Sensitivity Analysis of Single Hidden-Layer Neural Networks with Threshold Functions

Sang-Hoon Oh and Youngjik Lee
Electronics and Telecommunications Research Institute, Research Department

Abstract

An important consideration when applying neural networks to pattern recognition is the sensitivity to weight perturbation or to input errors. In this letter, we analyze the sensitivity of single hidden-layer networks with threshold functions. In case of weight perturbation or input errors, the probability of inversion error for an output neuron is derived as a function of the trained weights, the input pattern, and the variance of weight perturbation or the bit error probability of the input pattern. The derived results are verified with a simulation of the Madaline recognizing handwritten digits. The result shows that the sensitivity of trained networks is far different from that of networks with random weights.

1. Introduction

It is well known that neural networks can achieve arbitrary input-output mapping through a series of weighted sums and nonlinear functions[1]. For applications of neural networks, especially to pattern recognition, the sensitivity to input errors and to weight perturbation must be considered. Well-trained neural networks have low sensitivity to input errors. If there are several networks trained for a given problem, we can select a neural network with the lowest sensitivity to input errors as the best network[2]. When we implement neural networks with hardware of limited precision, the sensitivity to weight perturbation can be a criterion for performance degradation[3].

Recently, there have been some studies on the sensitivity of neural networks through simulations or theoretical approaches[2, 3, 4, 5, 6, 7]. Choi et al. used the first order Taylor approximation method[2], however this has limited range of application. Xie and Jabri proposed a statistical quantization model under the assumption that the inputs, the trained weights, and the weighted sums are distributed uniformly in certain ranges[4]. Notedly, Stevenson et al. derived the sensitivity, assuming that the input patterns are uniformly distributed over the input space and also the weight vectors are[3]. However, the distribution of the population from which the input patterns are sampled is unknown in real problems and the trained weight vectors are dependent upon the training patterns.

In this point of view, Orzechowski et al. derived the sensitivity for a certain input pattern in neural networks with threshold functions[5]. Nevertheless, it is time-consuming to calculate the probability of each event that a subset of hidden neurons become inverted due to noises.

In this letter, we analyze the sensitivity of a neural network with threshold functions, called Madaline[8], for a certain input pattern. In section II, the Madaline is described briefly, and the error probability of an output neuron due to weight perturbation and that due to input errors are derived. The derived results are verified with a simulation of the Madaline recognizing handwritten digits in section III, and section IV concludes this letter.

2. The sensitivity of the Madaline

2.1. The Madaline architecture

The Madaline is a layered feedforward network of neurons with threshold functions. The output of the jth neuron in layer l, s_j^l, is given by

$$s_j^l = \begin{cases} +1 & \text{if } w_{j0}^l + \sum_i w_{ji}^l s_i^{l-1} \geq 0 \\ -1 & \text{otherwise,} \end{cases} \tag{1}$$

where w_{ji}^l is a weight connecting neuron i in layer $l-1$ to neuron j in layer l and w_{j0}^l is the bias weight. Thus, each of the weight vectors associated with the first-layer neurons dichotomizes input space. The outputs of the first layer serve as inputs to the following layer, and so on. As a result, the layered structure with appropriate weights determined through training can achieve arbitrary binary-to-binary mapping[8].

2.2. Output inversion due to weight perturbation

In this subsection, we derive the probability of inversion error for an output neuron when the trained weights are perturbed randomly with additive Gaussian noises.

Consider a well-trained single hidden-layer Madaline with N inputs, H hidden neurons, and M output neurons. For a given input pattern represented by $\mathbf{x} = [x_1, x_2, \ldots, x_N]$, the output of the jth hidden neuron is

$$h_j = \begin{cases} +1 & \text{if } a_j = w_{j0} + \sum_{i=1}^N w_{ji} x_i \geq 0 \\ -1 & \text{otherwise,} \end{cases} \tag{2}$$

where w_{ji} is a weight connecting x_i to h_j and w_{j0} is the bias. Also, the weighted sum to the kth output neuron is given by

$$b_k = v_{k0} + \sum_{j=1}^{H} v_{kj} h_j, \tag{3}$$

where v_{kj} is a weight connecting h_j to the kth output neuron and v_{k0} is the bias. Finally, b_k determines the output of the kth output neuron.

