is that if $\Delta(t+1)$ absolute value is high; the features would have strong evidence for correct classification and vice versa. And, for the system to be profitable, it should classify high changes accurately while it is not so important to focus on small changes.

In this paper, three classifiers are used, namely; RBF, MLP, and SVM. The percentage performance of each classifier is calculated on the test data (252 days) by simply looking at the % correct classification.

5.3 Percentage normalized profit (PNP):

We have developed a normalized profit criterion which measures the success of the system predictions to follow the actual directions and values of $\Delta(t+1)$. Two classifiers may have the exact same classification performance but may give significantly different profits since one of them is predicting the important changes correctly and the other is not.

Let the parameter *P (t)* be *+1* if the prediction is correct and *P (t)* is *-2* if not. The reason is that, when a wrong decision is taken, the correct action is not taken, and further, since we have to take an action each day, the wrong action is taken. Let the ideal profit represents the profit in case all the predicted values were true. The normalized profit over a time duration T will be:

$$PNP = \frac{\sum_T P(t).\Delta(t+1)}{\sum_T \Delta(t+1)} \quad (1)$$

5.4 Experimental Results

Each of the feature selection and feature extraction techniques discussed in section 4 produce a subset of features. In all cases (except for the cluster-based extraction), we have tried subsets ranging from 5 to 20 features. Each subset was used to train machine-learning techniques, namely SVM, MLP and RBF. We experimented with different parameters for each classifier in order to reach the best performance on each dataset.

5.4.1 Feature Selection Results:

In this section we show the best percentage performance resulted from the feature selection techniques as shown in Fig. 2 below. Table 3 represents the best results for each dataset using the 3 classifiers over the test period.

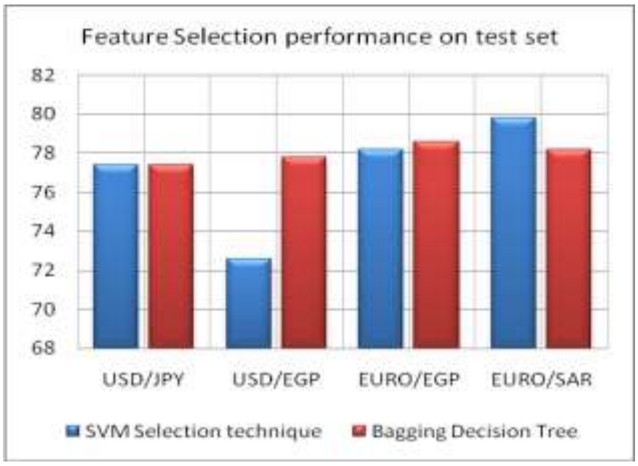

Fig. 2 Best Percentage Performance for Feature Selection Techniques

5.4.2 Feature Extraction Results:

In this section we show the best percentage performance resulted from the feature extraction techniques for all the datasets as shown in Fig 3 below. Table 4 represents the best results for each dataset using the three classifiers over the test period. For the feature extraction experiments using MCML, NCA and PCA, we used different number of mapped features (5, 10, 15 and 20). In Table 4 we show the best results for each technique among the datasets.

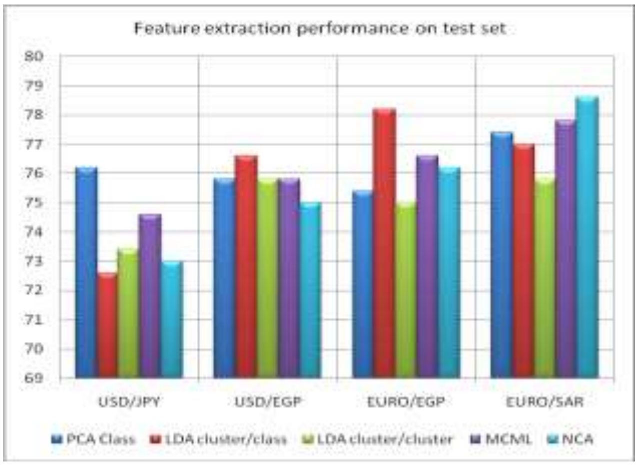

Fig. 3 Best Percentage Performance for Feature Extraction Techniques

5.4.3 Prediction-based versus Ideal Profits:

Fig. 4 below shows: $\sum_T P(t).\Delta(t+1)$ and $\sum_T \Delta(t+1)$, this is drawn over T from 1 to 252 to see how the actual profit is progressing compared to the ideal profit for the best profit resulted in each technique. Table 5 represents the techniques with the best percentage performance over T from 1 to 252 with the best PNP over the same time period.

