I.J. Intelligent Systems and Applications, **2012, 12, 26-31**
Published Online November 2012 in MECS (http://www.mecs-press.org/)
DOI: 10.5815/ijisa.2012.12.04

Margin Based Learning: A Framework for Acoustic Model Parameter Estimation

Syed Abbas Ali
Computer & Information Systems Engineering, N.E.D University of Engineering & Technology.
saaj@neduet.edu.pk

Najmi Ghani Haider
Computer Science & Information Technology, N.E.D University of Engineering & Technology.
nghaider@gmail.com

Mahmood Khan Pathan
Computer Science & Information Technology, N.E.D University of Engineering & Technology.
mkpathan@hotmail.com

Abstract—Statistical learning theory has been introduced in the field of machine learning since last three decades. In speech recognition application, SLT combines generalization function and empirical risk in single margin based objective function for optimization. This paper incorporated separation (misclassification) measures conforming to conventional discriminative training criterion in loss function definition of margin based method to derive the mathematical framework for acoustic model parameter estimation and discuss some important issues related to hinge loss function of the derived model to enhance the performance of speech recognition system.

Index Terms—Statistical Learning, Generalization Capability, Empirical Risk, Discriminative Training, Test Risk Bound

I. Introduction

Hidden Markov Model (HMM) is the most successful statistical pattern recognition approach to model the speech signal as stochastic pattern, over the last two decades. HMM parameters are estimated from training data according to certain criterion. There are several criteria for training HMMs, including the Maximum Likelihood Estimation (MLE) criterion [1] and a group of criteria called discriminative training (DT) such as Maximum Mutual Information (MMI) Estimation [2], Minimum Classification error (MCE) [3] and Minimum Word/Phone Error (MWE/MPE) [4]. The MLE criterion does not focus on minimizing classification error, while Discriminative training methods minimize classification error in training data as a model estimation criterion. In speech recognition, most of the discriminative training methods directly minimize the empirical risk on the training data sample and does not focus on the model generalization. Training and testing mismatches can often be measured by generalization capability of machine learning algorithm. SLT [5] defines the concept of test risk bound, which is bounded by the summation of two terms: An empirical risk (i.e risk on the training set) and a generalization function. Incorporating the margin concept into Hidden Markov Modeling (Acoustic Model) for speech recognition, Margin based DT methods demonstrate superior capability over any other conventional DT methods to improve generalization ability of the acoustic model by improving the margin of the model [6,7]. The main focus of this study is the loss function definition of conventional discriminative training methods in the empirical risk formulation and loss function definition of margin based methods based on misclassification measure (i.e distance between correct and competing hypothesis) to derive the separation measures corresponding to conventional DT criteria. In this paper separation (misclassification) measures corresponding to MMI, MCE and MWE/MPE are incorporated in loss function definition of Margin based method to develop the mathematical frame work for acoustic model parameter estimation. The derived mathematical framework combines the capability of conventional and margin based DT methods using concept of statistical learning theory. Rest of the paper is organized as follows. In section 2, discuss the framework of statistical learning theory for pattern recognition. Empirical risk formulation and loss functions corresponding to conventional discriminative training is presented in section 3. In section 4, we discuss the margin based DT criterions to understand the generalization problem of learning algorithms. We present our derived model for parameter estimation and discussed some issues related to hinge function to improve model generalization capability in section 5. Finally, the conclusion is drawn in section 6.

II. Statistical Learning Theory for Pattern Recognition

The main objective of statistical learning theory is to provide a framework for reading the problem of inference that is of gaining knowledge, making predictions, making decision or constructing models from set of data sample [8]. Statistical learning framework is associated with supervised learning for statistical pattern recognition. Supervised learning is method of machine learning to learn the function from the training data sample to construct an acoustic model for desired output. The standard framework of statistical learning problem can be defined as, consider a set of m training data set (x_1, y_1) , . . . (x_m , y_m) drawn from P(x , y) which is independent and identically distributed(iid) according to an unknown joint probability. By introducing the concept of loss function. We can define loss function "*L*" in classification problem as:

$$L(y, f(x,z)) = \begin{bmatrix} 0 & if y = f(x,z) \\ 1 & if y \neq f(x,z) \end{bmatrix} \quad (1)$$

Considered a risk function providing the true value of loss as follows:

$$R_{true}(z) = \int L\big(y, f(x,z)\big) dP(x,y) \quad (2)$$

The goal is to find out the function $f(x,z)$ that minimize the risk function $R_{true}(z)$.

