Erratum

Durch ein Versehen sind im Kapitel 6.1 („Schluß“) in der „Bewertung der Ergebnisse“ einige Absätze verlorengegangen. Die letzten zwei Absätze des Kapitels 6.1 auf Seite 180 sind durch den folgenden Text zu ersetzen:

Unter den Lernmodi erzielte das “Sum-of-Derivatives” Batch Learning (Batch1) trotz der (durch die Summation) möglichen gegenseitigen Auslöschung der partiellen Ableitungen teilweise ausgezeichnete Ergebnisse (speziell in Verbindung mit dem Standardverfahren), vor allem unter Berücksichtigung der Tatsache, daß dieser Modus nur soviele Update-Schritte benötigt wie Epochen. Bei größeren Problemen (wie dem AAT-Problem) war das Konvergenzverhalten jedoch sehr schlecht.

Das “Sum-of-Updates” Batch Learning (Batch2) erwies sich nur bei zwei Versuchen (XOR und Rotkäppchen) in Verbindung mit dem Verfahren von Schmidhuber bzw. meinem eigenen als dem Batch1 Learning überlegen, konvergierte bei den übrigen Versuchen jedoch nicht.

Das Periodical Learning schnitt im Vergleich mit dem Batch1 Learning teils besser, teils schlechter (unter Berücksichtigung der benötigten Anzahl von Updates) ab. Als On-Line-Lernmodus stellt es eine brauchbare Alternative zum Batch1 Learning dar, obwohl es eigentlich eine Abweichung vom „korrekten“ Gradientenabstieg (beim Standardverfahren) darstellt.

Das Sequential Learning hat den Nachteil, daß das Netzwerk zuvor gelernte Patterns wieder „vergißt“, während die übrigen Patterns trainiert werden, es benötigt daher eine (im Vergleich zu den anderen Verfahren) unverhältnismäßig hohe Anzahl von Update-Schritten. Erstaunlicherweise erwies es sich jedoch in Verbindung mit dem Verfahren von Schmidhuber sowie meinem eigenen in vielen Fällen als den anderen Lernmodi bei weitem überlegen.

Das von mir entwickelte Dynamic Learning hat sich in den meisten Versuchen hervorragend bewährt. Auch in Kombination mit dem Standardverfahren erwies es sich als besonders vorteilhaft (siehe z.B. im Versuch Sin5).

Besonders jedoch in Kombination mit dem Verfahren von Schmidhuber sowie meinem eigenen Verfahren erwies es sich als sehr effektiv.

Obwohl dadurch die Lernzeiten teilweise sehr schwankten, erwies sich die zufällige Auswahl der Patterns als besonders hilfreich bei der Überwindung lokaler Minima und führte zu einer generellen Verkürzung der Lernzeiten.

Die Beobachtung anderer Autoren, daß die Verwendung der symmetrischen Aktivierungsfunktion die Lernzeiten verkürzt, konnte in vielen, jedoch nicht in allen Fällen bestätigt werden.