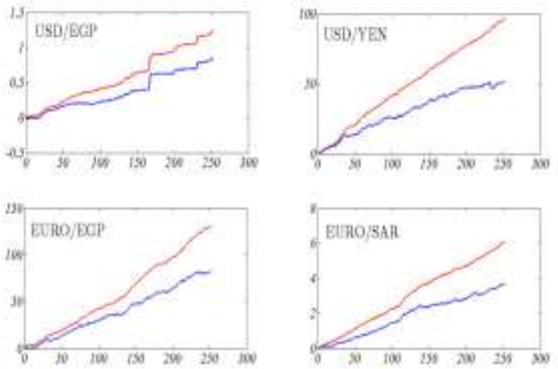

Fig. 4 Actual Profit vs. Ideal Profit for the best technique

Table 5: Techniques Resulted in Best Performance and the Highest Percentage Normalized Profit for Each Dataset

Dataset	Best Technique	Recognition rate	PNP
USD/JPY	Bagging Trees using (SVM) as classifier	77.8	61.9
USD/EGP	Bagging Trees using (MLP) as classifier	77.8	67
EURO/EGP	Bagging Trees using (SVM) as classifier	78.6	67.7
EURO/SAR	SVM – Based using (MLP) as classifier	79.8	60.8

Table 3: Test Results(Percentage Performance)
Comparison using Machine Learning Techniques for Feature Selection Techniques

Dataset / Selection technique	USD/JPY			USD/EGP			EURO/EGP			EURO/SAR		
	SVM	MLP	RBF	SVM	MLP	RBF	SVM	MLP	RBF	SVM	MLP	RBF
Selection using SVM	77.4	75.4	75.8	72.6	72.2	72	78.2	77.8	77	77.8	**79.8**	77.8
Selection using bagging tress	**77.8**	73	73	74.6	**77.8**	75.8	**78.6**	76.6	77	78.2	77	77.8

Table 4: Results(Percentage Performance)
Comparison using Machine Learning Techniques for Feature Extraction Techniques

Dataset / Extraction technique	USD/JPY			USD/EGP			EURO/EGP			EURO/SAR		
	SVM	MLP	RBF	SVM	MLP	RBF	SVM	MLP	RBF	SVM	MLP	RBF
PCA/Class	**76.2**	72.6	73	71.4	75.8	74.6	72.6	75	75.4	77	77.4	76.2
LDA(Cluster/ Class)	69.4	72.6	69	69.4	**76.6**	75	75	74.2	**78.2**	61.1	77	74.6
LDA(Cluster/ Cluster)	70.6	70.6	73.4	66.7	75.8	75.4	75	73	75	72.1	73.4	75.8
MCML	74.6	73	71.4	74.2	75	75.8	76.6	76.6	75.8	77.8	77.8	77.4
NCA	73	72.7	71.8	71	75	73.4	76.3	75.4	75	77	**78.6**	77

ISBN: 978-1-61804-039-8

6 Conclusion

In this paper we investigated Forex multiple time windows technical analysis and signal processing features for predicting the high rate daily trend. The prediction is posed as a binary classification problem for which the system predicts whether the high rate is going up or down. SVM-based and bagging trees feature selection methods as well as five feature extraction techniques are all used to find best feature subsets for the classification problem. Machine learning classifiers (RBF, MLP and SVM) are all trained using the many different feature subsets and the results are shown and compared based on the percentage classification performance. Also, a percentage normalized profit function is invented which shows the ratio between the profit based on the predicted trends and that using ideal predictions. The results show that the proposed systems can be used for practical trading purposes.

