

自乗値ペナルティ項を用いた2次学習アルゴリズム

齊藤和巳^{†,☆} 中野良平[†]

3層ニューラルネットの教師あり学習目的関数に3種のペナルティ項を付加したときのそれぞれの学習性能を、1次と2次の学習アルゴリズムを用いて比較評価する。非線型回帰問題と多クラス分類問題を用いた実験では、自乗値ペナルティ項と2次学習アルゴリズムの組合せが、適切なペナルティ係数値において、最も高い汎化性能を示すとともに、最も高速に収束したことを示す。また、ペナルティ項の種類により学習性能がなぜ異なるかを誤差曲面の観点から論じる。最後に、最適なペナルティ係数を決定するのに交差検証法が有効に利用できることを示す。

Second-order Learning Algorithm with Squared Penalty Term

KAZUMI SAITO[†] and RYOHEI NAKANO[†]

This paper compares three penalty terms with respect to the efficiency of supervised learning, by using first- and second-order learning algorithms. Our experiments showed that for a reasonably adequate penalty factor, the combination of the squared penalty term and the second-order learning algorithm drastically improves the convergence performance in comparison to the other combinations, at the same time bringing about excellent generalization performance. Moreover, in order to understand how differently each penalty term works, an error surface evaluation is described. Finally, we show how cross-validation can be applied to find an optimal penalty factor.

1. はじめに

ニューラルネット学習や機械学習における最重要課題の1つは、その汎化性能をいかにして向上させるかである。汎化 (generalization) とは、与えられた訓練事例を学習するだけで、未知の新たな事例に対してもできるだけ正確な (自乗誤差や分類誤り率が小さな) 出力を得られるようにすることである。

ニューラルネットの学習目的関数に、訓練事例を正確に学習するための誤差項だけでなく、結合重みの値が大きくなることを抑制するペナルティ項を付加することにより、多くの場合、汎化能力の高いニューラルネットを得られることが観測され、いくつかの視点に基づくペナルティ項が提案されている。それらは、weight-decay⁶⁾, regularization¹¹⁾, function-smoothing²⁾, weight-pruning^{5),8)}, そして Bayesian priors^{9),16)} などであり、単純な算術式で定義されるものや、高次微係数を用いて定義されるものもある。これらのペナル

ティ項に対する最も重要な評価尺度は、すでに述べたように、いかに汎化性能を向上できるかであるが、規模の大きな現実問題では、ペナルティ項を付加したことによる学習効率も重要な評価尺度となる。すなわち、収束までに多くの計算量を必要とするペナルティ項では、このような現実問題に適用することがきわめて困難になるからである。また、ペナルティ項の影響は、学習アルゴリズムのタイプに依存することが予想されるので、さまざまな比較評価が必要である。しかし、定量的で網羅的な評価は今後取り組むべき研究課題である¹³⁾。

本稿では、1次と2次の学習アルゴリズムを用いて、3種のペナルティ項を付加したときの学習性能について比較評価する。2章では、ペナルティ項を用いた学習の枠組みについて説明するとともに、ペナルティ項付き2次学習アルゴリズムについて述べる。3章では、非線型回帰問題を用いた実験結果を示す。4章では、現実の分類問題を用いた実験結果を示す。5章では、ペナルティ項の種類によりなぜ学習性能が異なるかを誤差曲面の観点から論じる。6章では、最適ペナルティ係数を決定するのに交差検証法が有効に利用できることを示す。

[†] NTT コミュニケーション科学研究所
NTT Communication Science Laboratories

[☆] 現在、NTT 研究開発推進部

Presently with NTT R&D Management Department

2. ペナルティ項を用いた学習

2.1 フレームワーク

$\{(\mathbf{x}_1, \mathbf{y}_1), \dots, (\mathbf{x}_m, \mathbf{y}_m)\}$ を事例集合, \mathbf{x}_t をニューラルネットへの n 次元入力ベクトル, \mathbf{y}_t をニューラルネットからの k 次元目標出力値ベクトルとする. また, 3層ニューラルネットにおいて, h を中間ユニット数, \mathbf{w}_j ($j = 1, \dots, h$) を全入力ユニットから中間ユニット j への結合重み, $\mathbf{c}_i = (c_{i0}, \dots, c_{ih})^T$ を全中間ユニットから出力ユニット i への結合重みとする. ここで, w_{j0} と c_{i0} はバイアスであり, $x_{t0} = 1$ とする. ただし, \mathbf{a}^T はベクトル \mathbf{a} の転置ベクトルを表す. 以下では, 全結合重み $(\mathbf{c}_1^T, \dots, \mathbf{c}_k^T, \mathbf{w}_1^T, \dots, \mathbf{w}_h^T)^T$ を $\Phi = (\phi_1, \dots, \phi_N)^T$ で表し, Φ の次元を $N (= h(n+1) + k(h+1))$ とする. そして, 出力ユニット i の出力を以下で定義する.

$$z_i(\mathbf{x}_t; \Phi) = c_{i0} + \sum_{j=1}^h c_{ij} \sigma(\mathbf{w}_j^T \mathbf{x}_t).$$

ただし, $\sigma(u)$ はシグモイド関数 $\sigma(u) = 1/(1 + e^{-u})$ である.