Assume that $\mathbf{W} = (w_{ji})$ and $\mathbf{V} = (v_{kj})$ are perturbed with independent, identically distributed(i.i.d.) Gaussian noises $N(0, \sigma_w)$ and $N(0, \sigma_v)$ respectively. Here, $N(\eta, \sigma)$ denotes the Gaussian distribution with mean η and standard deviation σ. Then a_j in (2) for given $\mathbf{x}$ is Gaussian because it is a summation of the Gaussian noises. The mean and the variance of a_j can be calculated as

$$\eta_{a_j} = w_{j0} + \sum_{i=1}^{N} w_{ji}x_i \quad \text{and} \quad \sigma_{a_j}^2 = \sigma_w^2(1 + \sum_{i=1}^{N} x_i^2). \tag{4}$$

The correlation coefficient r_{ij} between a_i and a_j is 0 since all w_{ji}'s are perturbed randomly without any dependency among them. After the threshold transformation in (2), the mean and the variance of h_j are

$$E[h_j] = 1 - 2G(-\tfrac{\eta_{a_j}}{\sigma_{a_j}}) \quad \text{and} \quad Var[h_j] = 4G(-\tfrac{\eta_{a_j}}{\sigma_{a_j}})[1 - G(-\tfrac{\eta_{a_j}}{\sigma_{a_j}})], \tag{5}$$

respectively, where $G(.)$ is the Gaussian distribution function. As i) $v_{ki}h_i$ and $v_{kj}h_j$ are indenpent, ii) $Var[\sum_{j=1}^{H} v_{kj}h_j] \to \infty$ as $H \to \infty$, and iii) all $v_{kj}h_j$ are bounded, $\sum_{j=1}^{H} v_{kj}h_j$ is asymptotically Gaussian by the central limit theorem[9]. Therefore, $b_k \approx N(\eta_{b_k}, \sigma_{b_k})$ with

$$\eta_{b_k} = v_{k0} + \sum_{j=1}^{H} v_{kj}E[h_j] \tag{6}$$

and

$$\sigma_{b_k}^2 = (1 + H)\sigma_v^2 + \sum_{j=1}^{H} v_{kj}^2 Var[h_j]. \tag{7}$$

Thus, the inversion probability of the kth output neuron due to the weight perturbation, for given $\mathbf{x}$, can be calculated as

$$P_I(\mathbf{x}) = \begin{cases} G(-\frac{\eta_{b_k}}{\sigma_{b_k}}) & \text{if } t_k = 1 \\ G(\frac{\eta_{b_k}}{\sigma_{b_k}}) & \text{if } t_k = -1, \end{cases} \tag{8}$$

where t_k is the output of the kth output neuron without weight perturbation.

2.3. Output inversion due to input errors

In applying the Madaline to real problems, it is important to understand the sensitivity of the Madaline outputs to changes in inputs. In this subsection, we derive the probability of inversion error for an output neuron when an input pattern $\mathbf{x}$ composed of -1 or $+1$ is subject to error.

Let $\mathbf{x}'$ denote a perturbed version of the input pattern $\mathbf{x}$. Assume that each element of $\mathbf{x}$ is inverted with probability p,

$$x'_i = \begin{cases} -x_i & \text{with probability } p \\ x_i & \text{with probability } 1-p. \end{cases} \tag{9}$$

Furthermore, assume that x'_i and x'_j are independent for $i \neq j$. As i) $Var[\sum_{i=1}^{N} w_{ji}x'_i] \to \infty$ as $N \to \infty$ and ii) all $w_{ji}x'_i$ are bounded, $a_j \approx N(\eta_{a_j}, \sigma_{a_j})$ for large N by the central limit theorem, where

$$\eta_{a_j} = w_{j0} + (1-2p)\sum_{k=1}^{N} w_{jk}x_k \tag{10}$$

and

$$\sigma_{a_j}^2 = 4p(1-p)\sum_{k=1}^{N} w_{jk}^2. \tag{11}$$

Also, a_i and a_j are jointly Gaussian with the correlation coefficient

$$r_{ij} = \frac{\sum_{k=1}^{N} w_{jk}w_{ik}}{\sqrt{\sum_{k=1}^{N} w_{jk}^2}\sqrt{\sum_{k=1}^{N} w_{ik}^2}}. \tag{12}$$

