MVA '96 IAPR Workshop on Machine Vision Applications, November. 12-14, 1996, Tokyo, Japan

Evaluation of the Noise Injection in High Dimensions

Yoshihiro Mitani, Yoshihiko Hamamoto * and Shingo Tomita
Faculty of Engineering, Yamaguchi University

Abstract

The noise injection into the training samples has been shown to lead to improvement of the generalization ability of artificial neural network(ANN) classifiers. In this paper, we investigate the positive effect of the noise injection on the generalization ability of ANN classifiers in high dimensions. We further show that the noise injection technique is very useful in situations where the true Bayes error is small.

1 Introduction

Raudys and Jain [1] point out that small sample size always leads to difficulties in designing artificial neural network(ANN) classifiers. Hence, in small training sample size situations, a considerable amount of effort has been devoted to improvement of the generalization ability of ANN classifiers [2] [3].

Some authors [4] [5] showed experimentally that the noise injection into the training samples improves the generalization ability of the resulting ANN classifiers. Moreover, Matsuoka [6] and Grandvalet et al. [7] studied theoretically the noise injection in terms of the smoothness of the mapping to be realized by a network. Holmström et al. [8] also studied theoretically the noise injection from viewpoint that using additive noise can be regarded as a kernel estimation. Bishop et al. [9] stated a relation between training with noise and regularization. However, these theoretical works do not address the generalization error directly. We are interested in practical situations where the sample size is small, or the feature size is large. In small training sample size situations, we experimentally studied the effect of the noise injection on the generalization error of ANN classifiers [10]. However, no attempt has been made to discuss the effect of the noise injection in high dimensions.

In this paper, we investigate the effect of the noise injection on the generalization error of ANN classifiers in high dimensions. Experimental results show the effectiveness of the noise injection even in high dimensions. Furthermore, we show that the noise injection is an effective means of improving the generalization ability of ANN classifiers, particularly when the true Bayes error is small.

*Address: 2557 Tokiwadai, Ube 755 Japan. E-mail: hamamoto@csse.yamaguchi-u.ac.jp

2 ANN Classifiers

We will consider ANN classifiers with one hidden layer. The input neurons correspond to the components of the pattern vector to be classified. The hidden layer has m neurons. The output neurons correspond to the pattern class labels. For simplicity we will focus on the two-class problem. Hence, the number of output neurons is 2. Each neuron of one layer except the output layer is fully connected to that of the only next layer. The back-propagation(BP) algorithm [11] was used to train the ANN classifiers. Initial weights of a network were distributed uniformly in -0.5 to 0.5. Learning was terminated when the mean-squared error over a training set dropped below a specified threshold, or when the mean-squared error was unchanged. Here, the maximum number of iterations was set to 10000. The rate of convergence is considerably affected by the learning rate c. From preliminary experiments, we used $c = 0.1$.

3 Noise Injection

We describe the noise injection technique below. Consider now N training samples $\{\boldsymbol{x}_1^i, \boldsymbol{x}_2^i, \cdots, \boldsymbol{x}_N^i\}$ from class ω_i, where $i = 1, 2$. The training samples with Gaussian noise, $\widehat{\boldsymbol{x}}_j^i$'s, are given by

$$\widehat{\boldsymbol{x}}_j^i = \boldsymbol{x}_j^i + \boldsymbol{n}, \tag{1}$$

where $\boldsymbol{n}$ is a random vector normally distributed with mean vector $\mathbf{o}$ and covariance matrix εI, and ε is a parameter determining the magnitude of noise. We assume that $\boldsymbol{n}$ is a random vector independent of $\boldsymbol{x}_j^i$. Hence, $\boldsymbol{n}$'s should be generated independently for any given ε. When $\boldsymbol{x}_j^i$ is normal with the mean vector $\boldsymbol{\mu}_i$ and covariance matrix Σ_i, then the mean vector and covariance matrix of $\widehat{\boldsymbol{x}}_j^i$ are

$$E\left[\widehat{\boldsymbol{x}}_j^i\right] = \boldsymbol{\mu}_i \tag{2}$$

$$E\left[\left(\widehat{\boldsymbol{x}}_j^i - \boldsymbol{\mu}_i\right)\left(\widehat{\boldsymbol{x}}_j^i - \boldsymbol{\mu}_i\right)^T\right] = \Sigma_i + \varepsilon I \tag{3}$$