z is the generalized parameter of function. In order to minimize the expected risk function in (2) on a test set drawn from the same distribution *P*(x, y), an empirical risk need to be obtained by induction principle and the expected risk function is substituted by Empirical Risk.

$$R_{emp}(f) = \frac{1}{m}\sum_{i=1}^{m} L(y_i, f(x_i, z)) \quad (3)$$

To directing the generalization capability of learning machine, there is a need to build a principle of induction for minimizing the risk function using small sample of training. A test risk bound $R_{true}(z)$ consist of empirical risk $R_{emp}(f)$ and VC have the dimension 'ɦ'(a measure of the capacity of set of function) and m(number of training sample). The probability "1-τ" having bound as:

$$R_{true}(z) \leq R_{emp}(f) + \frac{\sqrt{ɦ\left(\log\left(\frac{2m}{ɦ}\right) + 1 - \log\left(\frac{\tau}{4}\right)\right)}}{m} \quad (4)$$

There is a possibly to minimize the test risk bound by directly minimizing the right hand side of (4). Generalization function cannot be directly minimized due to monotonic increasing function and computing difficulty of ɦ. Vapnik show that, VC dimension "ɦ" is bounded by decreasing function of margin and can be reduced by increasing the margin [5]. Equation (4) consists of two optimization function: generalization function and other one is empirical risk.

III. Loss functions for Conventional DT Criteria

Discriminative training methods recently gaining remarkable attention in the field of machine learning whereas it does not make any explicit attempt to model underlying distribution of dataset and it directly optimizes a mapping function from input training set to the required output. DT methods only adjust the decision boundary without making a data producer in the entire feature space [9].In this section; we will mathematically define the loss function concept in empirical risk minimization and provide the loss function of conventional discriminative training criteria which is used to minimize empirical risks. Discriminative training criteria for acoustic model are used to minimize the empirical risk of speech recognition system. Discriminative training directly minimize the risk on training data sample in the application speech recognition and formulation of empirical risk can be define in term of loss function,

$$R_{emp}(\Lambda) = \frac{1}{N}\sum_{t=1}^{N} l(O_t, \Lambda)$$

Where *N*is the total number of training utterances and $l(O_t, \Lambda)$ is a loss function for utterance O_t. $\Lambda = (\pi, A, B)$ is a parameter set representing initial state probability, state transition probability and observation probability.

3.1 Maximum Mutual Information Estimation (MMI)

The main goal of MMI criterion is to minimize mutual information between training data set (O_1 , O_2 , ….. O_T) and their corresponding transcription (W_1 ,W_2 , ….. W_T) to establish the tightest possible relation between training data and their corresponding model [10, 11]. MMI criterion widely used in speech recognition can be defined as:

$$\sum_{t=1}^{T} \log \frac{\mathrm{P}(O_t|W_t).\mathrm{P}(W_t)}{\sum_{\hat{W}_t} \mathrm{P}(O_t|\hat{W}_t).\mathrm{P}(\hat{W}_t)} \quad (5)$$

The loss function value of the object function in (5) can be defined as:

$$1 - \log \frac{P(O_t|W_t).P(W_t)}{\sum_{\hat{W}_t} P(O_t|\hat{W}_t).P(\hat{W}_t)} \quad (6)$$

3.2 Minimum Classification Error (MCE)

The main objective of MCE is to explicitly minimize the total error counts in training data sample [12,13]. In MCE formulation, misclassification error is constructed for each training utterances $O_t$are as follows:

$$\mathrm{d}(O_t, \Lambda) = -log\left[P(O_t|W_t).P(W_t)\right] + log\sum_{\hat{W}_t \neq W_t}\left[P(O_t|\hat{W}_t).P(\hat{W}_t)\right] \quad (7)$$

by plugging misclassified function into sigmoid function:

$$l\left(\mathrm{d}(O_t,\Lambda)\right)=\frac{1}{1+e^{-d(O_t,\Lambda)}} \quad (8)$$

The minimum classification error can be represented in the form of loss function:

$$\arg\min\sum_{t=1}^{T} l(d(O_t,\Lambda))$$

as shown in [14], substituting the misclassification measure in (7) into the sigmoid function (8), equivalent form of MCE criterion can be obtained as:

$$\arg\max\sum_{t=1}^{T}\frac{P(O_t|W_t).P(W_t)}{\sum_{\hat{W}_t}P(O_t|\hat{W}_t).P(\hat{W}_t)} \quad (9)$$