まず, k 次元回帰問題を考える. 各目標出力値が独立に等分散ノイズを含むとすれば, 学習誤差項は自乗誤差を用いて以下のように定義できる.

$$f(\Phi) = \frac{1}{2} \sum_{t=1}^m \sum_{i=1}^k (y_{ti} - z_i(\mathbf{x}_t; \Phi))^2.$$

次に, k クラス分類問題を考える. ただし, 目標出力値はクラスラベルなので $y_{ti} \in \{0, 1\}$ である. 各事例がどれか1つのクラスに属するとすれば, 学習誤差項は soft-max 関数と cross-entropy を用いて以下のように定義できる.

$$f(\Phi) = - \sum_{t=1}^m \sum_{i=1}^k y_{ti} \log \left(\frac{\exp(z_i(\mathbf{x}_t; \Phi))}{\sum_{p=1}^k \exp(z_p(\mathbf{x}_t; \Phi))} \right).$$

なお, 分類問題でも, ニューラルネットの出力 z_i にシグモイド関数を施し, 目標出力値 y_i との差の自乗和を学習誤差項とする方法もある. しかし, z_i が入力事例に対する線型出力の場合には, 結合重みに対する2次微分 (Hesse) 行列は, cross-entropy を用いれば非負定値であるが, 自乗誤差ではそれが成立しないことが容易に確認できる. つまり, cross-entropy を用いれば, 学習誤差項の最小値は一意的である. z_i がニューラルネットによる非線型出力の場合にも, cross-entropy を用いて学習誤差項を定義すれば, 一般に学習が容易になると考えられるので, 上記のような学習誤差項と

表1 教師あり学習アルゴリズムの分類

Table 1 Classification of supervised learning algorithms.

教師あり 学習法		探索幅	
		固定	変動
探索 方向	1次	BP法など	Silva-Almeida法など
	2次	Newton法	SCG法, BPQ法など

した.

本稿では, 次の3種のペナルティ項を考える.

$$\begin{aligned} \Omega_1(\Phi) &= \frac{1}{2} \sum_{i=1}^N \phi_i^2, \\ \Omega_2(\Phi) &= \sum_{i=1}^N |\phi_i|, \\ \Omega_3(\Phi) &= \frac{1}{2} \sum_{i=1}^N \frac{\phi_i^2}{1 + \phi_i^2}. \end{aligned}$$

以下では, Ω_1, Ω_2 , および, Ω_3 のそれぞれを, 自乗項, 絶対値項, および, 正規項と呼ぶ. よって, ニューラルネットの学習目的関数は以下で定義できる.

$$F_i(\Phi) = f(\Phi) + \mu \Omega_i(\Phi). \quad (1)$$

ただし, μ はペナルティ係数である.

2.2 学習アルゴリズム

ニューラルネットの教師あり学習アルゴリズムには多くの提案があり, それらのすべてについてペナルティ項を付加したときの学習性能を評価するのは不可能である. そこで, 学習アルゴリズムの基本的な性質は, 探索方向 (search direction) と探索幅 (step-length) の計算法により特徴付けることができるので, 本稿では, 表1に示すように, 探索方向については1次と2次, 探索幅は変動と固定で分類したうえで, その代表的なアルゴリズムを用いて比較を行う. ただし, 探索幅一定のBPの収束効率は悪く, また, Newton法は大規模問題で実用的でないことが一般に知られているので, 以下では, 探索幅を変動させる1次と2次の学習アルゴリズムを用いた比較のみについて考える.

2.3 2次学習アルゴリズム

2次学習アルゴリズムでは, 大規模問題への適用性と非精密探索幅での効率が課題となり, 表2に示すように, 代表的枠組みには, それぞれの長所と短所が指摘されている. 本稿では, これらの課題解決に向けて開発した準ニュートン法³⁾に基づく2次学習アルゴリズムBPQ¹²⁾を用いる. すなわち, BPQの特徴は, 探索方向 $\Delta\Phi$ を小記憶BFGS法で計算し, 探索幅 λ を目的関数の探索幅に対する2次近似の最小点として効率良く計算することである.

式(1)に対して, 小記憶BFGS法は直接適用可能

表2 2次学習法の代表的枠組み
Table 2 Second-order learning algorithms.