Using the training samples with Gaussian noise, i.e., $\widehat{\boldsymbol{x}}_j^i$'s, ANN classifiers are trained. That is, the noise injected samples are repeatedly input to the network during the BP learning, independently for each iteration. Note that when $\varepsilon = 0$, ANN classifiers are

trained with the noiseless training samples. This means the conventional BP learning.

4 Experimental Results

We used the Ness data set [12] in which the dimensionality can be changed. The Ness data set consists of p-dimensional Gaussian data. The distribution parameters, μ_i and Σ_i, are shown below:

$$\mu_1 = [0, \cdots, 0]^T \qquad \mu_2 = [\Delta/2, \ \ 0, \cdots, 0, \ \ \Delta/2]^T$$

$$\Sigma_1 = I_p \qquad \Sigma_2 = \begin{bmatrix} I_{p/2} & O \\ O & \frac{1}{2} I_{p/2} \end{bmatrix}$$

where Δ is the Mahalanobis distance between class ω_1 and class ω_2, and I_p is the $p \times p$ identity matrix. In this data set, the true Bayes error can be controlled by varying the values of Δ and p.

It is recommended in general that the number of training samples per class should be at least five to ten times the dimensionality [13]. However, in many practical situations, the ratio of the training sample size to the dimensionality is small. Hence, our experiments were conducted in situations where the ratio is less than 5. On the other hand, in order to estimate the reliable generalization error of resulting ANN classifiers, we used 1000 test samples for each class. Note that the training samples are statistically independent of the test samples. The estimated generalization errors were averaged over 100 trials with the Monte Carlo method. Fresh samples were generated artificially by a computer on each trial. Through this paper, we assumed that each class has an equal prior probability.

Now, in order to evaluate the degree of improvement of the generalization ability, we will define the following criterion F:

$$F = \frac{E}{E + E_{noise}} \times 100, \tag{4}$$

where E is the averaged generalization error of the ANN classifier trained by the conventional BP learning, and E_{noise} denotes the averaged generalization error in the case of the noise injection. The result that $F > 50$ implies that use of the noise injection is effective in designing ANN classifiers. On the other hand, if $F < 50$, then it implies that the noise injection degrades the generalization ability of ANN classifiers.

4.1 Experiment 1

The purpose of this experiment is to study a behavior of the ANN classifier trained by the noise injection in high dimensions. Varying values of the dimensionality, we examined the influence of the noise injection with various values of variance ε on the generalization error. The following experiment was conducted.

No. of classes : 2
Values of Δ : 2, 8
Dimensionality p : 8, 16, 24, 32
Training sample size N : 32/class
Test sample size : 1000/class
Hidden unit size m : 8, 256
Variance ε : 0.1, 1.0, 5.0
Trials : 100

In this experiment, the true Bayes errors were ranging from 0.65% to 18.64%. Fig.1 shows the results. In almost all of results, the noise injection improves the generalization ability, even in high dimensions. These results suggest that the optimal value of ε should exist for a particular problem. Therefore, the selection of the value of ε is important in the noise injection.

