

Performance Evaluation of Loss Functions for Margin Based Robust Speech Recognition

Syed Abbas Ali

Dept. of Computer & Information Systems Engineering,
N.E.D University of Engineering & Technology,
Karachi, Pakistan

Raheela Asif

Computer Science & Information Technology
N.E.D University of Engineering & Technology,
Karachi, Pakistan

Maria Andleeb

Computer Science & Information Technology
N.E.D University of Engineering & Technology,
Karachi, Pakistan

Danish-ur-Rehman

Dept. of Computer & Information Systems Engineering,
N.E.D University of Engineering & Technology,
Karachi, Pakistan

Abstract—Margin-based model estimation methods are applied for speech recognition to enhance the generalization capability of acoustic model by increasing the margin. An important aspects of margin based acoustic model for parameter estimation is that, the acoustic models are derived from soft margin concept and hinge loss function used in SVM as loss function to attained enhanced speech recognition performance. In this study, performance evaluation of loss functions (Logistic, Savage, Sigmoid) have been computed in the presence of white noise, pink noise, and brown noise with and without SVM classifiers to analyze the impact of noise on loss functions in comparison with hinge loss function used in SVM for parameter estimation in margin based acoustic model. Experimental results show that hinge loss function in the presence of pink noise and white noise have significant effects on isolated digits (0-9) in both pre-conditioned and recorded data samples in comparison with brown noise. Whereas hinge loss functions show serious anomalies with savage loss and sigmoid loss in term of performance and sigmoid loss function provides exceptionally good results in term of percentage error for all prescribed conditions.

Keywords—Loss Functions; Statistical Learning; Automatic Speech Recognition (ASR); SVM Classifiers; Soft Margin Estimation (SME)

I. INTRODUCTION

The prime goal of pattern recognition is to find the parameters of recognizers or classifiers that can decrease the error rate by using the existing training data samples. To build an effective pattern recognizer or classifier, two different categories of learning algorithms in machine learning are generative model learning and discriminative model learning. MLE is considered as generative model or non-discriminative learning approach, which is focus on data distribution modeling instead of directly classifying class boundaries. In contrast, discriminative learning approach discriminately learns the parameters of joint probability model to minimize the recognition/classification error [1]. The main idea behind Discriminative training (DT) is to introduce a discriminative criterion to the training method of Hidden Markov Models (HMMs). Several discriminative training methods have been

proposed for ASR, such as maximum mutual information estimation (MMIE) [2,3,4], minimum classification error (MCE) [5,6,7]; and minimum word/phone error (MWE/MPE) [8,9]. For Hidden Markov (HMM) based speech recognition, conventional discriminative training criterions directly minimize the empirical risk on the training data sample and do not focus on the model generalization. In other words, the aim of discriminative training criterions is to minimize classification error on training sample as model estimation but do not show any significance performance to improve the generalization capability of the acoustic model for new unseen test data samples [10]. The generalization capability is an ability to translate gains in the training data set to test data set. In the past studies, the discriminative training achieved this generalization ability by optimizing the smoothed empirical error rate on training data samples [11]. Recently, many researches have been reported to incorporate margins (distance between the decision boundary and well classified data samples) into discriminative training method [12,13,14,15,16,17] to further enhance the generalization capability. The generalization problem of learning classifiers have been studied in the field of machine learning [18,19], whereas, machine learning using concept of statistical learning theory since last three decades to provide the framework for studying inference problem that is of making prediction, gaining knowledge, constructing models and making decisions for a set of data samples [20]. From the statistical learning [18] point of view, a test risk bound is defined by the summation of an empirical risk (i.e., training set risk) and a generalization function. Generalization function is often used to measure the possible mismatch between training and testing environments. ASR researchers at York University proposed the concept of large margin estimation (LME) for speech recognition based on the principle of large margin. Large margin estimation (LME) [10,12] and its variant large relative margin estimation (LRME) [21] of HMMs have been proposed with the concept of enhancing separation margin. The main crux of the LME and LRME is that only correctly classified data samples take part in update models whereas, it is important to note that misclassified data samples are also substantial for classifier learning. To address this issue in LME and LRME, the

extension of LRME [86] was proposed by considering all the training data samples, particularly moving misclassified data samples in the direction of correct decision boundary. Another margin based approach, Soft margin estimation (SME) was proposed by J.Li et al [22] from Georgia Tech University based on the idea of soft margin in support vector machines [23] to enhance the generalization capability of the learning classifiers. Soft margin estimation (SME) performs well as compared to Maximum likelihood estimation (MLE) and conventional discriminative criterion, and it is steadily better than Large margin estimation (LME) due to the well-defined separation(misclassification) measure and good optimized objective function for generalization [24]. In contrast, with LME, SME make use of both misclassified and correct classified data samples to update models and the performance of the SME can be improve when the distribution of testing and training data samples become quite comparable[25]. Two considerable issues have been identified in [26] related to hinge loss function in Soft margin (SME) 1) hinge loss function performs well when the noise in training sample is insignificant and 2) any misclassified training sample directly affects the time required for optimization and determines the label of the test sample. To improve this limitation of SME, X.Xiao et al [27] proposed feature domain method based on mean and variance normalization (MVN) [28] which showed that SME perform well with feature domain method and reduces the mismatch between training and testing data samples and suggested the combination of SME framework with other noise compensation methods e.g. model adaptation methods for future research. Issues related to hinge loss function, Geometric margin MCE criteria in soft margin estimation framework based on sigmoid loss function were presented to find the strength of robustness by increasing the geometric margin of the acoustic model [29]. Loss functions such as ramp loss and 0-1 loss also showed comparable noise tolerance capability like sigmoid loss function [30]. In this paper, demonstrative experiments have been performed to observe the behavior of three loss functions (Logistic, Savage, Sigmoid) in the presence of white noise, pink noise, and brown noise with and without SVM (Soft margin) classifiers in comparison with hinge loss function for preconditioned and recorded digit (0-9) taken from environment.

Rest of the paper is organized as follow. The consequent section discusses the loss function for soft margin (SME) including sigmoid, savage and logistic loss functions. In section III, Data collection and recording specifications are defined. The experimental results and discussions are presented with pre-conditioned and recorded digits in section IV. Finally, conclusions are drawn in section V.

