

This article was downloaded by: [New York University]

On: 17 October 2014, At: 22:24

Publisher: Taylor & Francis

Informa Ltd Registered in England and Wales Registered Number: 1072954 Registered office: Mortimer House, 37-41 Mortimer Street, London W1T 3JH, UK



## Connection Science

Publication details, including instructions for authors and subscription information:

<http://www.tandfonline.com/loi/ccos20>

### Simplifying Neural Network Soft Weight-sharing Measures by Soft Weight-measure Soft Weight Sharing

BARAK PEARLMUTTER<sup>a</sup>

<sup>a</sup> Siemens Corporate Research, Inc. , 755 College Road East, Princeton, NJ, 08540, USA E-mail:

Published online: 29 Oct 2007.

To cite this article: BARAK PEARLMUTTER (1994) Simplifying Neural Network Soft Weight-sharing Measures by Soft Weight-measure Soft Weight Sharing, Connection Science, 6:1, 105-105, DOI: [10.1080/09540099408915712](https://doi.org/10.1080/09540099408915712)

To link to this article: <http://dx.doi.org/10.1080/09540099408915712>

PLEASE SCROLL DOWN FOR ARTICLE

Taylor & Francis makes every effort to ensure the accuracy of all the information (the "Content") contained in the publications on our platform. However, Taylor & Francis, our agents, and our licensors make no representations or warranties whatsoever as to the accuracy, completeness, or suitability for any purpose of the Content. Any opinions and views expressed in this publication are the opinions and views of the authors, and are not the views of or endorsed by Taylor & Francis. The accuracy of the Content should not be relied upon and should be independently verified with primary sources of information. Taylor and Francis shall not be liable for any losses, actions, claims, proceedings, demands, costs, expenses, damages, and other liabilities whatsoever or howsoever caused arising directly or indirectly in connection with, in relation to or arising out of the use of the Content.

This article may be used for research, teaching, and private study purposes. Any substantial or systematic reproduction, redistribution, reselling, loan, sub-licensing, systematic supply, or distribution in any form to anyone is expressly forbidden. Terms & Conditions of access and use can be found at <http://www.tandfonline.com/page/terms-and-conditions>

## HUMOUR

**Simplifying Neural Network Soft Weight-sharing Measures by Soft Weight-measure Soft Weight Sharing**

---

---

BARAK PEARLMUTTER

It has been shown by Nowlan and Hinton (1991) that it is advantageous to construct weight complexity measures for use in weight regularization through the use of EM, instead of relying on some *a priori* complexity measure, or even worse, neglecting regularization by assuming a uniform distribution. Their work can be regarded as a generalization of the 'optimal brain damage' of Le Cun *et al.* (1990), in which the distribution of weights is estimated with a histogram, a peculiar functional form for a distribution. Nowlan and Hinton assume a much simpler functional form for the distribution, avoiding over-fitting and therefore over-regularization. However, they disregard the issue of regularization of the regularizer itself. Just as certain weights might be considered *a priori* quite unlikely, certain distributions of weights may be considered *a priori* quite unlikely. To solve this problem, we introduce a regularization term on the parameters of the weight distribution being estimated. This regularization term is itself determined by a distribution over these distributional parameters. In this light, Nowlan and Hinton (1991) make the uniform distributional parameter distribution assumption. Here, we estimate the distribution of distributions by running an ensemble of networks, with EM used to estimate the weight distribution of each network (following Nowlan and Hinton), but we then use EM to estimate the distribution of distributions across networks. Of course, each estimated distribution is used to regularize the parameters over which that distribution is defined, leading to regularization of the individual network regularizers.

We do not consider how to estimate the *a priori* distribution which might be used to regularize the distribution being used to regularize the distribution being used to regularize the weights being estimated from the data, which will be explored in a future paper.

**References**

- Le Cun (1990).  
Nowlan & Hinton (1991).