4.2 Experiment 2

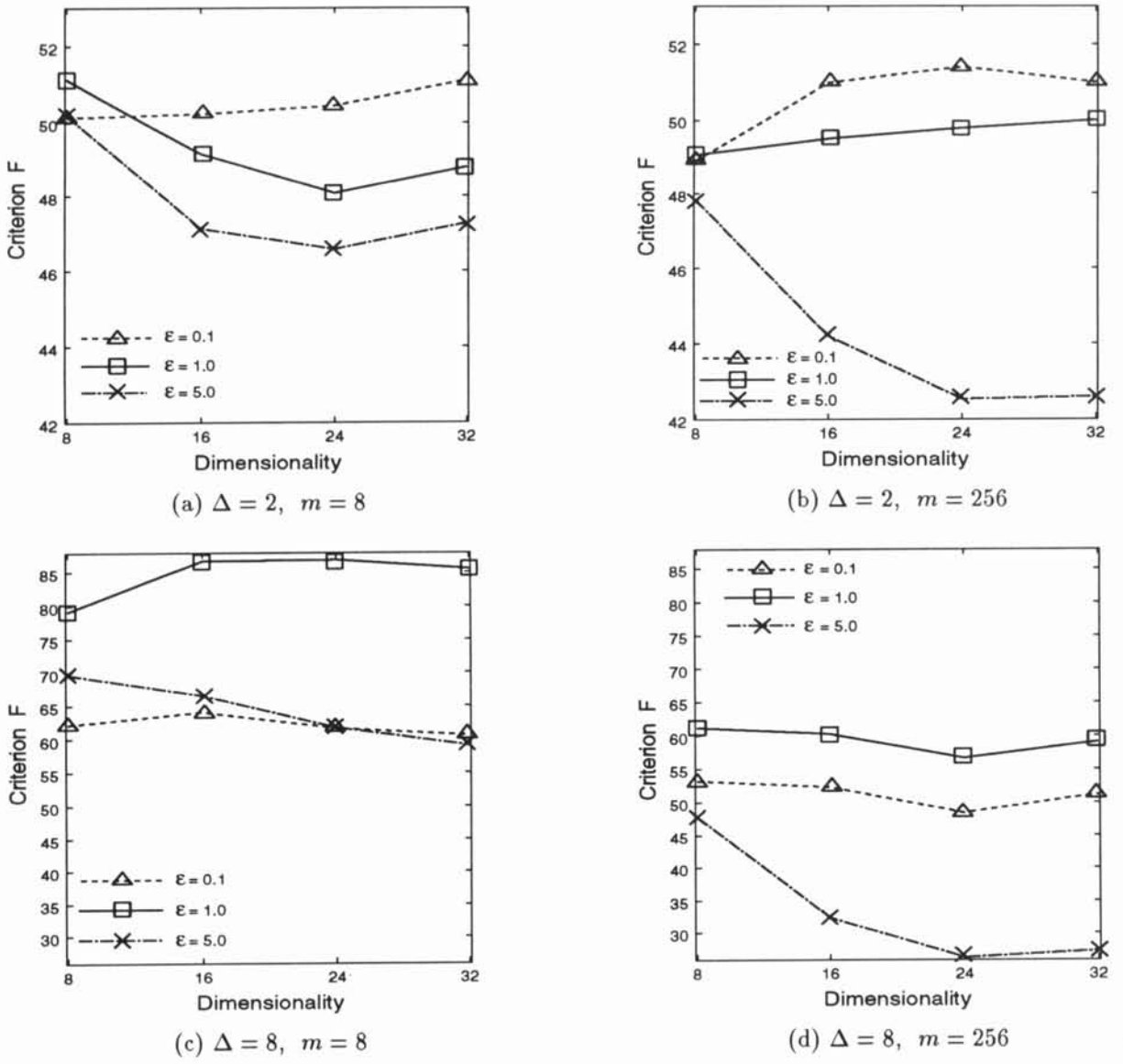
The purpose of this experiment is to study a relationship between the noise injection and the true Bayes error. Varying values of the true Bayes error, we examined the influence of the noise injection with various values of ε on the generalization error. Our experiment was conducted as follows.

No. of classes : 2
Values of Δ : 2, 3, 4, 5, 6, 8
Dimensionality p : 8, 32
Training sample size N : $N = p$/class
Test sample size : 1000/class
Hidden unit size m : 8, 256
Variance ε : 0.1, 1.0, 5.0
Trials : 100

In this experiment, the ratio of the training sample size to the dimensionality was set to one. Fig.2 shows the results. In all of results, the noise injection with $\varepsilon = 1.0$ improves the generalization ability, regardless of the dimensionality. It is also interesting to note that as the true Bayes error decreases, or as the hidden unit size decreases, the positive effect of the noise injection becomes clear.