3.3 Minimum phone (word) error (MPE/MWE)

In MCE (Minimum classification error) formulation, speech recognition error measured in sentence level. Whereas in large vocabulary speech recognition, performance of the system measured in the form of sub-string or word level. The research work motivated by Povey and Woodland [15] has improved MCE criterion for the sub-string level defined as:

$$\arg\max\sum_{t=1}^{T}\frac{\sum_{W_t}P(O_t|\hat{W}_t).P(\hat{W}_t).A(W_t,\hat{W}_t)}{\sum_{\hat{W}_t}P(O_t|\hat{W}_t).P(\hat{W}_t)} \quad (10)$$

Where $A(W_t,\hat{W}_t)$: row accuracy count, which is defined to calculate accuracy of sub-string between two sentences true transcription of each utterance W_t and all possible string sequence of utterances $\hat{W}_t$. The loss function value of (10) which reflects sub string error can be written as:

$$1-\frac{\sum_{\hat{W}_t}P(O_t|\hat{W}_t).P(\hat{W}_t).A(W_t,\hat{W}_t)}{\sum_{\hat{W}_t}P(O_t|\hat{W}_t).P(\hat{W}_t)} \quad (11)$$

In spite of major progress in discriminative learning methods, there are some limitation in the discriminative training criteria such as computational complexity is quite high, it is fail to cope with the temporal dynamic of the speech signal and one of the most important issue in context with this article that conventional DT methods only minimize the empirical risk $R_{emp}(\Lambda)$.

IV. Margin Based DT Criteria

Despite of the some significant progress in discriminative training methods, many issues still unsolved. One of the prominent issue arises in DT methods related to HMM based recognition is the poor generalization capability. Although, DT methods improve HMM based acoustic model and dramatically reduce error in training data sample but DT methods don't perform well into new unseen test data set. Theoretical framework has been studied in the field of machine learning to understand the generalization problem of learning algorithms. To address the generalization problem, the margin concept is introduced in pattern classification and incorporated into HMM for speech recognition. Several approaches were proposed for margin maximization [5,6]. In this section we will focus on LME (large margin estimation) and SME(soft margin estimation) methods to understand the separation (misclassification) measures concept in multi-class separation margin and log likelihood ratio respectively.

4.1 Large Margin Estimation (LME)

Discriminative training criteria based on the principle of large margin classifier called as LME (Large Margin estimation) criterion. The main objective of the LME criterion is to estimate acoustic model parameter based on minimum margin maximization criterion of training sample in the direction of improve generalization ability and robustness in designing learning classifier. The parameter estimation based on maximize their minimum margin is well established in [16]. In LME, given a set of training sample denoted as D, consist of utterances as D=(O_1 , O_2 , ….. O_T) and the true transcription for all utterances as L= (W_1 ,W_2 , ….. W_T).The separation measure for each training sample W_t based on the fact that by increasing margin of classifier, generalization ability improve accordingly. $\mathrm{d}(O_t,\Lambda)$ is the separation (misclassification) measure defined as the difference between correct and the closest competing transcription. The separation (misclassification) margin for each training sample W_t can be written as:

$$\mathrm{d}(O_t,\Lambda)=log\left[\mathrm{P}\left(O_t|W_t\right).\mathrm{P}\left(W_t\right)\right]$$
$$-\max_{\hat{W}_t,\hat{W}_t\neq W_t} log[\mathrm{P}(O_t|\hat{W}_t).\mathrm{P}(\hat{W}_t)] \quad (12)$$

The acoustic model Λ are estimated based on minimum margin maximization criterion for all training data samples cab be represent as:

$$\Lambda_{\mathrm{LME}}=\arg\max_{\Lambda}\min_{O_t\epsilon \mathrm{D}}\mathrm{d}(O_t,\Lambda) \quad (13)$$

If $\mathrm{d}(O_t,\Lambda)\leq 0$, W_t will be incorrectly classified by model parameter set Λ and if $\mathrm{d}(O_t,\Lambda)>0$, W_t will be correctly classified by model parameter set Λ. In (12), the max is applied on all competing transcription $\hat{W}_t$ ($\hat{W}_t\neq W_t$) for W_t which may provide word graph and we can make use Softmax concept of MCE in (12) which gives:

$$\mathrm{d}(O_t,\Lambda)=log[\mathrm{P}(O_t|W_t).\mathrm{P}(W_t)]$$
$$-log[\sum_{\hat{W}_t\neq W_t}P(O_t|\hat{W}_t).\mathrm{P}(\hat{W}_t)]$$

by placing (12) in (13), the LME criterion represented as maxi-min optimization problem:

$$\Lambda_{\mathrm{LME}} = \arg\max_{\Lambda} \min_{X_t \epsilon \mathrm{D}} log \frac{P(O_t|W_t).P(W_t)}{\sum_{\hat{W}_t, \hat{W}_t \neq St} P(O_t|\hat{W}_t).P(\hat{W}_t)} \quad (14)$$

