Error and Variance Bounds on Sigmoidal Neurons with Weight and Input Errors

David Lovell, Peter Bartlett & Tom Downs
Intelligent Machines Laboratory, Department of Electrical Engineering
University of Queensland, Queensland 4072, Australia

Abstract

We derive bounds on the expectation and variance of errors at the output of a multi-layer feedforward neural network with perturbed weights and inputs. It is assumed that errors in weights and inputs to the network are statistically independent and small. The bounds obtained are applicable to both digital and analog network implementations and are shown to be of practical value.

Introduction: This paper establishes bounds on the effect of errors in the weights and inputs of a neural network which arise in analog and limited precision digital realizations. Other researchers have investigated this problem [1, 2, 3] but their approaches have been aimed at *approximating* (rather than *bounding*) output error statistics.

The method described in this paper needs no information about the distribution of the weight and input errors other than bounds on their expectation and variance. This is a more general approach than that of Xie and Jabri [1] who assumed the weights to be uniformly distributed over the quantization width. Xie and Jabri also made assumptions about the distribution of the output of each neuron (a restriction which we do not impose) and used a piecewise linear transfer function in their derivation (whereas we employ the more commonly used sigmoid function).

The approximate effect of weight and input perturbations on the output variance has been expressed in system sensitivity terms by Choi and Choi [3], but these authors did not consider the effect of perturbations on the expected value of the output. The results we obtain indicate that the error in the expected value of the output can be significant and should not be disregarded. We also present results that show the bounds we obtain to be of practical value.

Preamble: Consider a layer of n_{out} neurons, each of which is fully connected to n_{in} inputs. The value of the i^{th} weight is given by $\hat{w}_i = w_i + e_{\hat{w}_i}$ where w_i is the nominal value of the weight and the error term has zero mean and a fixed (known) variance: $\mathrm{E}[e_{\hat{w}_i}] = 0$, $\mathrm{var}(e_{\hat{w}_i}) = v_w$.

The i^{th} input is denoted by x_i and its value is given by $\hat{x}_i = x_i + e_{\hat{x}_i}$, where the expectation and variance of the error term $e_{\hat{x}_i}$ are bounded[1]: $|\mathrm{E}[e_{\hat{x}_i}]| \leq e_x$, $\mathrm{var}(e_{\hat{x}_i}) \leq v_x$.

The following discussion is based around neurons described by

$$\begin{aligned} a &= \sum_{i=1}^{n_{in}} w_i x_i & (1) \\ y &= \tanh(a/n_{in}). & (2) \end{aligned}$$

Normalizing the activation a of the neuron by its fan-in ensures that, in a fixed precision implementation, the activation of every neuron has the same range and precision, regardless of the number of inputs to it. It will be seen later that normalization of activation results in tighter statistical bound on the neuron's output y.

Bounding the Expectation and the Variance of the Activation Error: We wish to determine how the estimated activation deviates from the activation given by Equation (1) in a neuron with errors in its inputs and weights. We start by considering the expectation of the activation error

$$\mathrm{E}[e_{\hat{a}}] = \mathrm{E}\left[\sum_{i=1}^{n_{in}} (w_i x_i + w_i e_{\hat{x}_i} + e_{\hat{w}_i} x_i + e_{\hat{w}_i} e_{\hat{x}_i})\right] - \sum_{i=1}^{n_{in}} w_i x_i. \quad (3)$$

Since $\mathrm{E}[e_{\hat{w}_i}] = 0$, this reduces to

$$\begin{aligned} |\mathrm{E}[e_{\hat{a}}]| &= \sum_{i=1}^{n_{in}} w_i \mathrm{E}[e_{\hat{x}_i}] \\ &\leq n_{in} W e_x = e_a & (4) \end{aligned}$$

where W is the magnitude of the largest weight connecting the neuron to its input (this results in a general bound that can be improved upon for a specific set of weights).