II. LOSS FUNCTION FOR SOFT MARGIN ESTIMATION (SME)

The objective of recognition and classification systems is to minimize the classification risk on testing data samples by developing a classifier f . The concept behind the risk minimization is to measure the performance of estimator by its risk, in order to select best estimator function we should have a measure of inconsistency between an estimated classification $f(x, z)$ and true classification $Y(x)$ of x as shown in (1) and (2) respectively,

$$f(x, z) = y' \quad (1)$$

$$Y(x) = y \quad (2)$$

The performance of classifier f can be measure using loss function $L(y, f(x, z))$, which can be defined as;

$$L(y, f(x, z)) = \begin{cases} 0 & \text{if } y = f(x, z) \\ 1 & \text{if } y \neq f(x, z) \end{cases} \quad (3)$$

Consider a risk or estimator function providing the true or expected value of loss as follows:

$$R_{\text{true}}(z) = \int L(y, f(x, z)) dP(x, y) \quad (4)$$

Where $R_{\text{true}}(z) = \int L(y, y') dP(x, y)$ and $P(x, y) = P(x) P(y/x)$

There is a need to find function $f(x, z)$ that minimize the risk function $R_{\text{true}}(z)$ (for all functions $f(x, z), z \in \Lambda$), but we don't know $P(x, y)$. Soft Margin Estimation (SME) [13,22] is a margin-based model estimator applied for speech recognition with an objective to enhance the generalization capability and decision feedback learning by increasing the margin and to enhancing the separation measures of the model in the classifier design respectively. Concept of test risk bound has been defined in statistical learning theory bounded by the summation of two terms: A generalization function and an empirical risk (i.e. risk on the training set) [18]. The generalization of a model is a monotonically increasing function of its VC dimension which is used to measures the complexity of model bounded by decreasing function of margin[18]. Soft margin estimation (SME) combines two target optimization function in a single object function based on soft margin estimation,

$$\Lambda_{\text{SME}} = \frac{\lambda}{\rho} + R_{\text{emp}}(\Lambda) = \frac{\lambda}{\rho} + \frac{1}{N} \sum_{t=1}^N l(O_t, \Lambda) \quad (5)$$

Λ represent the set of HMM based model parameters, $R_{\text{emp}}(\Lambda)$ represent empirical risk, $l(O_t, \Lambda)$ defines the loss function for utterance O_t and N is the total number of training utterances. Whereas, λ and ρ are the coefficient used to balance the empirical risk minimization and margin maximization and soft margin respectively. Margin based acoustic model derived from soft margin concept and hinge loss function used in SVM is defined as loss function to attained enhanced speech recognition performance. Hinge loss function does not perform well in the presence of significant amount of noise. This experimental setup evaluate the performance of hinge loss function in the presence of noise in comparison with three other loss functions with and without SVM classifier for preconditioned isolated digit and digit taken from real environment. The hinge loss function used in SVM can be defined as [22]:

$$l(d(O_t, \Lambda)) = d(O_t, \Lambda) \cdot \frac{1}{1 + e^{-\gamma d(O_t, \Lambda)}} \quad (6)$$

γ is a positive value number relating to smoothness of loss function and $d(O_t, \Lambda)$ represent the for soft margin (SME). Similarly savage loss [31], standard sigmoid loss and logistic loss function can be written as (7), (8) and (9) respectively.

$$l(d(O_t, \Lambda)) = 1/(1 + e^{-\gamma d(O_t, \Lambda)})^2 \quad (7)$$

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

$$l(d(O_t, \Lambda)) = \ln(1 + e^{-\gamma d(O_t, \Lambda)}) \quad (9)$$

III. DATA COLLECTION AND RECORDING SPECIFICATIONS

The methodology of experiments includes the collection of data that comprises of two individual sets; the set of TI-digits (0-9) standard isolated digit corpus and digits recorded from real environment. For the recording specification of recorded isolated digit corpus, ITU recommendations based standardized procedure was adopted for speech corpora development. Standard recording environment has been used having SNR (signal to noise ratio) greater than and equal to 45dB. We made use of Microsoft Windows 7 built-in sound recorder to record the 10 utterances of each isolated digit (0-9). The recording format is Mono, 32 bit PCM with sampling rate of 8000Hz using microphone with impedance of 32 Ω , Max Input power=40mW, Drive Unit=30mm, Plug Type=3.5MM, Frequency Response=20Hz ~ 20 KHz. Microphone with specified configuration were used to take input digits 0 to digit 9 and recorded in noise free recording studio environment. Afterward white noise, brown noise and pink noise were mixed with both sets of data with the help of audacity software. The purpose of noise addition in isolated TI-Digit and recorded digit samples is to study the behavior of each digit under prescribed conditions with and without white, brown and pink noises. The number of experiments was performed with the evaluation of cepstrum coefficient values for each digit in clean and noisy conditions and four graphs of each digit for both recorded and isolated TI-Digit databases were analyzed separately.

IV. EXPERIMENTAL RESULTS AND DISCUSSION

In this section, we present demonstrative experiments to show the performance of four loss function such as sigmoid, hinge, savage and logistic in the presence of noise with and without SVM (Soft margin) classifiers to observe the behavior of four different loss functions with pre-conditioned and recorded data samples. Data samples were used in experiments consist of isolated digits taken from TI-Digit corpus [33] and recorded digits taken from real environment with and without SVM (Soft Margin) classifier in the presence of three different types of noises (White, Brown and Pink) taken from NOISEX-92 noise-in-speech database [32,34]. The experimental frame work was divided in two phases: two sets of data have been used with each phase of experiments comprises on isolated TI-Digits (0-9) samples of recorded digits taken from real environment. First stage shows the results of pre-recorded isolated Ti-digits (0-9) without classifier & with addition of three noises. Second stage contains the same results with performance of SVM classifier. In the next phase of

experiment, we studied behavior of different loss functions: hinge loss, sigmoid loss, savage loss and logistic individually for data samples of each set in the presence of white, brown and pink noise with and with SVM (Soft Margin) classifiers. The results obtained from experimental analysis represented in loss function comparative analysis charts clearly evident that behavior of loss functions significantly changes for clean & noisy conditions. Furthermore, we present the graphical analysis of different loss function with and without noise to observe behavior of hinge loss, sigmoid loss, savage loss and logistic loss function for both sets of isolated TI-Digit and recorded digit. SVM (Soft Margin) Classifier took values of loss functions values as an input for both data sets separately. Based on the results obtained in the previous steps, classifier classifies the clean sample from noisy samples. For graphical representation of the experimental results in the proceeding sections, digit "one" was selected with and without SVM (soft margin) classifiers in the presence of white, brown and pink noises for both isolated TI-Digit and recorded data samples. Results generated from the experimental framework were based on numerous pieces of code that were implemented & observed in MATLAB tool version 10.0 and speech processing toolbox.