LME criterion in (14) is calculated based on the hypothesis that all training sample is perfectly recognized by current model [17]. From the above said statement, it can be concluded that LME updates models only with correctly classified data samples. While, misclassified sample may cause crucial impact on the learning classifier.

4.2 Soft Margin Estimation (SME)

Soft margin estimation [18] was proposed to make the direct use of an idea of margin in SVM [19] to improve the generalization capability by increasing margin. The framework of SME based on the concept of Statistical learning theory. The SLT is bounded by combination of two terms: an empirical risk function and generalization function. Here, we are interesting to define the separation (misclassification) measure formulation related to soft margin estimation. To define the separation measure of SME make use of log likelihood ratio (LLR) as discussed in [2] and defined as:

$$\mathrm{d}(O_t, \Lambda) = log\left[\frac{P(O_t|W_t)}{P(O_t|\hat{W}_t)}\right] \quad (15)$$

If LLR based separation (misclassification) measure $\mathrm{d}(O_t, \Lambda) > 0$ the classification will be correct, otherwise incorrect classification by model parameter set Λ. Where as $P(O_t|W_t)$ and $P(O_t|\hat{W}_t)$ are the likelihood values for the true and competing transcription respectively. To define more precise separation model define in (16) for each utterance, there is a need to select frames having different acoustic model labels in true and competing transcription. Selected frame will provide the discriminative information and separation (misclassification) measure for each utterance will be average value of frame LLRs. The formulation of the precise model can be representing as:

$$\mathrm{d}(O_t, \Lambda) = \frac{1}{n_t}\sum_j log\left[\frac{P(O_{tj}|W_t)}{P(O_{tj}|\hat{W}_t)}\right]\Gamma(O_{tj} \epsilon \mathrm{F_t}) \quad (16)$$

Where $\mathrm{F_t}$ is the frame set having frame with different labels in competing transcription, O_{tj} is the jth frame for utterance O_t, $\mathrm{n_t}$ is the number of frame in $\mathrm{F_t}$. Soft margin estimation use margin concept to enhance the generalization capability in learning classifier, if the shift cause due to mismatch of training and testing sample is less than margin, a true decision can be obtained by classifier. When the value of soft margin is greater than separation (misclassification) measures, loss will be happen and the loss function definition in term of hinge loss functions:

$$l((O_t, \Lambda) = \begin{cases} \rho - \mathrm{d}(O_t, \Lambda) \text{if } \rho > \mathrm{d}(O_t, \Lambda) \\ 0, \quad \text{otherwise} \end{cases} \quad (17)$$

Plugging the value of (16) into (17), we can get the loss function of soft margin estimation.

$$l(O_t, \Lambda) = \left\{\rho - \frac{1}{n_t}\sum_j log\left[\frac{P(O_{tj}|W_t)}{P(O_{tj}|\hat{W}_t)}\right]\Gamma(O_{tj} \epsilon \mathrm{F_t})\right\} \quad (18)$$

Test risk bound defined in (4) has two functions for optimization: margin maximization and empirical risk minimization, it is not tightly bound that why, it is not mandatory to follow Vapnik's theorem. The test risk bound can be estimated by the combination of two optimization function in a single object function based on soft margin estimation,

$$\Lambda_{\mathrm{SME}} = \frac{\lambda}{\rho} + \frac{1}{N}\sum_{t=1}^{N} l(O_t, \Lambda) \quad (19)$$

Introducing (17) into (19), the soft margin objective function can be reformulated as:

$$\Lambda_{\mathrm{SME}} = \frac{\lambda}{\rho} + \frac{1}{N}\sum_{t=1}^{N}(\rho - d(O_t, \Lambda))\Gamma(O_t \epsilon \mathrm{U}) \quad (20)$$

Γ is the indicator function and U is the set of utterance having separation (misclassification) measure less than soft margin. LME works only on correctly classified sample, whereas Soft margin estimation consider all training samples including both correctly classified and misclassified sample.