Now we bound the variance of the activation error

$$\mathrm{var}(e_{\hat{a}}) = \sum_{i=1}^{n_{in}} \mathrm{var}(\hat{w}_i \hat{x}_i) \quad (5)$$

where

$$\begin{aligned} \mathrm{var}(\hat{w}_i \hat{x}_i) &= w_i^2 \mathrm{var}(e_{\hat{x}_i}) + \mathrm{var}(e_{\hat{w}_i}) x_i^2 + \mathrm{var}(e_{\hat{w}_i} e_{\hat{x}_i}) \\ &\leq W^2 v_x + v_w + v_w(v_x + e_x^2). & (6) \end{aligned}$$

[1]The expected error value is not assumed to be zero for inputs because of the way in which errors propagate through a network. This is justified by the analysis which follows.

(We have simplified $\mathrm{var}(e_{\hat{w}_i} e_{\hat{x}_i})$ using the fact that that $\mathrm{E}[e_{\hat{w}_i}] = 0$ and $\mathrm{var}(XY) = (\mathrm{var}(X) + \mathrm{E}[X]^2)\mathrm{var}(Y)$ for X, Y independent and $\mathrm{E}[Y] = 0$). Equation (6) gives the i^{th} term in the summation of Equation (5) and lets us bound the variance of $e_{\hat{a}}$ by

$$\mathrm{var}(e_{\hat{a}}) \;\leq\; n_{in}(W^2 v_x + v_w(1 + v_x + e_x^2)) \;=\; v_a. \tag{7}$$

Bounding the Expectation of the Output Error: Now that the expected value and variance of the activation error have been bounded it is possible to bound the expected value of the output. In the following derivation, we refer to the transfer function $\tanh(\cdot)$ as $f(\cdot)$.

The expected value of the output is given by $\mathrm{E}[\hat{y}] = \mathrm{E}[f(a + e_{\hat{a}})]$. By Taylor's expansion, this may be written as

$$\begin{aligned} \mathrm{E}[\hat{y}] &= \mathrm{E}[f(a) + f'(a)e_{\hat{a}} + \frac{f''(a)}{2!}e_{\hat{a}}^2 + \cdots + \frac{f^{(n)}(a)}{n!}e_{\hat{a}}^n + \cdots] & (8) \\ &\leq f(a) + f'_{max}e_{\hat{a}} + \frac{f''_{max}}{2!}e_{\hat{a}}^2 + \cdots + \frac{f^{(n)}_{max}}{n!}e_{\hat{a}}^n + \cdots & (9) \end{aligned}$$

where $f^{(n)}_{max}$ is the maximum value of the n^{th} derivative of $f(\cdot)$.

The estimate of the activation is the sum of n_{in} random variables, and (by the Central Limit Theorem) will become approximately normally distributed as n_{in} becomes sufficiently large[2]. The higher order terms of Equations (8) and (9) will, therefore, approach zero so it is reasonable to use only the first and second terms of Equation (9) to bound the estimate of the activation. From Equation (2) we know that

$$\begin{aligned} |f'(a)| &\leq \frac{1}{n_{in}} = f'_{max} & (10) \\ |f''(a)| &\leq 0.7698\frac{1}{n_{in}^2} = f''_{max}, & (11) \end{aligned}$$

hence,

$$\begin{aligned} |\mathrm{E}[\hat{y}] - f(a)| &\leq \frac{1}{n_{in}}|\mathrm{E}[e_{\hat{a}}]| + 0.3849\frac{1}{n_{in}^2}\mathrm{var}(e_{\hat{a}}) & (12) \\ &= \frac{e_a}{n_{in}} + 0.3849\frac{v_a}{n_{in}^2} & (13) \end{aligned}$$

Bounding the Variance of the Output: An obvious property of the sigmoidal transfer function (Equation (2)) is that $f^2(a) \leq a^2/n_{in}^2$. Hence, the variance of the output error

[2] Papoulis [4, p. 214] suggests that $n_{in} \geq 30$ is adequate

can be bounded:

$$\begin{aligned} \mathrm{var}(e_{\hat{y}}) &= \mathrm{var}(\hat{y}) \\ &= \mathrm{E}[f^2(\hat{a})] - \mathrm{E}^2[f(\hat{a})] \;\leq\; \frac{v_a}{n_{in}^2}. \end{aligned} \tag{14}$$