After the threshold transformations as in (2), the mean and the variance of h_j are given by substituting (10) and (11) into (5). The mean of h_ih_j can be derived as

$$E[h_ih_j] = 2\int_{-\infty}^{-\eta_{a_j}} f(v)dv - 2\int_{-\eta_{a_j}}^{\infty} f(v)dv + 1 - 2G(-\tfrac{\eta_{a_j}}{\sigma_{a_j}}) \tag{13}$$

where

$$f(v) = \frac{1}{\sigma_{a_j}\sqrt{2\pi}}\exp[-\frac{v^2}{2\sigma_{a_j}{}^2}]G(-\tfrac{1}{\sqrt{1-r_{ij}^2}}(\tfrac{\eta_{a_i}}{\sigma_{a_i}} + \tfrac{r_{ij}v}{\sigma_{a_j}})). \tag{14}$$

Thus, the correlation coefficient between h_i and h_j is by definition

$$r_{ij}^{(h)} = \frac{C_{ij}^{(h)}}{Var[h_i]Var[h_j]} = \frac{E[h_ih_j] - E[h_i]E[h_j]}{Var[h_i]Var[h_j]}. \tag{15}$$

However, $|r_{ij}^{(h)}|$ may have small value because of the orthogonal tendency of the weight vectors to hidden neurons and the threshold functions which reduce $|r_{ij}|$[10, 11, 12]. Furthermore, $Var[\sum_{j=1}^{H} v_{kj}h_j] \to \infty$ as $H \to \infty$ and all $v_{kj}h_j$ are bounded, the central limit theorem can be established that $b_k \approx N(\eta_{b_k}, \sigma_{b_k})$ for large H. The mean η_{b_k} is as in (6) and the variance is

$$\sigma_{b_k}^2 = \sum_{j=1}^{H} v_{kj}^2 Var[h_j] + 2\sum_{j=1}^{H-1}\sum_{i=j+1}^{H} v_{ki}v_{kj}C_{ij}^{(h)}. \tag{16}$$

Thus, the inversion probability of the kth output neuron due to the input errors, for given $\mathbf{x}$, can be calculated by substituting the above η_{b_k} and σ_{b_k} into (8).

3. Simulation results

In order to verify the results of the previous section, the sensitivity is simulated on the single hidden-layer Madaline recognizing handwritten digits. 3,000 handwritten digitized images gathered from 40 persons are used for training after skeletonization. A digit image is 16x16 pixels and each pixel takes on binary values of either $+1$ or -1. The Madaline, which consists of 256 inputs, 35 hidden neurons, and 10 output neurons, is trained to output the target pattern associated with the input pattern. The target pattern consists of only one '$+1$' element and nine '-1' elements such that the index of '$+1$' element represents the digit class of the input pattern. The output neuron corresponding to the index of '$+1$' element is named as "target neuron."

After successful training, we add zero-mean i.i.d. Gaussian noises to the trained weights, and measure the inversion probability, P_I, of the target neuron. Since the range of w_{ji} and that of v_{kj} are different, we use noises with the same relative standard deviation which is defined as

$$\sigma_R \triangleq \sigma_w / Aver[|w_{ji}|] = \sigma_v / Aver[|v_{kj}|]. \tag{17}$$

Here, $Aver[|w_{ji}|]$ and $Aver[|v_{kj}|]$ denote the average of absolute weight values. For investigating P_I of the target neuron for class "4," 70 patterns are randomly selected from 300 training patterns of class "4." When a pattern is presented to the input layer, P_I of the target neuron for given σ_R is estimated by a simulation of 100 times, and the simulated results for the 70 patterns are averaged to draw the solid line in Fig. 1. The dotted line of Fig. 1 shows the calculated results using (8). The simulation results closely match with the calculations. Assume $\mathbf{n}$ is a random variable with $N(0, \sigma)$, then the mean of $|\mathbf{n}|$ is $\sigma\sqrt{2/\pi}$. Thus, P_I in Fig. 1 whose horizontal axis is scaled with $\sqrt{2/\pi}(\approx 0.8)$ may be compared to that of networks with randomly assigned weight values (Fig. 7 in [3]). While the former shows robustness when $\sigma_R < 0.2$, the latter shows sharp increase. This fact implies that the weight accuracy is not critical in hardware implementation of trained Madalines.