A. Graphical Representation and Loss Functions Comparative Analysis of Isolated Ti-Digit Without and with SVM Classifiers

To evaluate the performance of data sets in clean and noisy conditions, the cepstrum of the each digit in the data set with and without noise were obtained to determine the peak values of cepstrum coefficient. We made use of these cepstrum coefficients to distinct clean digit from the noisy data sample. Isolated TI-Digit "1" was selected for the graphical representation of entire experimental results in this and later sections to illustrate the behavior of the different loss function under certain conditions. The interpretation of graphical results demonstrated by blue line and red line. The blue line/curve represent the plot of loss function value without noise and the red line/curve represent the plot of loss function value with noise. The gap between two lines or curves obtained from the different loss functions indicate the resultant value of error between loss functions with and without noise. When the gap increases between two line/curve, it provides higher value of error which reflects the poor performance in the presence of noise.

The red and blue lines/ curves in the above figures clearly show that behavior of the different loss functions with three different noises. The gap between lines/curve of sigmoid and savage loss is lesser than the hinge and logistic loss. Loss function comparative analysis for all isolated digits (0-9) in the next section indicates that the savage and sigmoid loss functions perform well in comparison with hinge and logistic loss function except for some anomalies. Similarly, the performance of loss functions have been evaluated with isolated TI-Digit in the presence of white, brown and pink noise using SVM (soft margin) classifiers. We made use of SVM classifiers to separate clean TI-Digit from noisy TI-Digit as shown in Fig. 5, Fig. 6 for digit 1 with white and brown noise respectively.

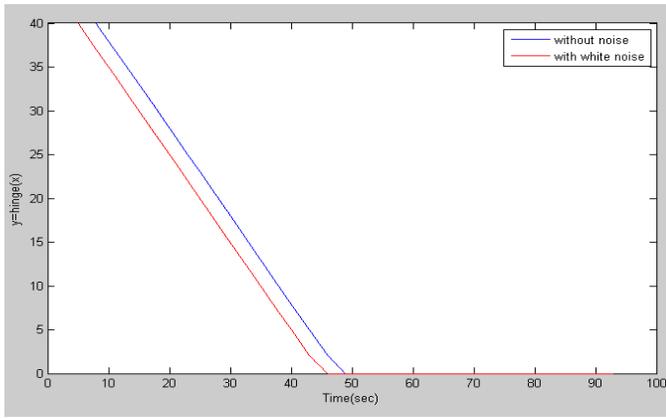


Fig. 1. TI-Digit 1 in the presence of white noise without SVM classifier using hinge loss

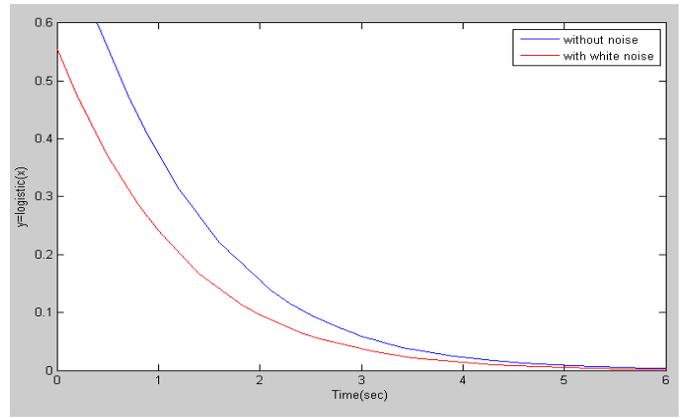


Fig. 4. TI-Digit 1 in the presence of white noise without SVM classifier using logistic loss

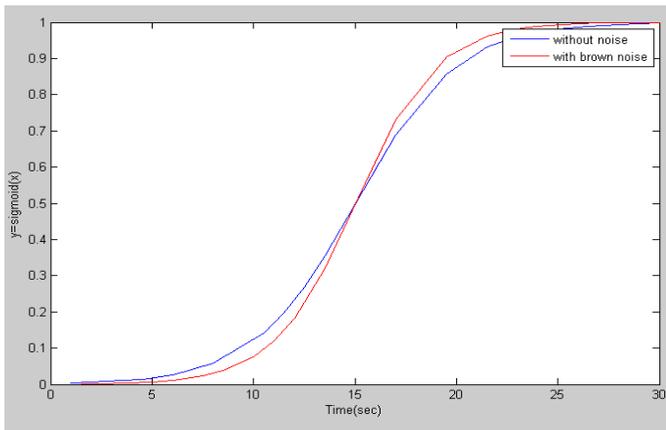


Fig. 2. TI-Digit 1 in the presence of brown noise without SVM classifier using sigmoid loss

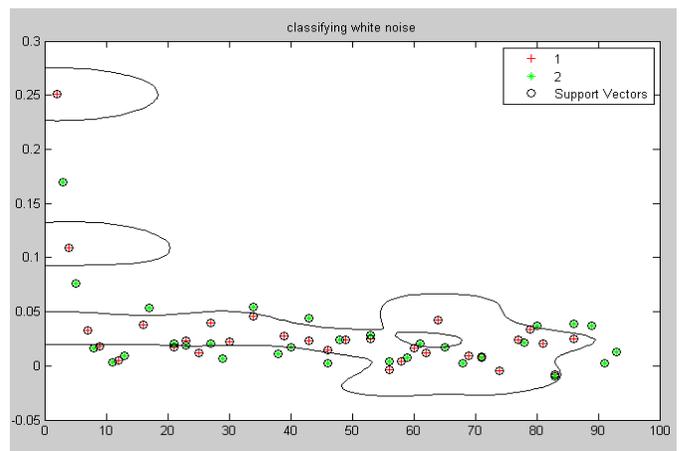


Fig. 5. Plot of TI-digit 1 using SVM classifier with white noise

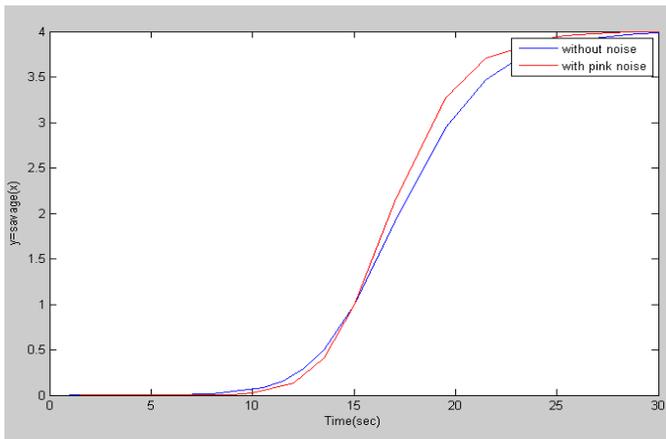


Fig. 3. TI-Digit 1 in the presence of pink noise without SVM classifier using savage loss

