

This document is also available in PDF¹ format.

Purpose: Basic techniques for training MLPs

Material: Papers by Lippman and Hush. Course notes by Jordan.

General: There are a number of algorithms for training ANNs. The most frequently encountered is the so-called *backpropagation* algorithm which minimises the square of the errors at the output nodes of the network. This will be the topic of this lecture. (This algorithm is closely related to the well-known LMS algorithm encountered in adaptive filters.)

Topics:

- Minimising Least Squares Error criterium by means of gradient descent. Matrix equations in article by Jordan.
- Batch versus online training.
- Use of momentum in weight adaptation can prevent oscillations and supply some control over effective step lengths.
- Adaptable step lengths provide for somewhat faster convergence, and alleviate difficult decisions on step lengths. Based on consistency of direction of change for weights. An interesting discussion on momentum and adaptive weights: L.W. Chan and F. Fallside, "An adaptive training algorithm for back propagation networks", Computer Speech and Language vol 2, nr 3/4, Sep/Dec 1987. Conjugate Gradient optimisation is a popular alternative. (Search the net and let me know if you find a good article.)
- Advantages: Very flexible, can map arbitrary functions.
- Some limitations: Horribly slow. Can get stuck in a local minimum. To guarantee a smaller error at each iteration an impractical small step length must be used. The whole process can be seen as estimating the parameters of a function mapping the input to the output. Unpredictable behaviour may occur in regions of the input space not adequately covered during training. Generalisation.

Task: (Hand in 2 weeks from lecture)

Implement a backpropagation ANN. Use it to model and classify some of the non-temporal data used in previous tasks. Although any of the datasets for which you would consider using a GMM should do, computational restrictions probably make the simulated vowels a more attractive option. Compare! What advantages could this technique have over the previously encountered ones?

ALTERNATIVE TASK: Use an RBF network instead of the backprop net. See Haykin, "Neural Networks: A Comprehensive Foundation", (published by McMillan).

¹http://www.dsp.sun.ac.za/pr813/lectures/11_mlp_train/11_mlp_train.pdf