5 Conclusions

We have investigated the effect of the noise injection on the generalization ability of ANN classifiers in high dimensions. Experimental results showed that if the value of ε is properly selected, the noise injection is a very useful method for improving the generalization ability of ANN classifiers, even in high dimensions. Therefore, special attention should be paid to the problem of selecting the optimal value of ε. Experimental results suggest that in general, the use of small values of ε (say, $\varepsilon \leq 1$) is recommended. We believe that this is independent of the hidden unit size, the dimensionality and the true Bayes error. Furthermore, it is shown that as the true Bayes error decreases, or as the hidden unit size decreases, the positive effect of the noise injection becomes clear.



(a) $\Delta = 2,\ m = 8$ (b) $\Delta = 2,\ m = 256$

(c) $\Delta = 8,\ m = 8$ (d) $\Delta = 8,\ m = 256$

Fig. 1: Dependence of the effect of the noise injection on the dimensionality.

References

[1] S. Raudys and A. K. Jain, "Small sample size problems in designing artificial neural networks ", in Artificial Neural Networks and Pattern Recognition: Old and New Connections, I.Sethi and A.K. Jain(eds.), Elsevier, 1991.

[2] R. P. W. Duin, "Superlearning capabilities of neural networks", Proc. of the 8th Scandinavian Conf. Image Analysis, Tromso, pp. 547-554, 1993.

[3] S. Raudys, "Why do multilayer perceptrons have favorable small sample properties?", in Pattern Recognition in Practice IV, E. S. Gelsema and L. N. Kanal (eds.), Elsevier Science B. V., pp. 287-298, 1994.

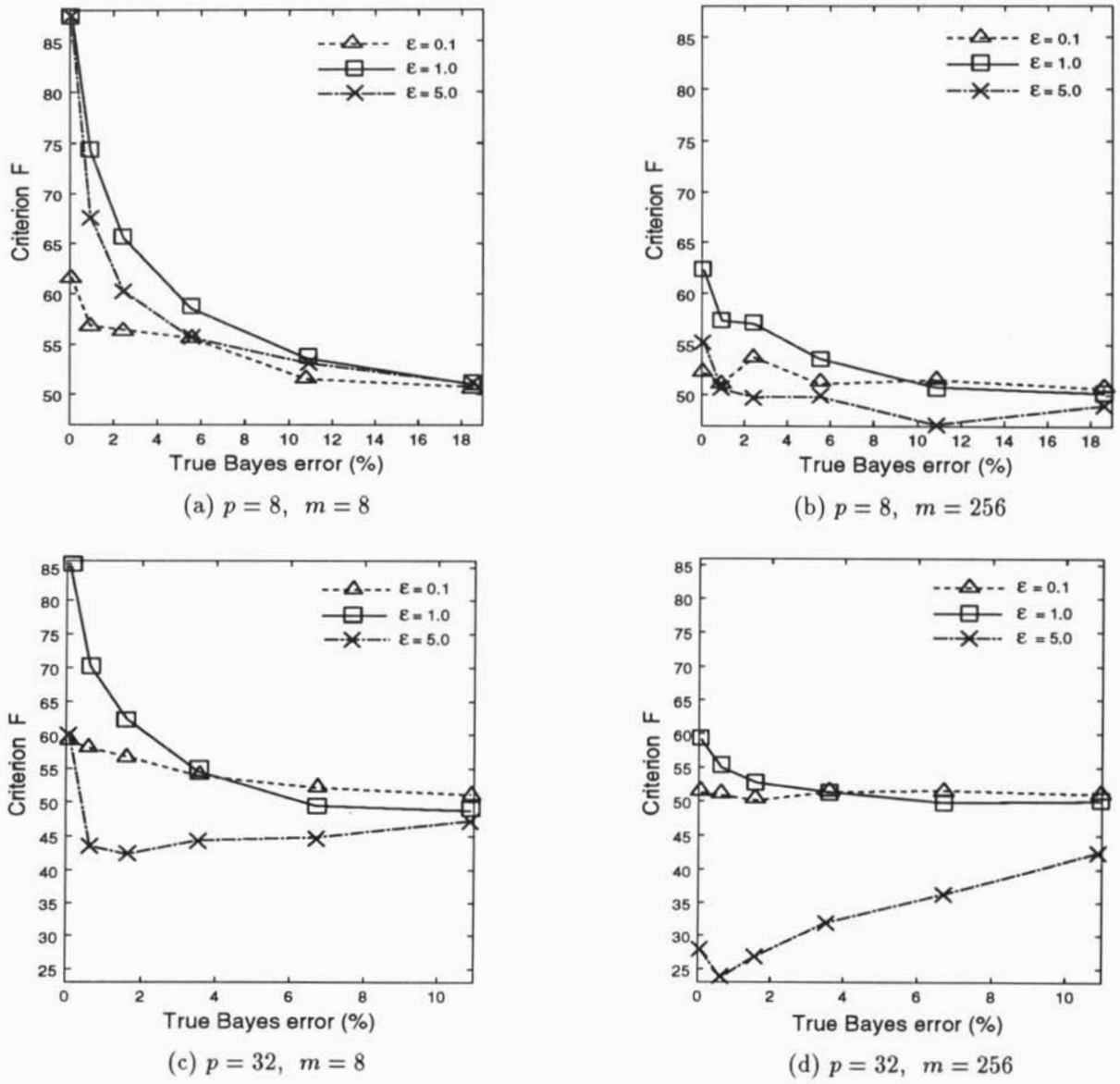
[4] D.C. Plaut, S.J. Nowlan and G.E. Hinton, "Experiments on learning by back propagation", Tech. Rep. CMU-CS-86-126, Carnegie-Mellon Univ., 1986.