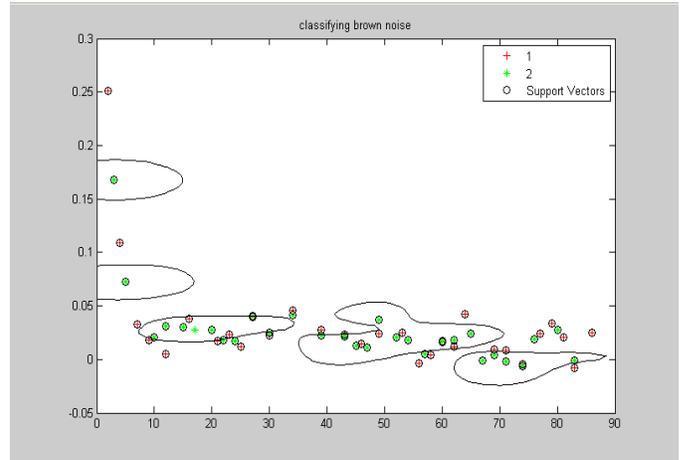


Fig. 6. Plot of TI-digit 1 using SVM classifier with brown noise

The interpretation of the above plots demonstrated by green and red dots which were used to represent the separation of clean and noisy signal respectively using SVM classifiers. The implementation of hinge loss, sigmoid loss, savage loss and logistic loss functions have been done through SVM classifier.

Fig. 7 and Fig. 8 represent hinge and sigmoid plot of Isolated TI-Digit 1 with brown and white noise respectively using SVM classifiers.

Loss functions comparative analysis have been performed among sigmoid loss, hinge loss, savage loss and logistic loss in the presence of white noise, brown noise and pink noise without SVM (Soft Margin) classifiers for Isolated TI-Digit.

The interpretation of experimental results demonstrated by blue bar, red bar, green bar and purple bar for hinge function, sigmoid function, savage function and logistic function respectively.

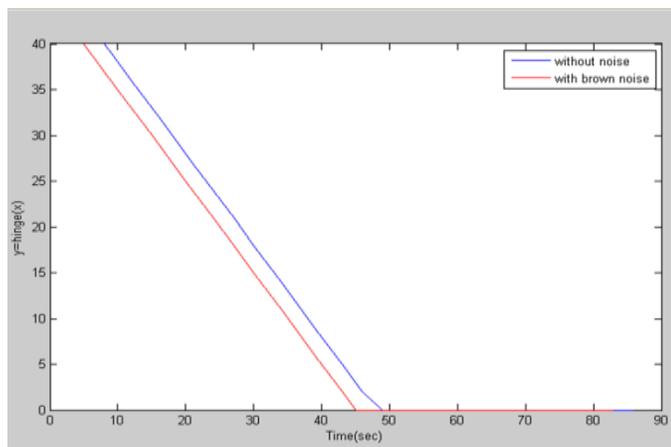


Fig. 7. TI-Digit 1 in the presence of brown noise with SVM classifier using hinge loss

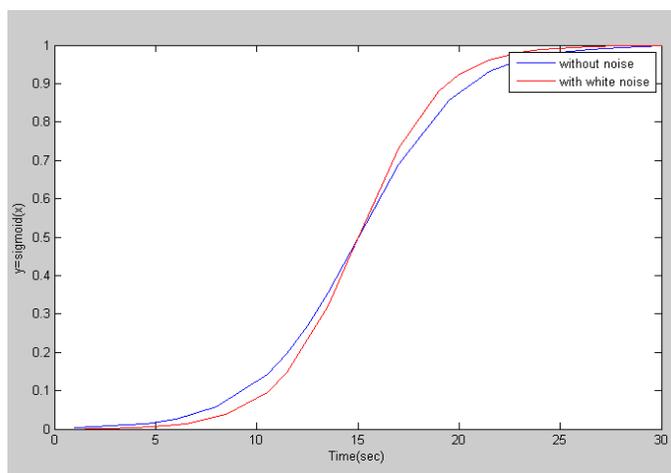


Fig. 8. TI-Digit 1 in the presence of white noise with SVM classifier using sigmoid loss

TABLE I. LOSS FUNCTIONS COMPARATIVE ANALYSIS OF WHITE NOISE FOR ISOLATED TI-DIGIT WITHOUT USING CLASSIFIER

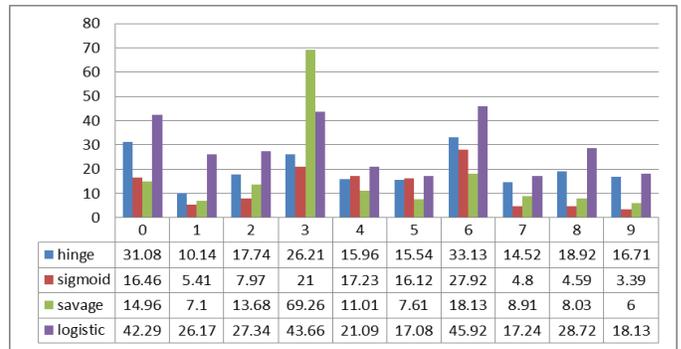


TABLE II. LOSS FUNCTION COMPARATIVE ANALYSIS OF BROWN NOISE FOR ISOLATED TI-DIGIT WITHOUT USING CLASSIFIER

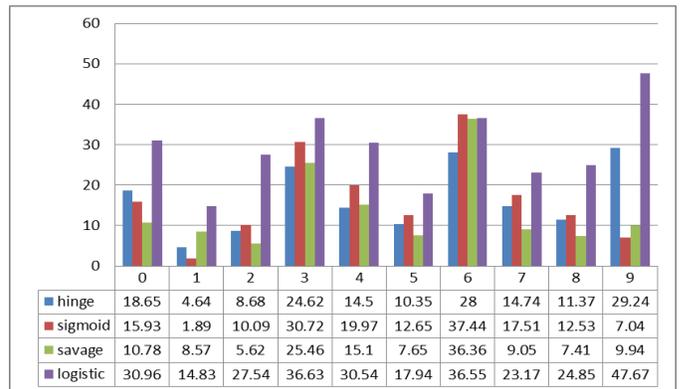
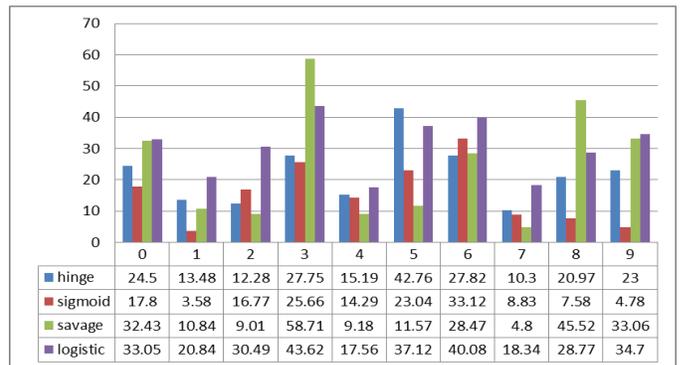