2次学習法	大規模問題への適用性	非精密探索幅での効率
ガウス-ニュートン法	×	○
準ニュートン法	△	△
共役勾配法	○	×

であり、また、探索幅 λ は次式で計算できる。

$$\lambda = \frac{-\nabla F_i(\Phi)\Delta\Phi^T}{\Delta\Phi^T\nabla^2 F_i(\Phi)\Delta\Phi} = \frac{-\nabla F_i(\Phi)\Delta\Phi^T}{\Delta\Phi^T\nabla^2 f(\Phi)\Delta\Phi + \mu\Delta\Phi^T\nabla^2\Omega_i(\Phi)\Delta\Phi}$$

学習誤差項に対する2次形式 $\Delta\Phi^T\nabla^2 f(\Phi)\Delta\Phi$ の値は、BPQの方法を用いて、計算量 $Nm + O(hm)$ で効率良く計算できる。一方、ペナルティ項に対する2次形式の計算は以下となる。

$$\begin{aligned} \Delta\Phi^T\nabla^2\Omega_1(\Phi)\Delta\Phi &= \sum_{i=1}^N \Delta\phi_i^2, \\ \Delta\Phi^T\nabla^2\Omega_2(\Phi)\Delta\Phi &= 0, \\ \Delta\Phi^T\nabla^2\Omega_3(\Phi)\Delta\Phi &= \sum_{i=1}^N \frac{(1-3\phi_i^2)\Delta\phi_i^2}{(1+\phi_i^2)^3}. \end{aligned}$$

探索幅計算では、2次形式 $\Delta\Phi^T\nabla^2 F_i(\Phi)\Delta\Phi$ の値が、基本的に、正であることが仮定される。これに対して、3種のペナルティ項は異なる影響を及ぼす。すなわち、自乗値項では、つねに正の値が加えられ、絶対値項では、影響がなく、一方、正規値項については、多くの結合重みが $\sqrt{1/3}$ より大きければ、負の値が加えられる可能性がある。よって、自乗値項には、望ましい性質のあることが分かる。なお、2次学習アルゴリズムとして SCG¹⁰⁾ や OSS¹⁾ などを採用することもできるが、これら既存法と比較して、BPQは最も効率良く誤差を減少できたことが報告されている¹²⁾。

3. 回帰問題による評価

3.1 非線型回帰問題

ペナルティ項を付加したニューラルネット学習の汎化性能と収束性能を評価するため、関数

$$y = (1 - x + 2x^2)e^{-0.5x^2}$$

を学習する非線型回帰問題を考える。実験では、入力変数 x には $[-4, 4]$ の範囲でランダムな値を与え、対応する y の値を x より計算した。ただし、各 y の値には、平均 0、標準偏差 0.2 の正規分布に基づく独立なノイズを加えた。また、訓練事例数を 30 とし、中間ユニット数を 5 に設定した。ただし、各結合重みの

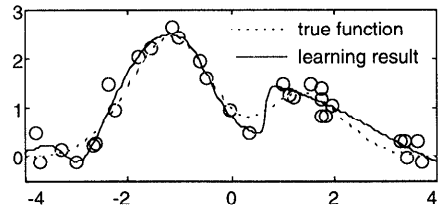


図1 非線型回帰問題
Fig. 1 nonlinear regression problem.

初期値は、平均 0、標準偏差 1 の正規分布に基づいて独立に生成し、勾配ベクトルの大きさが十分小さいとき、すなわち、

$$\|\nabla F_i(\Phi)\|^2/N < 10^{-12}$$

または、合計 CPU 処理時間が 100 秒を超えたときに、学習を終了させた。ペナルティ係数 μ は 2^0 から 2^{-19} まで 2^{-1} を乗じて徐々に変化させ、各ペナルティ項に対して、それぞれ 20 回の試行を行った。汎化能力評価には、各試行の平均 RMSE (root mean squared error) 値を用い、各試行では訓練事例とは独立な 5,000 個のテスト事例で RMSE 値を推定した。なお、テスト事例に対しても訓練事例と同じ量のノイズを加えたので、RMSE 値の下限は 0.2 である。

図1に、訓練事例、真の関数、および、ペナルティ項なし学習で得た関数を示す。ペナルティ項なし学習で得た関数は訓練事例にオーバーフィットしていることが分かる。

3.2 2次学習アルゴリズムの結果

図2に、RMSE 値の平均と標準偏差による汎化性能比較を示す^{*}。どのペナルティ項でも、適切なペナルティ係数を用いれば汎化性能が向上したが、正規値項 Ω_3 では、しばしば望ましくない局所最適解に収束し、最も不安定な挙動を示した。正規値項 Ω_3 が不安定であることは、その標準偏差からも分かる。