Fig. 2 shows P_I of the target neuron for class "4" when input patterns are subject to errors. The horizontal axis denotes the bit error probability p in (9). The curves shown in Fig. 2 represents averages over the 70 patterns. As can be seen from Fig. 2, the simulation results are in good agreement with the results predicted by the calculations. In application of the Madaline to pattern recognition, $|r_{ij}^{(h)}|$ has small value and may be omitted to calculate P_I. This can be supported by Fig. 2 where the dotted line(P_I with $r_{ij}^{(h)}$ omitted) is close to the thin solid line(P_I with $r_{ij}^{(h)}$). Therefore, it is reasonable that Orzechowski et al. did not consider $r_{ij}^{(h)}$ to derive the sensitivity to input errors[5]. They, however, considered the events that less than 3 hidden neurons become inverted due to input errors or weight perturbation; for example, there are 631 events in case of the Madaline with 35 hidden neurons. It is time-consuming to estimate the probability of each event.

In contrast to the monotonically increasing characteristics in Fig. 1, the functional dependence of the probability of inversion error on the probability of bit error in Fig. 2 appears to be characterized by a third order polynomial in p, which is typical of all classes. This is because the ratio η_{a_j}/σ_{a_j} from (10) and (11) has four kinds of functional shape(increasing, decreasing, convex, and concave) for $0 < p < 1$ according to the values of w_{j0} and $\sum_{k=1}^{N} w_{jk}x_k$. Since $G(.)$ is a monotonic increasing function, $E[h_j]$ in (5) can be traced as a function of p according to the shape of η_{a_j}/σ_{a_j}. Also, the variance $Var[h_j]$ in (5) can be traced as a concave function of p. Therefore, the variance $\sigma_{b_k}^2$ in (16) is concave, if we neglect the second term of (16). The mean η_{b_k} is a weighted sum of $E[h_j]$. After all, the ratio η_{b_k}/σ_{b_k} is not a monotonic function of p and $G(-\eta_{b_k}/\sigma_{b_k})$ shows a curve as in Fig. 2.

4. Conclusion

In this letter, we derive the sensitivity of single hidden-layer neural networks with threshold functions, called "Madaline," for a certain input pattern, and verify the results with a simulation of the Madaline recognizing handwritten digits. The sensitivity to weight perturbation in Fig. 1 shows that the weight accuracy is not critical in hardware implementation of trained Madalines. The sensitivity to probabilistic inversion error of binary input pattern is not monotonically increasing as shown in Fig. 2.

5. Acknowledgements

This research was partly funded by Korea Telecom. and the Ministry of Communication, Korea. The authors wish to thank Dr. El Hang Lee for his continuing support and guidance and Dr. Jin Young Choi for his helpful discussions. Also, the authors wish to thank the reviewer for his/her helpful criticisms.

References

[1] Y. Ito, "Approximation of continuous functions on r^d by linear combinations of shifted rotations of a sigmoid function with and without scaling," *Neural Networks*, vol. 5, pp. 105–115, 1992.

[2] J. Y. Choi and C.-H. Choi, "Sensitivity analysis of multilayer perceptron with differentiable activation functions," *IEEE Trans. Neural Networks*, vol. 3, pp. 101–107, 1992.

[3] M. Stevenson, R. Winter, and B. Widrow, "Sensitivity of feedforward neural networks to weight errors," *IEEE Trans. Neural Networks*, vol. 1, pp. 71–90, Mar. 1990.

[4] Y. Xie and M. A. Jabri, "Analysis of the effects of quantization in multilayer neural networks using a statistical model," *IEEE Trans. Neural Networks*, vol. 3, pp. 334–338, Mar. 1992.

[5] N. S. Orzechowski, S. R. T. Kumara, and C. R. Das, "Performance of multilayer neural networks in binary-to-binary mappings under weight errors," *Proc. ICNN'93 San Francisco*, pp. 1684–1689, Mar. 1993.

[6] A. von Lehmen, E. G. Paek, P. F. Liao, A. Marrakchi, and J. S. Patel, "Influence of interconnection weight discretization and noise in an optoelectronic neural networks," *Opt. Lett.*, vol. 14, pp. 928–930, Sept. 1989.

[7] S.-H. Oh and Y. Lee, "An analysis on the classification performance of multilayer perceptrons," *Proc. IJCNN'92 Beijing*, vol. II, pp. 787–792, Nov. 1992.

