Bayesian Neural Networks: Case Studies in Industrial Applications

Aki Vehtari[1] **and Jouko Lampinen**[1]

[1]Laboratory of Computational Engineering, Helsinki University of Technology, P.O.Box 9400, FIN-02015 HUT, FINLAND

Abstract: We demonstrate the advantages of using Bayesian neural networks in industrial applications. The Bayesian approach provides consistent way to do inference by combining the evidence from data to prior knowledge from the problem. A practical problem with neural networks is to select the correct complexity for the model, i.e., the right number of hidden units or correct regularization parameters. The Bayesian approach offers efficient tools for avoiding overfitting even with very complex models, and facilitates estimation of the confidence intervals of the results. In this contribution we review the Bayesian methods for neural networks and present comparison results from case studies in prediction of the quality properties of concrete (regression problem), electrical impedance tomography (inverse problem) and forest scene analysis (classification problem). The Bayesian neural networks provided consistently better results than other methods.

1 Introduction

In classification and non-linear function approximation neural networks have become very popular in recent years. With neural networks the main difficulty is in controlling the complexity of the model. Another problem of standard neural network models is the lack of tools for analyzing the results (confidence intervals, like 10 % and 90 % quantiles, etc.).

The Bayesian approach provides consistent way to do inference by combining the evidence from data to prior knowledge from the problem. Bayesian methods use probability to quantify uncertainty in inferences and the result of Bayesian learning is a probability distribution expressing our beliefs regarding how likely the different predictions are. Predictions are made by integrating over the posterior distribution. The main advantages of using Bayesian methods are:

- Automatic complexity control: Values of regularization coefficients can be selected using only the training data, without the need to use separate training and validation data.
- Possibility to use prior information and hierarchical models for the hyperparameters.
- Predictive distributions for outputs.

Y. Suzuki et al. (eds.), *Soft Computing in Industrial Applications*

In this contribution we demonstrate the advantages of Bayesian neural networks [1] in three case problems. First we briefly review Bayesian methods for MLP neural networks in section 2. In sections 3, 4 and 5 we present results using Bayesian neural network in prediction of the quality properties of concrete (regression problem), electrical impedance tomography (inverse problem) and forest scene analysis (classification problem).

2 Bayesian Learning for MLP

Bayesian methods can be used for many types of neural networks, but we concentrate here to one hidden layer MLP network with hyperbolic tangent (tanh) activation function. See [2] for thorough introduction to MLP. Basic MLP network model with m hidden units and k outputs is

$$f_k(\mathbf{x}, \mathbf{w}) = w_{k0} + \sum_{j=1}^{m} w_{kj} \tanh\left(w_{j0} + \sum_{i=1}^{d} w_{ji} x_i\right), \tag{1}$$

where $\mathbf{x}$ is a d-dimensional input vector, $\mathbf{w}$ denotes the weights, and indices i and j correspond to hidden and output units, respectively.

Next we review main ideas of Bayesian learning briefly. See, e.g., [3] for good introduction to Bayesian methods.

2.1 Bayesian Learning

Consider a regression or classification problem involving the prediction of a noisy vector $\mathbf{y}$ of target variables given the value of a vector $\mathbf{x}$ of input variables.

The process of Bayesian learning is started by defining a model, $\mathcal{M}$, and prior distribution $p(\theta)$ for the model parameters θ. Prior distribution expresses our initial beliefs about parameter values, before any data has observed. After observing new data $D = \{(\mathbf{x}^{(1)}, \mathbf{y}^{(1)}), \ldots, (\mathbf{x}^{(n)}, \mathbf{y}^{(n)})\}$, prior distribution is updated to the posterior distribution using Bayes' rule

$$p(\theta|D) = \frac{p(D|\theta)p(\theta)}{p(D)} \propto L(\theta|D)p(\theta), \tag{2}$$

where the likelihood function $L(\theta|D)$ gives the probability of the observed data as function of the unknown model parameters.

To predict the new output $\mathbf{y}^{(n+1)}$ for new input $\mathbf{x}^{(n+1)}$, predictive distribution is obtained by integrating the predictions of the model with respect to the posterior distribution of the model parameters

$$p(\mathbf{y}^{(n+1)}|\mathbf{x}^{(n+1)}, D) = \int p(\mathbf{y}^{(n+1)}|\mathbf{x}^{(n+1)}, \theta)p(\theta|D)d\theta. \tag{3}$$