図3に、収束までの処理時間の比較を示す^{**}。ただし、標準偏差のグラフにおいて、その値が 0 の箇所は図示していない。ペナルティ項なしの学習結果と比較すれば、自乗値項 Ω_1 では、特に μ が大きいとき、処理時間を大幅に短縮できたが、絶対値項 Ω_2 では、 μ が大きいときに収束した試行がなく、正規値項 Ω_3 では、一般に多くの処理時間を必要とした。つまり、自乗値項 Ω_1 だけが、適切なペナルティ係数で高い汎化能力を示しながら、同時に 2 ~ 100 倍ほど収束性能を改善した。

^{*} 本稿の汎化性能評価は、処理時間の上限で収束しなかった試行結果も含めている。

^{**} 本節の実験には SUN S-4/20 を用いた。

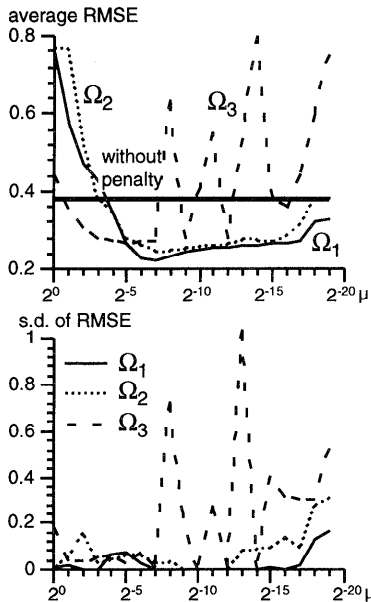


図2 2次学習法BPQを用いた汎化性能比較
Fig. 2 Comparisons of generalization performance using BPQ.

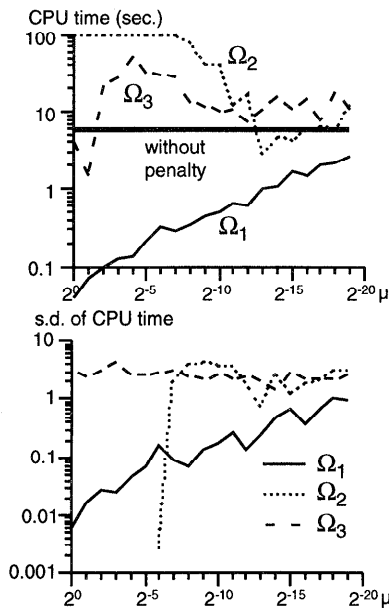


図3 2次学習法BPQを用いた収束性能比較
Fig. 3 Comparisons of convergence performance using BPQ.

図4に、自乗値ペナルティ項の異なるペナルティ係数値における学習結果の例を示す。図から、ペナルティ係数が大きすぎればアンダフィットとなり、逆に小さすぎればオーバーフィットとなることが分かる。また、適切なペナルティ係数値を用いれば、真の関数を

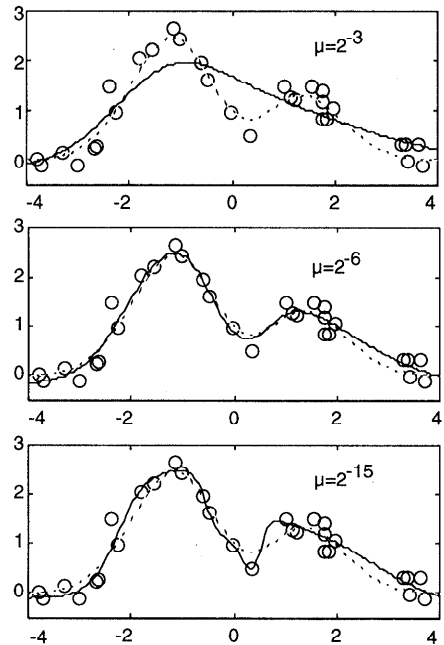


図4 自乗値ペナルティ項の学習結果
Fig. 4 Learning results of squared penalty term.

かなり精度良く近似できることも分かる。

3.3 1次学習アルゴリズムの結果

1次学習アルゴリズムBPを用いて同様の評価を行った。ただし、BPにはSilva-Almeidaの学習定数適応規則¹⁴⁾を適用した。なお、適応規則では、各結合重み ϕ_i に対する学習定数 η_i は

$$\eta_i^p = \begin{cases} \eta_i^{p-1} \times u, & \text{if } \frac{\partial f(\phi^p)}{\partial \phi_i} \times \frac{\partial f(\phi^{p-1})}{\partial \phi_i} \geq 0, \\ \eta_i^{p-1} \times u^{-1}, & \text{otherwise} \end{cases}$$