TABLE III. LOSS FUNCTION COMPARATIVE ANALYSIS OF PINK NOISE FOR ISOLATED TI-DIGIT WITHOUT USING CLASSIFIER



In Table 1, savage and sigmoid loss function represents substantial anomalies when compared with hinge loss function for digits 3, digit 4 and digit 5. Whereas, perform of hinge loss function quite well than Logistic function. Table 2 indicate that hinge loss function perform well than Logistic loss function in the presence of brown noise, while hinge loss function shows considerable anomalies when compared with savage and sigmoid loss function for digit 1, digit 3, digit 4 and digit 6.

In Table 3, hinge function in the presence of pink noise not performs well than sigmoid function except for digit 2 and digit 6. Whereas, some anomalies have been observed with savage loss function in comparison with hinge loss for digit 0, digit 3, digit 8 and digit 9. Similarity, loss functions comparative analysis have been performed among sigmoid loss, hinge loss, savage loss and logistic loss in the presence of white noise, brown noise and pink noise with SVM (Soft Margin) classifiers for Isolated TI-Digit.

TABLE IV. LOSS FUNCTIONS COMPARATIVE ANALYSIS OF WHITE NOISE FOR ISOLATED TI-DIGIT USING CLASSIFIER

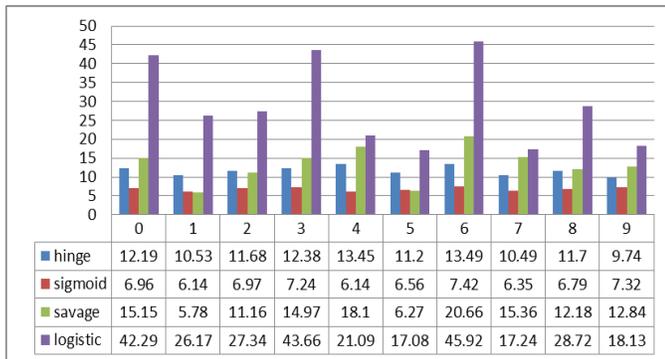


TABLE V. LOSS FUNCTIONS COMPARATIVE ANALYSIS OF BROWN NOISE FOR ISOLATED TI-DIGIT USING CLASSIFIER

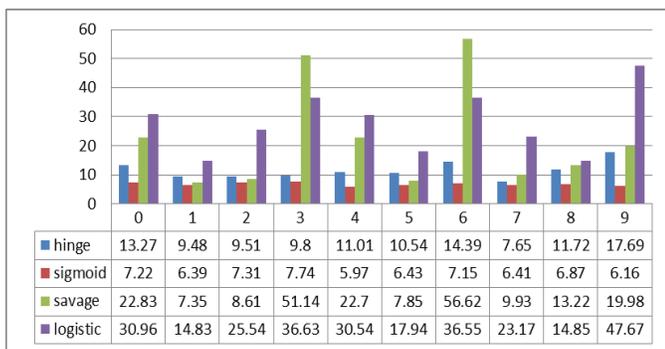
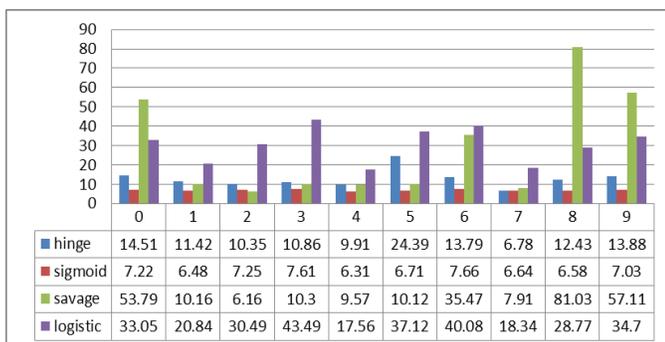


TABLE VI. LOSS FUNCTIONS COMPARATIVE ANALYSIS OF PINK NOISE FOR ISOLATED TI-DIGIT USING CLASSIFIER



In Table 4, hinge is better than Logistic function for all digits but severe anomalies can be seen when comparing performance with other loss functions. Table 5 shows that hinge is better than Logistic function for all digits but severe anomalies can be seen when comparing performance with other loss functions. In Table 6, logistic loss function not

performs well than hinge loss, but considerable anomalies have been observed when hinge loss function compared with savage loss function.

B. Graphical Representation of Loss Functions Comparative Analysis of Recorded Digit without and with SVM Classifiers

The performance of the loss functions have been evaluated in this section, using recorded digit samples taken from environment to study the behavior of the hinge loss, sigmoid loss, savage loss, and logistic in the presence of noise. Fig. 9 and Fig. 10 represent hinge and logistic plot of Isolated TI-Digit 1 in the presence of pink and pink noise respectively without SVM classifiers.

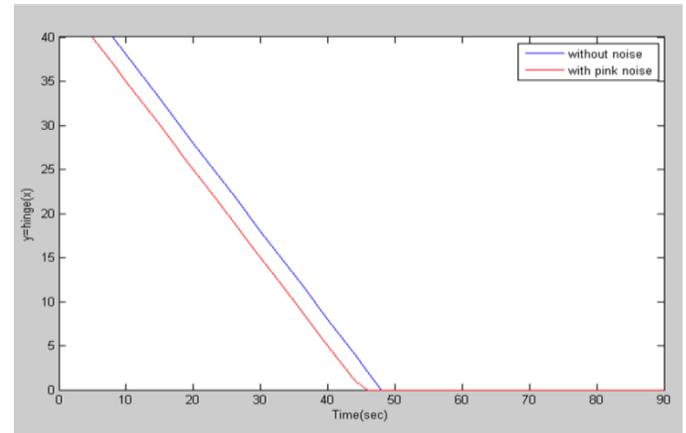


Fig. 9. Recorded digit 1 in the presence of pink noise without SVM classifier using hinge loss