[8] B. Widrow, R. G. Winter, and R. A. Baxter, "Layered neural nets for pattern recognition," *IEEE Trans. Acoust., Speech, Signal Proc.*, vol. 36, pp. 1109–1118, July 1988.

[9] A. Papoulis, *Probability, Random variables, and Stochastic Processes.* McGraw-Hill, New York, 2nd edition, 1984.

[10] S.-H. Oh and Y. Lee, "Effect of nonlinear functions on correlation between weighted sums in multilayer perceptrons," *IEEE Trans. Neural Networks*, accepted for publication.

[11] B. Widrow and M. A. Lehr, "30 years of adaptive neural networks: Perceptron, Madaline and backpropagation," *Proc. IEEE*, vol. 78, pp. 1415–1442, Oct. 1990.

[12] Y. Lee, M. W. Kim, H. K. Song, and S. C. Park, "Analysis of classification performance for Hopfield network with predefined correlated exemplar patterns," *Proc. IJCNN'90 San Diego*, vol. I, pp. 803–808, June 1990.

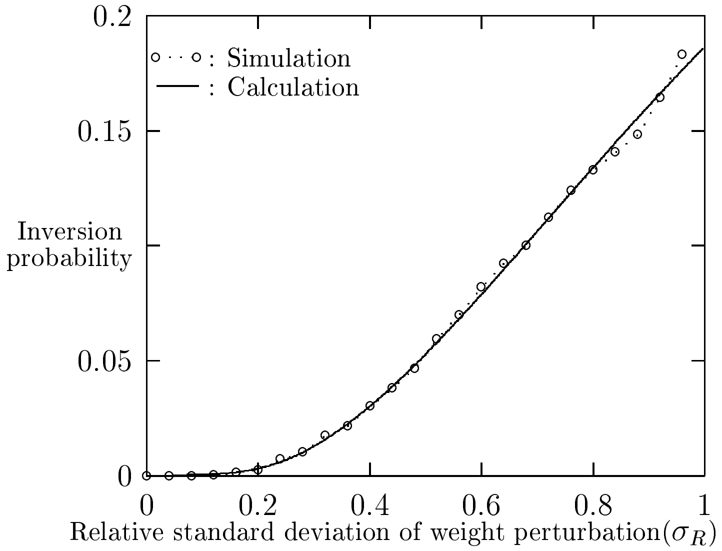

Figure 1: Inversion probability of the target neuron versus relative standard deviation of weight perturbation for class “4” ($\sigma_R = \sigma_w / Aver[|w_{ji}|] = \sigma_v / Aver[|v_{kj}|]$; $Aver[|w_{ji}|]$ is the average of absolute weight values from the input layer to the hidden layer and $Aver[|v_{kj}|]$ is that from the hidden layer to the output layer).

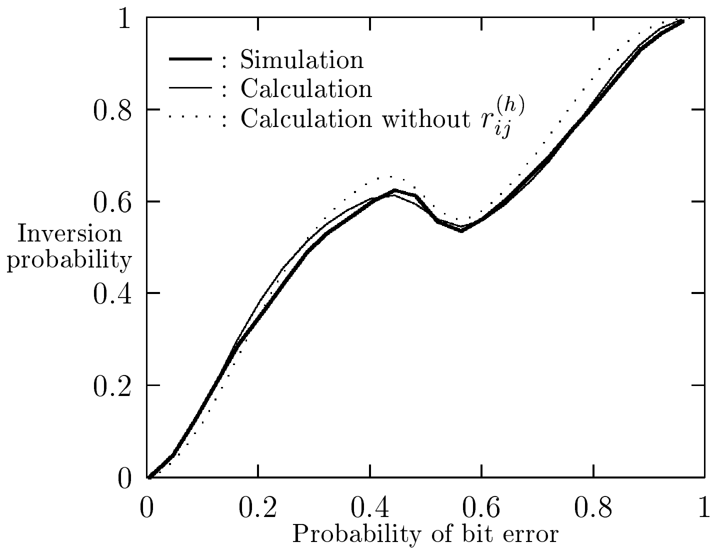

Figure 2: Inversion probability of the target neuron versus input errors for class “4.” $r_{ij}^{(h)}$ is the correlation coefficient between neuron i and neuron j in the hidden layer.