で調整される。ここで、 p は反復回数を表し、パラメータ u は提案者が推奨するように1.1に設定した¹⁴⁾。ただし、目的関数の値が減少しないときには、すべての学習定数の値はその半分の値に設定される。

図5と図6に、汎化性能と収束までの処理時間の比較を示す。なお、ペナルティ項なしBPの処理時間は図6に示していないが、ほとんどの試行が100秒以内では収束しなかった。また、ここでも標準偏差のグラフにおいて、その値が0の箇所は図示していない。BPを用いた場合、どのペナルティ項でも、適切なペナルティ係数を用いれば汎化性能が向上した。BPQの学習結果と比較すれば、自乗値項 Ω_1 と絶対値項 Ω_2 では、汎化性能はほぼ同じ曲線となったが、BPを用いた場合、自乗値項 Ω_1 の半数以上の試行と絶対値項 Ω_2 のすべての試行が100秒以内では収束しなかった。正規値項 Ω_3 では、BPの汎化性能は安

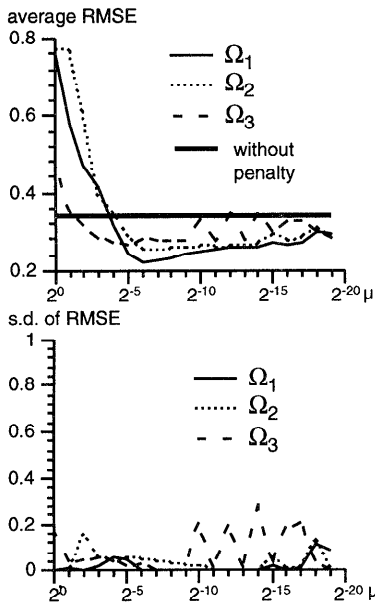


図5 1次学習法BPを用いた汎化性能比較
Fig. 5 Comparisons of generalization performance using BP.

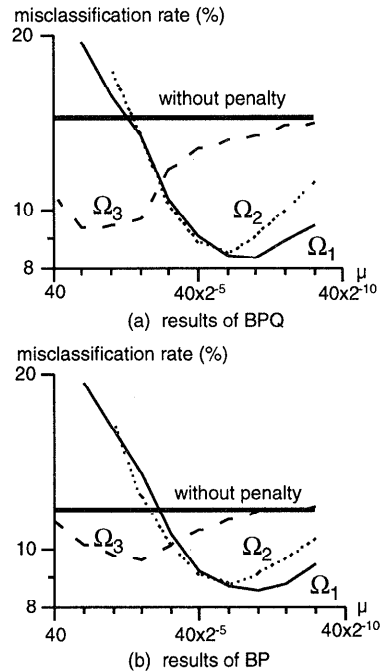


図7 分類問題における汎化性能比較
Fig. 7 Comparisons of generalization performance on classification problem.

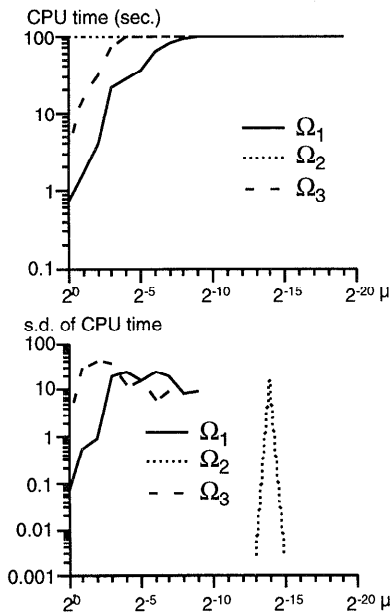


図6 1次学習法BPを用いた収束性能比較
Fig. 6 Comparisons of convergence performance using BP.

定した結果となったが、その多くの試行は100秒以内では収束しなかった。

回帰問題を用いた評価の結論として、実験を行った組合せでは、汎化性能と収束性能の2つを考慮すれば、BPQと自乗値項 Ω_1 の組合せが最も高い性能を示し

た。なお、ペナルティ項なしBPQと比較して、ペナルティ項なしBPは若干高い汎化性能を示した。これはBPの学習時間に上限を設定したことによる early stopping²⁾が働いたためと考えられる。実際、訓練事例に対して、ペナルティ項なしBPQとBPの平均RMSE値は、それぞれ0.133と0.138であった。

4. 分類問題による評価

縮退したGlucksmanの特徴^{4),7)} ($n = 16$ 次元)を用いた手書き数字0~9 ($k = 10$ クラス)の分類問題による評価を行った。実験では、各クラス400ずつの4,000事例の中から、各クラス200ずつの2,000事例を訓練に用い、残りをテスト事例として分類誤り率で汎化能力を評価した。また、中間ユニット数を20、学習におけるCPU処理時間の上限を200秒に設定したこと、および、ペナルティ係数 μ を40から 40×2^{-9} まで 2^{-1} を乗じて変化させたことを除いて、実験設定は3章と同じものを採用した。