V. Acoustic Model Parameter Estimation

The separation (misclassification) measure is defined as the distance between true and competing hypothesis. In this section, separation measures corresponding to discriminative criteria are presented, which is obtained from the optimization objectives of MMI (Maximum mutual information estimation),MCE (Minimum classification error) and MWE/MPE (Minimum word/phone error) in (5),(9) and (10) respectively. The equations of separation (misclassification) measure can be represented as

$$\log \frac{P(O_t|W_t).P(W_t)}{\sum_{\hat{W}_t} P(O_t|\hat{W}_t).P(\hat{W}_t)} \quad (21)$$

$$\frac{P(O_t|W_t).P(W_t)}{\sum_{\hat{W}_t} P(O_t|\hat{W}_t).P(\hat{W}_t)} \quad (22)$$

$$\frac{\sum_{\hat{W}_t} P(O_t|\hat{W}_t).P(\hat{W}_t).A(W_t, \hat{W}_t)}{\sum_{\hat{W}_t} P(O_t|\hat{W}_t).P(\hat{W}_t)} \quad (23)$$

by placing (21), (22) & (23) in (20). We can estimate the acoustic model parameter in the context of margin based learning.

$$\Lambda_{\mathrm{SME}} = \frac{\lambda}{\rho} + \frac{1}{N}\sum_{t=1}^{N}(\rho - log\frac{P(O_t|W_t).P(W_t)}{\sum_{\hat{W}_t} P(O_t|\hat{W}_t).P(\hat{W}_t)})\ \Gamma(O_t \epsilon \mathrm{U}) \tag{24}$$

$$\Lambda_{\mathrm{SME}} = \frac{\lambda}{\rho} + \frac{1}{N}\sum_{t=1}^{N}(\rho - \frac{P(O_t|W_t)\ .P(W_t)}{\sum_{\hat{W}_t} P(O_t|\hat{W}_t).P(\hat{W}_t)})\Gamma(O_t \epsilon \mathrm{U}) \tag{25}$$

$$\Lambda_{\mathrm{SME}} = \frac{\lambda}{\rho} + \frac{1}{N}\sum_{t=1}^{N}\left(\rho - \frac{\sum_{\hat{W}_t} P(O_t|\hat{W}_t).P(\hat{W}_t).A(W_t,\hat{W}_t)}{\sum_{\hat{W}_t} P(O_t|\hat{W}_t).P(\hat{W}_t)}\right)\Gamma(O_t \epsilon \mathrm{U}) \tag{26}$$

(24), (25) and (26) representing the misclassification measure corresponding to conventional DT criteria which are incorporated in loss function definition of Margin based estimation for acoustic model parameter estimation.

One of the important aspects of (21), (22) and (23) is that, these equations are derived from the concept of soft margin estimation and hinge loss function used in SVM is defined as loss function to achieved improved speech recognition performance. The hinge function is susceptible to outliers and imposes no bound on the maximum penalty. There are two main issues related to hinge function need to be addressed here. First, hinge loss function performs well when the noise in training sample and kernel used for training is very small and appropriately tuned respectively. This is not the case in many real world data sample, so, even a few points vary away from the margin hyper plane can severely impact the cost of that hyper plane and in turn, affect the final result of optimization. This reflects the limitation of soft margin estimation, when the mismatch match between training and testing data increases. One approach in this direction is to investigate the different loss function for pattern recognition or develop a new loss function for minimum classification error to reduce the mismatch between training and testing data sample and improve the performance of the derived model. Secondly, any misclassified training sample contributes a support vector, which directly affects the time required for optimization and determines the label of the test sample. This may decreases the convergence speed and arises the shallow local optima problem. One interesting approach in this direction is to study the convex optimization methods in context with the derive framework to accelerate the convergence speed by reducing the free variables present in non-convex object function and avoid shallow local optimal point, because any local optimum always global optimal point in convex optimization problem.

VI. Conclusion

In this paper, we incorporate the separation (misclassification) measures corresponding to conventional discriminative training criteria in the loss function definition of margin based methods to derive the mathematical framework for acoustic model parameter estimation. This paper present initial study of the derived model. We have re-examined SME based discriminative learning framework and working on issues related to hinge function to reducing mismatch between training and testing data sample and time required for optimization due to misclassified training sample and identify two possible approaches related to hinge loss function. In future research work, authors have high expectation from this derived framework for achieving the goal of enhancing the generalization capability of robust speech recognition system.