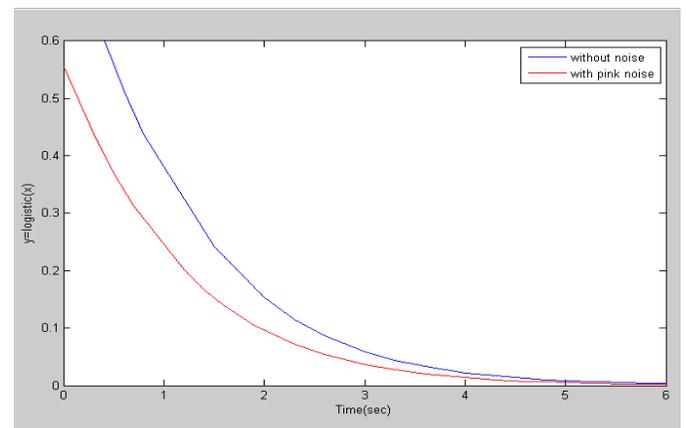


Fig. 10. Recorded digit 1 in the presence of pink noise without SVM classifier using logistic loss

Similarly, we evaluated the performance of loss functions with recorded digit taken from environment in the presence of white, brown and pink noise using SVM classifiers. We made use of SVM classifiers to separate clean recorded digit sample from noisy samples. Fig. 11 and Fig. 12 represent hinge and sigmoid plot of recorded digit 1 with brown and pink noise respectively with SVM classifiers.

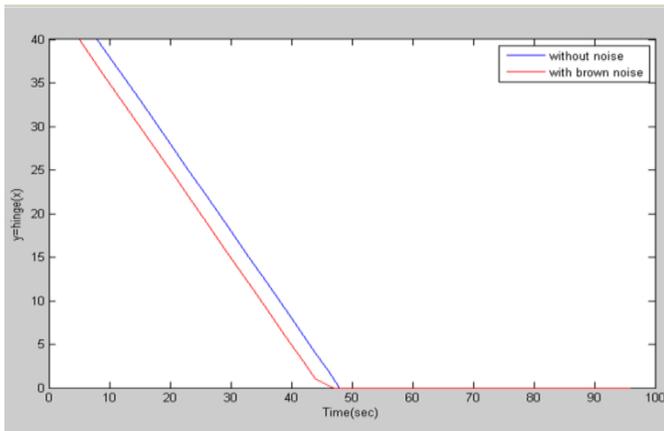


Fig. 11. Recorded digit 1 in the presence of brown noise with SVM classifier using hinge loss

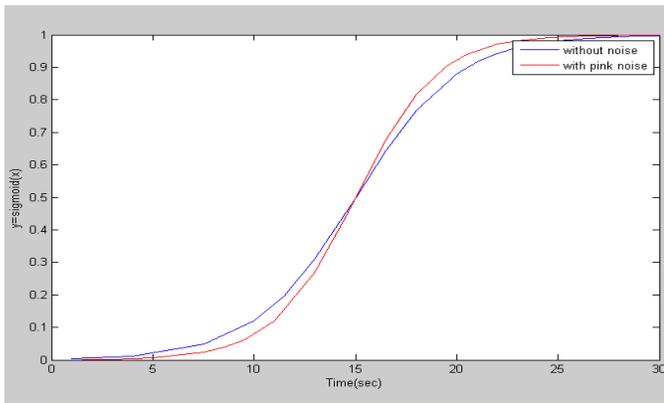


Fig. 12. Recorded digit 1 in the presence of pink noise with SVM classifier using sigmoid loss

Loss function comparative analysis among hinge loss, sigmoid loss, savage loss and logistic loss have been performed in the presence of white noise, pink noise and brown noise without SVM (Soft Margin) classifiers for recorded digits taken from real environment.

TABLE VII. LOSS FUNCTIONS COMPARATIVE ANALYSIS OF WHITE NOISE FOR RECORDED DIGIT WITHOUT CLASSIFIER

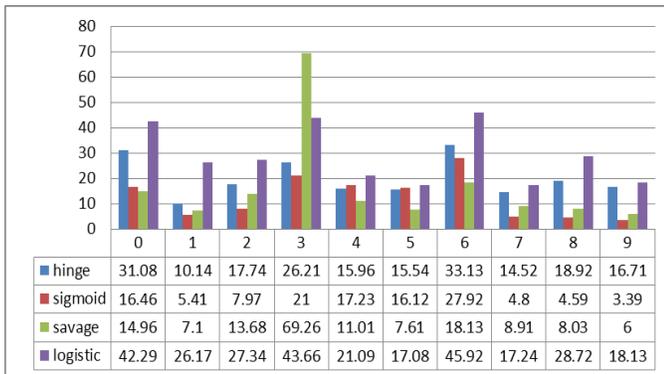


TABLE VIII. LOSS FUNCTIONS COMPARATIVE ANALYSIS OF BROWN NOISE FOR RECORDED DIGIT WITHOUT CLASSIFIER

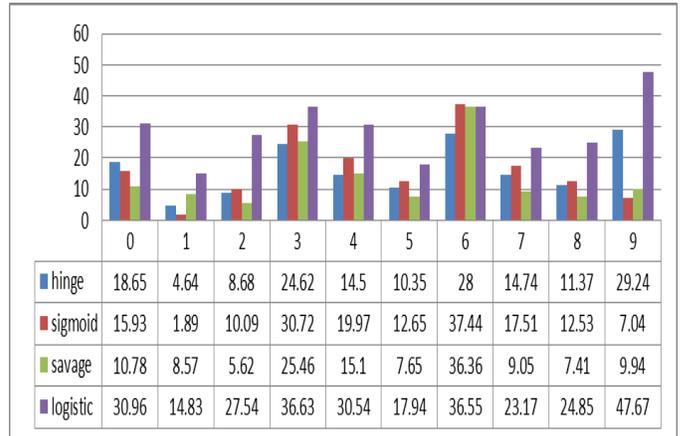


TABLE IX. LOSS FUNCTIONS COMPARATIVE ANALYSIS OF PINK NOISE FOR RECORDED DIGIT WITHOUT CLASSIFIER

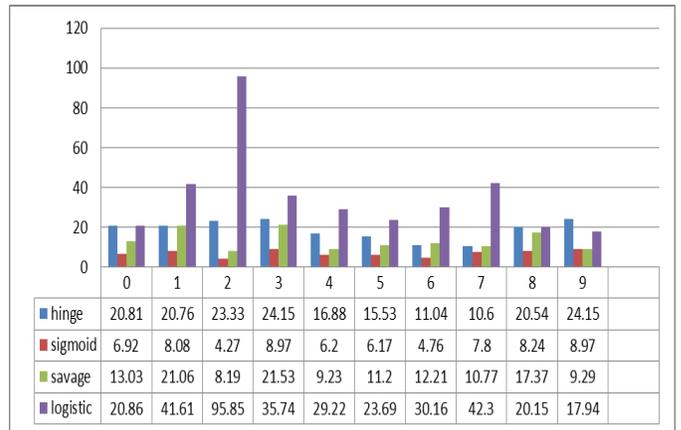


Table 7, Table 8 and Table 9 provide the comparative analysis of hinge, sigmoid, savage and logistic loss function for recorded digit without SVM (Soft Margin) Classifiers. In Table 7, savage and sigmoid loss function performs well than hinge loss function in the presence of white noise except digits 3, digit 4 and digit 5.