図7(a)と(b)に、3種のそれぞれのペナルティ項に対する、1次と2次の学習アルゴリズムの結果を示す。なお、ペナルティ係数値が40と20のときを除いて、すべての組合せの結果において、汎化誤り率の分散は1%未満であった。図より、この実験においても、どのペナルティ項を用いても、適切なペナルティ

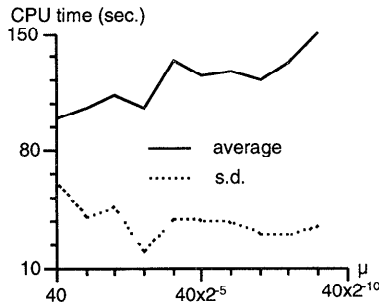


図8 自乗値項を用いたBPQの収束性能
Fig. 8 Convergence performance of BPQ with Ω_1 .

係数で汎化性能が向上したことが分かる。

収束性能評価については、2次学習アルゴリズムBPQと自乗値項 Ω_1 の組合せを除いた他の組合せでは、勾配ベクトルの値を十分に小さくできず、学習処理時間の上限でアルゴリズムが終了した。図8に、自乗値項を用いたBPQの収束性能を示す^{*}。よって、この分類問題を用いた実験でも、汎化性能と収束性能の2つを考慮すれば、BPQと自乗値項 Ω_1 の組合せが最も高い性能を示した。

5. 誤差曲面評価

ペナルティ項の種類により学習性能がなぜ異なるかを誤差曲面を用いて考察するため、関数

$$y = \sigma(w_1x) + \sigma(w_2x)$$

において、2種の変数結合重み w_1 と w_2 だけを学習する非線型回帰問題について考える(図9)。すなわち、この3層ネットワークでは、入力層と出力層にそれぞれ1つのユニットがあり、一方、中間層には2つのユニットがある。ただし、中間ユニットと出力ユニット間の結合重みは1に固定され、各ユニットでのバイアス重みは考えず、中間ユニットの活性化関数はシグモイド $\sigma(x) = 1/(1 + \exp(-x))$ とする。実験では、 $(w_1, w_2) = (1, 3)$ に設定し、入力値 $x_i \in \{-0.2, -0.1, 0, 0.1, 0.2\}$ から各目標値 y_i をそれぞれ計算して訓練事例を生成した。すなわち、最適結合重みは、上記の値か $(w_1, w_2) = (3, 1)$ で与えられる。

図10に、 w_1 と w_2 に対する各ペナルティ項の誤差曲面、および、 $(w_1, w_2) = (-1, -3)$ を初期値として100反復させた学習軌跡を示す。ただし、ペナルティ係数 μ は0.1または0.01に設定し、学習アルゴリズムにはBPQを用いた。自乗値項 Ω_1 の誤差曲面

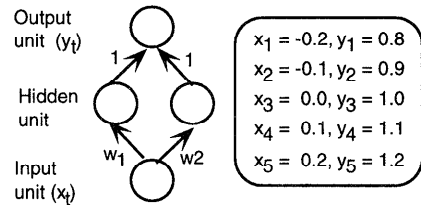


図9 2-結合重み学習問題
Fig. 9 2-weight learning problem.

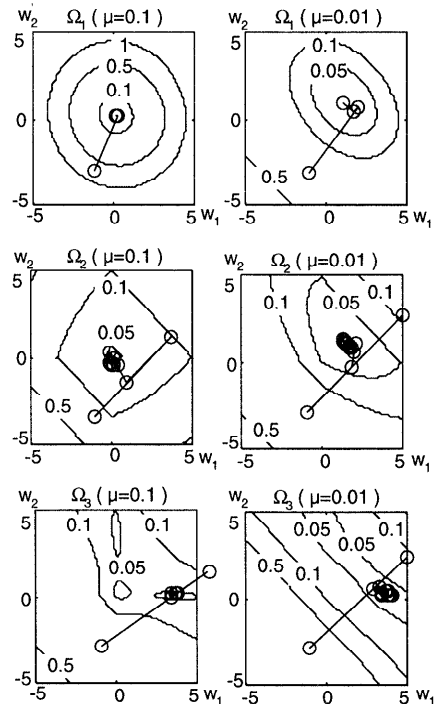


図10 誤差曲面の視覚的評価
Fig. 10 Graphical evaluation of error surfaces.