Application of Bounds to a Specific Network Structure: The following results apply to a network that receives noiseless inputs but uses weights throughout with variance V_w (this is typical of a fixed precision digital implementation). The network has ℓ layers and the number of inputs to the i^{th} layer is denoted n_i. Let the bounds on the expected value and the variance of the output error at the i^{th} layer be written e_i, v_i respectively. Equations (4), (6), (13) and (14) can be used to derive the recursion relations

$$e_i = We_{i-1} + 0.3849 v_i \tag{15}$$

$$v_i = \frac{1}{n_i}(W^2 v_{i-1} + V_w(1 + v_{i-1} + e_{i-1}^2)). \tag{16}$$

The nature of the inputs to the network lets us calculate the first terms of these relations:

$$e_1 = 0.3849\frac{V_w}{n_1} \tag{17}$$

$$v_1 = \frac{V_w}{n_1}. \tag{18}$$

Equations (15) to (18) can be used to obtain bounds on the expectation and variance of output values for arbitrary feedforward networks (so long as the activation and transfer function of Equations (1) and (2) apply). With certain assumptions, an interesting closed form expression can be derived. The major assumption is that the number of nodes in any hidden layer n_i is related to the maximum weight by $n_i \geq bW^2$, where b is a constant greater than unity (note that this applies to a large class of feedforward networks). With this supposition, one can show, after some algebra, that:

$$e_\ell \leq 0.3849\frac{V_w W^{\ell-1}}{n_\ell}(1 + \frac{bc}{(b-1)(W-1)}) \tag{19}$$

$$v_\ell \leq \frac{V_w}{n_\ell}\frac{bc}{b-1}, \tag{20}$$

where $c = 1 + v_\ell + e_\ell^2$.

Note that the bound on e_ℓ is exponentially dependent upon the number of layers in the network. This indicates that results such as those in [3], which treat variance as the

8 Bits	e_y	v_y	4 Bits	e_y	v_y
Layer 1	0.000002557	0.000006643	Layer 1	0.0006546	0.001701
Layer 2	0.00001362	0.000008813	Layer 2	0.003488	0.002259
Layer 3	0.00005814	0.000009521	Layer 3	0.01489	0.002442
Layer 4	0.0002363	0.000009752	Layer 4	0.06053	0.002503
Layer 5	0.0009491	0.000009828	Layer 5	0.2431	0.002528

Table 1: Output error and variance bounds for networks with quantized weights.

sole indicator of network performance, should be treated with caution. We remark that the proportionality between v_ℓ and the variance of the network weights in Equation (20) is in agreement with Choi and Choi's approximation [3].

Table 1 shows bounds on the expectation and variance of the output error (calculated using Equations (15) to (18)) at each layer of a five layer, fully interconnected network with 49 nodes in each layer (as in [2, 3]). The weights are quantized into N-bit representations over the range $[-W, W] = [-4, 4]$; thus, the error in the weights is uniformly distributed over the quantization width (*ie.* the size of the least significant bit) and $\mathrm{E}[e_{\hat{w}_i}] = 0$ and $v_w = (4 \cdot 2^{-(N-1)})^2/3$. We assume that the inputs to the network have no errors.

Table 1 shows that even a 4 bit implementation of this network will give acceptable performance, despite the exponential increase in the output error bound with the number of layers. We point out that the values shown in Table 1 are bounded by Equations (19) and (20).

Conclusions: This paper has described bounds on the expectation and variance of the output error in a neural network with perturbed weights and inputs. The bounds we have obtained are in general agreement with approximate results obtained elsewhere. The exponential dependence of the output error bound on the number of layers suggests that performance estimates of neural networks with perturbed weights and inputs should not disregard the effect of this quantity.

References

[1] XIE, Y., and JABRI, M. A.:'Analysis of Effects of Quantisation in Multilayer Neural

Networks Using Statistical Model', *Electronics Letters*, 1991, **27**, pp. 1196–1198.

[2] STEVENSON, M., *et al.*:'Sensitivity of Feedforward Neural Networks to Weight Errors', *IEEE Trans. Neural Networks*, 1990, **1**, pp. 71–80.

[3] CHOI, J. Y., and CHOI, C.:'Sensitivity Analysis of Multilayer Perceptron with Differentiable Activation Functions', *IEEE Trans. Neural Networks*, 1992, **3**, pp 101–107.

[4] PAPOULIS, A.: 'Probability, Random Variables, and Stochastic Processes' (3rd Edition, McGraw-Hill 1991).