In Table 8, serious anomalies have been observed with hinge loss function when compared with savage and sigmoid in some digits with brown noise. Table 9 displays that savage and sigmoid perform well in comparison with hinge loss except digit 1 and digit 7 whereas logistic function not performs well than hinge loss function.

Similarly, loss function comparative analysis among hinge loss, sigmoid loss, savage loss and logistic loss in the presence of white noise, brown noise and pink noise with SVM (Soft Margin) classifiers for recorded digits taken from real environment.

TABLE X. LOSS FUNCTIONS COMPARATIVE ANALYSIS OF WHITE NOISE FOR RECORDED DIGIT WITH CLASSIFIER

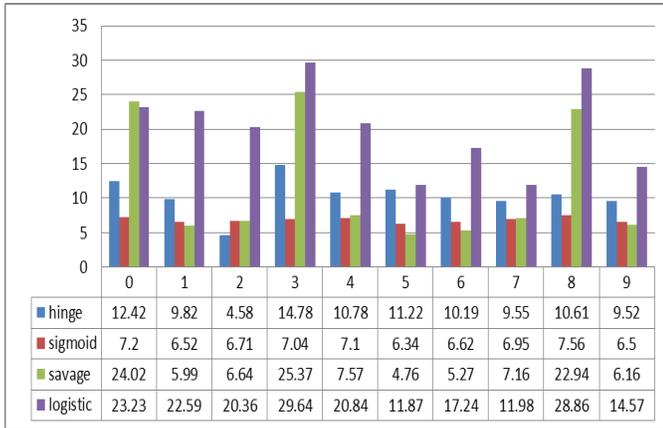


TABLE XI. LOSS FUNCTIONS COMPARATIVE ANALYSIS OF BROWN NOISE FOR RECORDED DIGIT WITH CLASSIFIER

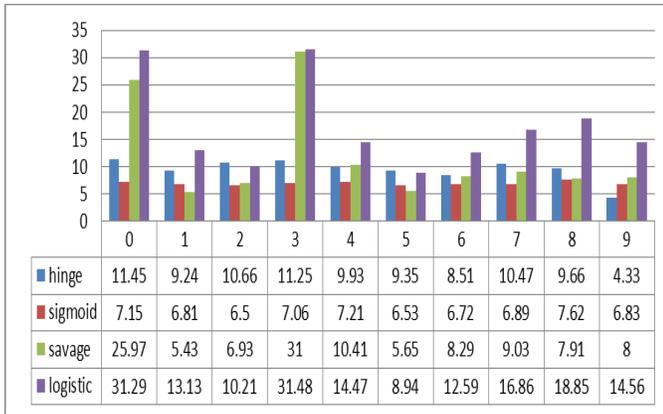


TABLE XII. LOSS FUNCTIONS COMPARATIVE ANALYSIS OF PINK NOISE FOR RECORDED DIGIT WITH CLASSIFIER

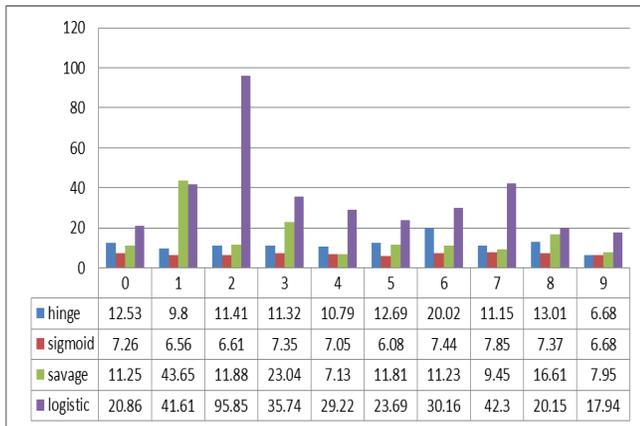


Table 10, Table 11 and Table 12 provide the comparative analysis of hinge, sigmoid, savage and logistic loss function for recorded digit with SVM (Soft Margin) Classifiers. In Table 10, hinge loss not perform well as compared to savage and sigmoid loss function but some anomalies have been observed

with digit 0, digit 2, digit 3 and digit 8 in the presence of white noise. In Table 11, hinge loss performs well than logistic loss except for digit 2 whereas, some anomalies have been observed with digit 0, digit 3, digit 4 and digit 9 in the presence of brown noise when compared with savage and sigmoid function. Table 12 indicates that logistic loss function not perform well in comparison with hinge loss but savage function and sigmoid function perform well than hinge loss function except for digit 1, digit 3 and digit 8. The following observation has been acquired from demonstrative experiments:

- Pink noise and white noise illustrate significant effects on isolated digits (0-9) in both pre-conditioned and recorded conditions in comparison with brown noise. Hinge loss function doesn't perform well than sigmoid loss and savage loss functions but it performs better than logistic loss function in the presence of white and pink noise, however, some anomalies are observed in the presence of brown noise.
- In all four prescribed conditions in demonstrative experiments for both recorded digits taken from environment and pre-conditioned TI-Digits with and without classifiers, logistic loss function not perform well in comparison with hinge loss function whereas hinge loss function show serious anomalies with savage loss sigmoid and functions in term of performance.
- In comparison with hinge loss and Logistic loss functions, sigmoid loss function provides exceptionally good results in term of percentage error for all prescribed conditions in experiments. Whereas, few inconsistencies can be seen in the performance of savage loss function in comparison with hinge loss.

V. CONCLUSION

Motivated by the issue related to hinge loss function used in SVM for parameter estimation in margin based acoustic model, this paper presented the comparative analysis of three loss functions (Logistic, Savage, Sigmoid) in comparison with hinge loss to observe the behavior of loss functions in the presence of white noise, pink noise, and brown noise with and without SVM (Soft margin) classifiers for preconditioned and recorded data samples. Demonstrative experiments have been made on NOISEX-92 (speech and noise-in speech) databases, TIDIGIT corpus and recorded data samples (0-9) taken from real environment. The demonstrative experiments indicated that hinge loss function doesn't perform well than savage loss and sigmoid loss functions but it performs better than logistic loss function in the presence of pink and white noise as compared to brown noise for all prescribed conditioned. Whereas, sigmoid loss function shows remarkably better results in comparison with hinge and other loss function in term of percentage error.