は楕円を形成し、BPQの学習を容易にしている。一方、 $\mu = 0.1$ のとき、絶対値項 Ω_2 の誤差曲面はほとんど四角い形状となり、原点 ($w_1 = w_2 = 0$) において1次微分関数が不連続となるため、学習軌跡は原点付近で振動した。また、正規値項 Ω_3 の誤差曲面は谷を形成し、BPQの学習を困難にしている。

ここでは、学習結果を可視化するため結合重み数を2としたが、上述の各ペナルティ項の性質は、いずれも多次元そのまま拡張できるものであり、結合重み数が2より大きい場合でも、同様な性質があると考えられる。また、ここでの考察は、3章や4章の実験結果において、2次学習アルゴリズムBPQと自乗値項 Ω_1 の組合せが最も高い収束性能であったことと符合する。

^{*} 本章の実験にはHP C180を用いた。

6. 最適ペナルティ係数の決定法

一般に、与えられた問題に対して、あらかじめ最適なペナルティ係数を知ることはできない。ここでは、ペナルティ係数を変えて学習したニューラルネットを適切に評価するアプローチを考える。汎化性能評価には、交差検証 (cross-validation) 法¹⁵⁾を用いる。以下では、回帰問題について考え、最も効率良く学習できた2次学習アルゴリズムBPQと自乗値項 Ω_1 の組合せを学習アルゴリズムとする。

検証には任意の1つの訓練事例を用いる leave-one-out 法を採用した。すなわち、 h 番目の事例を除いた誤差項を

$$f^{(h)}(\Phi) = \frac{1}{2} \sum_{t=1, t \neq h}^m (y_t - z(x_t; \Phi))^2$$

で定義し、自乗値項を付加した目的関数を最小にするパラメータを

$$\hat{\Phi}^{(h)} = \arg \min_{\Phi} (f^{(h)}(\Phi) + \mu \Omega_1(\Phi))$$

とすれば、交差検証誤差は

$$RMSE^{(CV)}(\mu) = \sqrt{\frac{1}{m} \sum_{h=1}^m (y_h - z(x_h; \hat{\Phi}^{(h)}))^2}$$

である。ペナルティ係数を変えながら学習を進め、交差検証誤差を最小にするものを最適ペナルティ係数とすればよい。すなわち、

$$\mu_{opt} = \arg \min_{\mu} RMSE^{(CV)}(\mu)$$

である。なお、各検証評価の結合重み $\hat{\Phi}^{(h)}$ を求めるには、すべての訓練事例を用いた学習結果の結合重み

$$\hat{\Phi} = \arg \min_{\Phi} (f(\Phi) + \mu \Omega_1(\Phi))$$

を初期値として与えた。この初期値設定法の利点としては、一般に $\hat{\Phi}$ と $\hat{\Phi}^{(h)}$ があまり大きく離れていないと考えられるので、比較的軽い計算量で $\hat{\Phi}^{(h)}$ が求まるからである。また、ニューラルネットの学習には一般に複数の局所最適解が存在するので、 $\hat{\Phi}$ の近傍で $\hat{\Phi}^{(h)}$ を求めることにより、交差検証誤差の評価が安定すると考えられる。

以下、3章と同じ非線型回帰問題を用いて上記の方法の有効性を評価する。ただし、実験条件も3章と同じ設定にした。図11に、交差検証誤差と汎化誤差の平均と標準偏差による比較を示す。交差検証誤差は実際の汎化誤差より悪い評価を与えるが、両者はほぼ同じペナルティ係数で最小になったことが分かる。また、

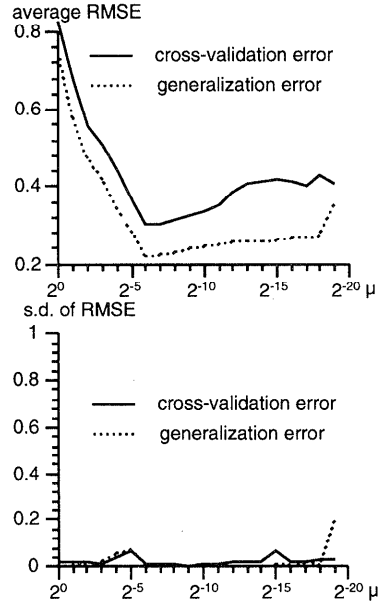


図11 交差検証法を用いた汎化性能
Fig. 11 Generalization performance using cross-validation.

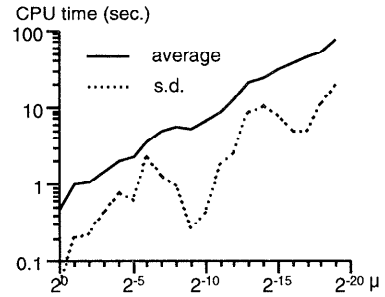


図12 交差検証法を用いた収束性能
Fig. 12 Convergence performance using cross-validation.