[5] J. Sietsma and R.J.F. Dow, "Neural net pruning - Why and How",Proc. IEEE Int. Conf. Neural Networks, **1**, pp. 325-333, 1988.

[6] K. Matsuoka, "Noise injection into inputs in back-propagation learning", IEEE Trans. SMC-22, 3, pp. 436-440, 1992.

[7] Y. Grandvalet and S. Canu, "Comments on "Noise injection into inputs in back propagation learning" ", IEEE Trans. **SMC-25**, 4, pp. 678-681, 1995.

[8] L. Holmström and P. Koistinen, "Using additive noise in back-propagation training", IEEE Trans. Neural Networks, **3**, 1, pp. 24-38, 1992.



(a) $p = 8,\ m = 8$

(b) $p = 8,\ m = 256$

(c) $p = 32,\ m = 8$

(d) $p = 32,\ m = 256$

Fig. 2: Dependence of the effect of the noise injection on the true Bayes error.

[9] C. M. Bishop, "Training with noise is equivalent to Tikhonov regularization", Neural Computation, **7**, 1, pp. 108-116, 1995.

[10] Y. Hamamoto, Y. Mitani and S. Tomita, "On the effect of the noise injection in small training sample size situations", Proc. Int. Conf. Neural Information Processing, **1**, pp.626-628, Seoul, 1994.

[11] D.E. Rumelhart, G.E. Hinton and R.J. Williams, "Learning internal representations by error propagation", in Parallel Distributed Processing: Explorations in the Microstructure of Cognition, ch.8, MIT Press, 1986.

[12] J. Van Ness, "On the dominance of non-parametric Bayes rule discriminant algorithms in high dimensions", Pattern Recognition, Vol. 12, pp.355-368, 1980.

[13] A. K. Jain and B. Chandrasekaran, "Dimensionality and sample size considerations in pattern recognition practice", in Handbook of Statistics, Vol. 2, P. R. Krishnaiah and L. N. Kanal, eds., North-Holland, pp.835-855, 1982.