References

[1] Deboeck, ***Trading on the Edge: Neural, Genetic and*** *F u q* ***Systems for Chaotic Financial Markets,*** New York Wiley, 1994.

[2] G. E. P. Box and G. M. Jenkins, ***Time Series Analysis: Forecasting and Control,*** Holden-Day, San Francosco, CA

[3] H. B. Hwamg and H. T. Ang, "A Simple Neural Network for ARhlA@,q) Time Series," *OMEGA: lnt Journal of Management Science,* vol. 29, pp 319-333, 2002

[4] G. Peter Zhang, "Time series forecasting using a hybrid ARIMA and neural network model", Department of Management, Georgia State University, January 2002

[5] Ching-Chun Wei, "Using the Component GARCH Modeling and Forecasting Method to Determine the Effect of Unexpected Exchange Rate Mean and Volatility Spillover on Stock Markets", International Research Journal of Finance and Economics, © EuroJournals Publishing, Inc. 2009

[6] Dr. S S S Kumar, "Forecasting Volatility – Evidence from Indian Stock and Forex Markets", 2006

[7] Yu, J., (2002) "Forecasting Volatility in the New Zealand Stock Market," *Applied Financial Economics*, 12: 193-202.

[8] Patrik Idvall, Conny Jonsson, " Algorithmic Trading
Hidden Markov Models on Foreign Exchange Data ", Department of Mathematics, Link¨opings Universitet, LiTH - MAT - EX - - 08 / 01 - - SE, 2008

[9] Dong-xiao Niu; Bing-en Kou; Yun-yun Zhang;
Coll. of Bus. & Adm., North China Electr. Power Univ., Beijing, China, "Mid-long Term Load Forecasting Using Hidden Markov Model", pp 481 – 483, 2009

[10] L Cao, Francis E H Tay, "Financial Forecasting Using Support Vector Machines", Neural Computing Applications (2001) vol. 10, Springer, pp: 184-192

[11]Joarder Kamrwzaman' and *Ruhul* A Sarkeg, "Forecasting of currency exchange rates using ANN: A case study ", IEEE Int. Conf. Neural Networks *8* Signal Processing, 2003

[12] Zhang, G.P.; Kline, D.M., "Time-Series Forecasting With Neural Networks", Neural Networks, IEEE, pp 1800 – 1814, 2007

[13] Woon-Seng Gad and Kah-Hwa Ng2, "Multivariate FOREX Forecasting using Artificial Neural Networks", IEEE Xplore, 2010

[14]http://www.earningmoneyonlinearticles.com/index.php/trade-the-forex-market-without-any-challenges/

[15] Steven B. Achelis, Technical Analysis from A to Z, McGraw-Hill, United States, 2nd Edition, 2001

[16] http://forexsb.com/wiki/indicators/source/start

[17]http://www.stockinvestingideas.com/technical-analysis/technical-indicators.htm

[18] Isabelle Guyon, Andr´e Elisseeff, " An Introduction to Variable and Feature Selection", Journal of Machine Learning Research, pp 1157-1182, 2003

[19] Alain Rakotomamonjy, " Variable Selection Using SVM-based Criteria", *JMLR*, 1357-1370, 2003

[20] B. Ghattas, A. Ben Ishak, " An Efficient Method for Variables Selection Using SVM-Based Criteria".

[21] http://www.dtreg.com/

[22] (MATLAB reference for the tree forest feature selection)

[23]http://homepage.tudelft.nl/19j49/Matlab_Toolbox_for_Dimensionality_Reduction.html

[24] Amir Globerson, Sam Roweis, " Metric Learning by Collapsing Classes", Advances in Neural Information Processing, 2006

[25]http://en.wikipedia.org/wiki/Neighbourhood_components_analysis

[26] P. Belhumeur, J. Hespanha, and D. Kriegman "Eigenfaces vs. Fisherfaces: Recognition Using Class Specific Linear Projection". *IEEE PAMI, 1997.*