学習結果の標準偏差は非常に小さく、安定した学習であったことが分かる。図12に、平均処理時間とその標準偏差を示す。この処理時間には、交差検証評価が含まれるが、かなり効率の良い学習を行えたことが分かる。

7. おわりに

本稿では、3層ニューラルネットの教師あり学習目的関数に3種のペナルティ項を付加したときのそれぞれの学習性能を、1次と2次の学習アルゴリズムBPとBPQを用いて比較評価した。回帰問題と分類問題を用いた実験では、自乗値ペナルティ項と2次学習アルゴリズムの組合せは、適切なペナルティ係数値において、最も高い汎化性能を示すとともに、他の組合せ

と比較して最も高速に収束することを示した。なお、2次学習アルゴリズムとしてSCGやOSSなどを用いても、収束効率が異なるだけなので、同様な結果が得られると考える。今後は、提案したアルゴリズムをさらに幅広い問題に適用して評価する予定である。

謝辞 手書き文字データを提供していただいたNTT基礎研究所石井健一郎氏に感謝いたします。

参考文献

- 1) Battiti, R.: First- and second-order methods for learning between steepest descent and Newton's method, *Neural Computation*, Vol.4, No.2, pp.141-166 (1992).
- 2) Bishop, C.: *Neural networks for pattern recognition*, Oxford Press (1995).
- 3) Gill, P., Murray, W. and Wright, M.: *Practical optimization*, Academic Press, London (1981).
- 4) Glucksman, H.: Classification of mixed-font alphabets by characteristic loci, *IEEE Computer Conference*, pp.138-141 (1967).
- 5) Hanson, S. and Pratt, L.: Comparing biases for minimal network construction with backpropagation, *Advances in Neural Information Processing Systems 1*, Touretzky, D. (Ed.), pp.177-185, Morgan Kaufmann, San Mateo, CA (1989).
- 6) Hinton, G.: Learning translation invariant recognition in massively parallel networks, *Proc. PARLE Conference on Parallel Architectures and Languages Europe*, deBakker, J., Nijman, A. and Treleaven, P. (Eds.), pp.1-13, Berlin (1987).
- 7) 石井健一郎: 変形文字を用いた識別辞書の設計, 信学論, Vol.J72-D, No.5, pp.669-677 (1989).
- 8) Ishikawa, M.: A structural learning algorithm with forgetting of link weight, Technical Report, TR-90-7, Electrotechnical Lab., Tsukuba (1990).
- 9) MacKay, D.: Bayesian interpolation, *Neural Computation*, Vol.4, No.3, pp.415-447 (1992).
- 10) Møller, M.: A scaled conjugate gradient algorithm for fast supervised learning, *Neural Networks*, Vol.6, No.4, pp.525-533 (1993).
- 11) Poggio, T. and Girosi, F.: Regularization algorithms for learning that are equivalent to multilayer networks, *Science*, Vol.247, pp.978-982 (1990).
- 12) Saito, K. and Nakano, R.: Partial BFGS update and efficient step-length calculation for three-layer neural networks, *Neural Computation*, Vol.9, No.1, pp.123-141 (1997).
- 13) Saito, K. and Nakano, R.: Second-order Learning Algorithm with Squared Penalty Term, *Advances in Neural Information Processing Systems 9*, pp.627-633, MIT Press (1997).
- 14) Silva, F. and Almeida, L.: Speeding up backpropagation, *Advanced Neural Computers*, Eckmiller, R. (Ed.), pp.151-160, North-Holland, Amsterdam (1990).
- 15) Stone, M.: Cross-validation: A review, *Operationalforsch. Statist. Ser. Statistics B*, Vol.9, No.1, pp.111-147 (1978).
- 16) Williams, P.: Bayesian regularization and pruning using a Laplace prior, *Neural Computation*, Vol.7, No.1, pp.117-143 (1995).

(平成9年3月3日受付)

(平成9年9月10日採録)

斉藤 和巳 (正会員)



昭和38年生。昭和60年慶応義塾大学理工学部数理科学科卒業。同年NTT入社。平成3年より1年間オタワ大学客員研究員。神経回路網、機械学習の研究に従事。現在、NTT研究開発推進部担当課長。情報処理学会論文賞受賞(平成9年)。電子情報通信学会、人工知能学会、日本神経回路学会、日本認知科学会各会員。

中野 良平 (正会員)



昭和22年生。昭和46年東京大学工学部計数工学科卒業。工学博士。同年、日本電信電話公社(現NTT)入社。以来、統計解析、データベース、人工知能、神経回路網の研究に従事。現在、NTTコミュニケーション科学研究所主幹研究員。電気通信普及財団賞(テレコムシステム技術賞)受賞(平成9年)。情報処理学会論文賞受賞(平成9年)。人工知能学会、日本神経回路学会、日本応用数理学会各会員。