REFERENCES

- [1] X. He and L. Deng, Discriminative Learning for Speech Recognition: Theory and Practice. Morgan & Claypool, 2008.

- [2] Y. Normandin, "Maximum Mutual Information Estimation of Hidden Markov Models," In Automatic Speech and Speaker Recognition, edited by C. H. Lee, F. K. Soong and K. K. Paliwal, Eds. Kluwer Academics Publishers, Norwell, M.A, 1996, pp.1-159.
- [3] L. R. Bahl, P. F. Brown, P. V. Desouza and R. L. Mercer, "Maximum Mutual Information Estimation of Hidden Markov Models parameters for speech recognition," In Proc. ICASSP, vol.1, pp. 49-52, 1986.
- [4] V. Valtchev, J. Odell, P. Woodland and S. Young, "Maximum Mutual Information Estimation training for large vocabulary recognition systems," Speech Communication, vol.22, no.4, pp. 303-314, 1997.
- [5] 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.
- [6] B.-H. Juang, and S. Katagiri, "Discriminative learning for Minimum Error Classification," IEEE Trans. on Signal Processing, vol.40, pp.3043-3054, 1992.
- [7] R. Schlueter, W. Macherey, B. Muller, and H. Ney "Comparison of discriminative training criteria and optimization methods for speech recognition," Speech Communication, vol.34, pp.287-310, 2001.
- [8] D. Povey and P. Woodland, "Minimum Phone error and I-smoothing for improved discriminative training," In Proc. ICCASP, vol.1, pp. 105-108, 2002.
- [9] D. Povey, Discriminative training for large vocabulary speech recognition, Ph.D. dissertation, Cambridge University, Dept. Eng., Cambridge, UK, 2004.
- [10] H. Jiang, X. Li and C. Liu, "Large Margin Hidden Markov models for speech recognition," IEEE Trans. on Audio, Speech and Language Processing, vol.14, no.5, pp.1584-1595, 2006.
- [11] D. Yu, L. Deng, X. He and A. Acero, "Large Margin minimum classification training for Large-Scale Speech Recognition Tasks," In Proc. ICASSP, 2007.
- [12] H. Jiang and X. Li, "Solving large margin HMMs estimation via semi-definite programming," In Proc. ICASSP, vol.4, no.5, pp.IV-629-IV-632, 2007.
- [13] J. Li, M. Yuan and C.-H. Lee, "Soft margin estimation of Hidden Markov Model parameters," In Proc. Interspeech, pp.2422-2425, 2006.
- [14] X. Li, and H. Jiang, "A constrained joint optimization methods for large margin HMM estimation," In Proc. ASRU Workshop, pp.151-156, 2005.
- [15] C. Liu, H. Jiang and L. Rigazio, "Recent improvement of minimum relative margin estimation of HMMs for speech recognition," In Proc. ICASSP, vol.1, pp.269-272, 2006.
- [16] F. Sha, and L. Saul, "Large-Margin Gaussian mixture modeling for phonetic classification and recognition," In Proc. ICASSP, vol.1, pp.265-268, 2006.
- [17] D. Yu, L. Deng, X. He and A. Acero, "Use of incrementally regulated discriminative margins in MCE training for speech recognition," In Proc. Interspeech, pp. 2418-2421, 2006.
- [18] V. Vapnik, The nature of Statistical Learning Theory. New York: Springer-Verlag, 1995.
- [19] B. Scholkopf and A. J. Smola, Learning with Kernels: support vector machine, regularization, optimization, and beyond. Cambridge: The MIT Press, 2002.
- [20] O. Bousquet, S. Bouchern and G. Lugosi, "Introduction to statistical learning theory. Advanced lectures on machine learning lecture notes in artificial intelligence 3176, pp.167-207. (Eds) Springer, Heidelberg, Germany, 2004.
- [21] C. Liu, H. Jiang and X. Li "Discriminative training of CDHMMs for maximum relative separation margin," In Proc. ICASSP, pp.1101-1104, 2005.
- [22] J. Li and C. -H. Lee, "Soft margin feature extraction for automatic speech recognition", Proc. Interspeech, 2007.
- [23] C. Burges, "A tutorial on support Vector machine for pattern recognition," Data Mining and Knowledge Discovery, vol.2, no. 2, pp.121-167, 1998.
- [24] J. Li, C.-H. Lee and R. H. Wang, "A Study of Soft margin estimation for LVCSR," In Proc. IEEE automatic speech recognition and understanding workshop, pp.268-271, 2007.
- [25] R. K. Aggarwal and M. Dave, "Acoustic modeling problem for ASR system: advances and refinements (Part II)," International Journal of Speech Technology, Springer, pp.309- 320, 2011.
- [26] S.A. Ali, N. G. Haider and M. K. Pathan., "Margin Based Learning: A Framework for Acoustic Model Parameter Estimation", I.J. Intelligent Systems and Applications, vol. 2, no.12, pp. 26-31, November 2012.
- [27] X. Xiao, J. Li, E. S. Chng, H. Li, C. H. Lee, " A study on the generalization capability of Acoustic modes for Robust Speech Recognition," IEEE Trans. on Audio, Speech and Language Processing, Vol. 18, No.6, August 2010.
- [28] O. Viikki, K. Laurila, "Cepstral domain segmental feature vector normalization for noise robust speech recognition," Speech Communication, vol.25, pp.133-147, 1998.
- [29] S.A. Ali and N.G. Haider, "Margin Based Learning Framework with Geometric Margin Minimum Classification Error for Robust Speech Recognition," International Journal of Sciences: Basic and Applied Research (IJSBAR), vol.11, No 1, pp. 39-48, 2013.
- [30] A. Gosh, N. Manwani, and P. S. Sastry. Making risk minimization tolerance to Label Noise. CoRR, abs/1403.3610, March 2014.
- [31] H. Masnadi-Shirazi and N. Vasconcelos, "On the design of loss functions for classification: theory, robustness to outliers, and Savage Boost", Advances NIPS, 2007.
- [32] A.P. Varga, H. J. M Steeneken, M. Tomlinson, and D. Jones, "The NOISEX-92 Study on the Effect of Additive Noise on Automatic Speech Recognition," In Technical Report, DRA Speech Research Unit, 1992.
- [33] Clean digits are from "TIDIGIT Corpus", available at www.ee.columbia.edu/~dpwe/sounds/tidigits/
- [34] Noise Corpus are from "NOISEX-92" http://Spib.rice.edu/spib/select_noise.