References

[1] A.P. Dempster, N. M. Laird and D. B. Gopinath, "Maximum Likelihood from incomplete data via the EM algorithm," J. Roy.Stat.Soc., 39(1), 1-38, 1977.

[2] B. -H. Juang, W. Chou, and C.-H. Lee, "Minimum Classification Error rate methods for speech recognition," IEEE Trans. on Speech and Audio Proc., vol.5, no.3, pp.257-265, 1997.

[3] Y. Normandin, "Maximum Mutual Information Estimation of Hidden Markov Models," In Automatic Speech and Speaker Recognition, Kluwer Academics Publishers, 1996.

[4] D. Povey and P. Woodland, "Minimum Phone error and I-smoothing for improved discriminative training," Proc ICCASP, vol.1, pp. 105-108, 2002.

[5] V. Vapnik, "The nature of Statistical Learning Theory," Springer-Verlag, New york, 1995.

[6] H. Jiang, X. Li and C. Liu, "Large Margin Hidden Markov models for speech recognition," IEEE Trans. Audio, Speech, and Language Processing, vol.14, no.5, pp.1584-1595, 2006.

[7] J. Li and C. -H. Lee, "Soft margin feature extraction for automatic speech recognition," Proc. Interspeech, 2007.

[8] O. Bousquet, S. Bouchern and G. Lugosi, "Introduction to statistical learning theory. Advanced lectures on machine learning lecture notes in artificial intelligence 3176, 167-207. (Eds) Springer, Heidelberg, Germany (2004).

[9] J. Hui, " Discriminative training of HMMs for automatic speech recognition: A survey," Computer speech and language, Elsevier Ltd. 2010

[10] L. R. Bahl, P. F. Brown, P.V. De souza, R. L. Mercer, " Maximum mutual information estimation of hidden markov model parameters for speech recognition," in Proc IEEE, International conference on Acoustic, Speech and Signal processing (ICASSP 86), Tokyo, Japan, pp. 49-52.

[11] A. Nadas, D. Nahamoo, M. A. Picheny, "On a model-robust training method for speech recognition," IEEE Transaction on Acoustic, Speech and Signal Processing 36(9), 1432-1436. 1988.

[12] B. -H. Juang, W. Chou, C.-H. Lee, "Minimum classification error rate methods for speech recognition," IEEE Transaction on Speech and Audio Processing 5(3), 257-265. 1997.

[13] S. Katagiri, B. -H .Juang, C.-H. Lee, "Pattern recognition using a generalized probabilistic decent method," In Proc IEEE 86(11), 2345-2373. 1998.

[14] X. He, L. Deng, W. Chou, "Discriminative learning in sequential pattern recognition: a unifying view for optimization-based speech recognition," IEEE Signal Processing Magazine, 14-36. 2008.

[15] D. Povey, "Discriminative training for large vocabulary speech recognition," Ph.D. Dissertation, Cambridge University. 2004.

[16] H. Jiang, X. Li and C. Liu, "Large margin hidden markov model for speech recognition," IEEE Trans. On Audio, speech and Language Proc., vol.14, no.5, pp.1548-1595, 2006.

[17] X. Li, H. Jiang and C. Liu, "Large margin for speech recognition," Proc. ICASSP, pp. V513-V516, 2005.

[18] J. Li, M. Yuan and C.-H. Lee, "Soft margin estimation of hidden markov model parameters," Proc. Interspeech, pp.2422-2425, 2006.

[19] C. Burges, "A tutorial on support vector machines for pattern recognition," Data Mining and Knowledge Discovery, 2(2),121-167. 1998.

SYED ABBAS ALI (1977－), male, Karachi, Pakistan, Research Scholar, Department of Computer Science & Information Technology, his research directions include Machine learning algorithms and automatic speech recognition.

NAJMI GHANI HAIDER (1964－), male, Karachi, Pakistan, Professor, Department of Computer Science & Information Technology supervisor for research scholar, his research directions include Machine learning, robust speech recognition and image processing.

MAHMOOD KHAN PATHAN (1952 －), male, Karachi, Pakistan, Professor, Department of Computer Science & Information Technology co-supervisor for research scholar, his research directions include Computation finite fields and data mining (EDM).

How to cite this paper: Syed Abbas Ali, Najmi Ghani Haider, Mahmood Khan Pathan,"Margin Based Learning: A Framework for Acoustic Model Parameter Estimation", International Journal of Intelligent Systems and Applications(IJISA), vol.4, no.12, pp.26-31, 2012. DOI: 10.5815/ijisa.2012.12.04

