In this paper, we use the evidence framework of MacKay as an example implementation of Bayesian learning. An interesting additional feature of this framework is the automatic relevance determination (ARD) method which allows us to assess the relative importance of various inputs by adding weight regularization terms to the objective function.

**Input ranking using automatic relevance determination (ARD)**

Selecting the best subset of a set of \( n \) input variables as predictors for a neural network is a non-trivial problem. This follows from the fact that the optimal input subset can only be obtained when the input space is exhaustively searched. When \( n \) inputs are present, this would imply the need to evaluate \( 2^n - 1 \) input subsets. Unfortunately, as \( n \) grows this very quickly becomes computationally infeasible. For that reason, heuristic search procedures are often preferred. A multitude of input selection methods have been proposed in the context of neural networks. These methods generally rely on the use of sensitivity heuristics, which try to measure this impact of input changes on the output of the trained network. Inputs may then be ranked (soft input selection) and/or pruned (hard input selection) according to their sensitivity values. In this paper, we focus on input ranking as a means assess the relative importance of the various inputs for the direct marketing case at hand.

(未完)
industry and academia. We compared this benchmark with two state-of-the-art techniques from the machine learning and data mining domain.

First, given the widespread use of decision trees in prediction problems where the user seeks insight into the predictive process, we have implemented Random Forest (RF, Breiman, 2001). This technique focuses on growing an ensemble of decision trees using a random selection of features to split each node (i.e. the random subspace method), where the final prediction is computed as the average output from the individual trees. RF models have been argued to possess excellent properties for feature selection, and to avoid overfitting given that the number of trees is large (Breiman, 2001). In this approach, we will grow 5000 trees, as in other applications.

3.2. Cross-validation
3.3. Variable selection

5. Results

5.1. Survey response
5.2. Predictive performance
5.3. Usefulness of the variable-selection technique
5.4. Variable importance
**NN applications**

App2. (2006) Interval estimation of urban ozone level and selection of influential factors by employing automatic relevance determination model

(香港空氣污染研究應用)

3. Methodology

3.1. Background of ANN and its limitation

3.2. Bayesian neural network (BNN)

MacKay (1992) and Neal (1992) laid the foundation of Bayesian neural network (BNN) by combining the concept of Bayesian evidence framework and ANN. It provides a unified theoretical treatment of learning in ANN and also a potential solution to the problems mentioned in Section 3.1. The advantage of BNN includes:

1. Automatic complexity control. Bayesian inference techniques allow the values of regularization coefficients to be selected using only the training data, without the need of using CV techniques.
2. Predictive distribution or an error bar is obtained for output instead of only a ‘point’ estimation for output in model based on ML conception.
3. A soft feature selection scheme for input variables. The ARD model, an extension of evidence framework, is able to pick out relevance inputs and to correctly rank them in order of relevance with respect to outputs.

3.2.1. The evidence framework

3.2.2. Automatic relevance determination (ARD)

Automatic relevance determination method is an extension of Bayesian evidence framework. For a two-layer MLP, in the standard evidence framework. For a two-layer MLP, in the standard evidence framework, four hyperparameters related to different weight group are needed and introduced normally. These are $\alpha_{w1}, \alpha_{h1}, \alpha_{w2}$ and $\alpha_{b2}$, which are used to control the magnitude of input-layer weights, input-layer biases, hidden-layer weights, and hidden-layer biases respectively. Besides above mentioned, in the ARD framework for MLP, it assigns one parameter $\alpha_{w1}$ for each $n$ input variable (input neuron), which can control the magnitude of weight fanning out from the input neuron (here, $j$ ranges from 1 to $n$). Therefore, in ARD
framework, there are total \( n + 3 \) hyperparameters to control the corresponding weight group.

### 3.2.3. Prediction for evidence framework

#### 5. Conclusions

Three major advantages:

1. Automatic control of model’s complexity in evidence framework and solve over-fitting problem without using technology like cross-validation.
2. By using different error bar for each output, the ARD model was more robust and reliable to capture the wild fluctuation of \( O_3 \) level especially during \( O_3 \) episodes.
3. The ARD model was able to select relevant inputs and rank them in order of relevance with respect to output by comparing the hyperparameters associate with each input variable.

**App4. (2004) Selection of input parameters to model direct solar irradiance by using artificial neural networks**

(太陽能照明預測)

In this work, the Bayesian framework for ANN, named as automatic relevance determination method (ARD), was employed to obtained to select the optimum input parameters to the neural network. For that, a multi-layer feed-forward perception is trained on these data. The results reflect the relative importance of the inputs selected.


(生化應用)

The relative importance of the molecular descriptors for the most predictive BBB model were determined by the use of automatic relevance determination (ARD), and compared with the important descriptors from other literature models of BBB partitioning.

#### 2. Methods

2.1. Blood-brain barrier partition data
2.2. Molecular descriptors

2.3. Bayesian neural nets and automatic relevance determination (ARD)

We reported theory of Bayesian regularized neural networks to QSAR modeling, and a number of applications of this robust SAR mapping method, in several prior publications. Consequently, we will present only a brief account here.

Conventional back-propagation neural net training methods are usually variations of maximum likelihood algorithms that aim to find a single set of network weights that maximize the fit to training data. Bayesian regularization considers a probabilities being varied during training in response to how well a particular set of weights models the data. When predicting net output for new data Bayesian regularized nets provide a **predictive output distribution** for the new cases, whereas conventional neural nets give a **single value**. Thus the Bayesian neural net can provide a measure of uncertainty or quality of the prediction.

**ARD** model uses multiple regularization constants, one associated with each input. By applying Bayes’ theorem, the regularization constants for non-informative inputs are automatically inferred to be large, preventing those inputs from causing overfitting.

2.4. QSAR modeling

3. Results and discussions

3.1. QSAR modeling

3.2. Model interpretation

**App6.** (2003) A Bayesian neural network approach for modelling censored data with an application to prognosis after surgery for breast cancer

4. Automatic relevance determination

4.1. Bayesian regularization framework for ARD

4.2. Setting the regularization parameters

4.3. Model selection with the neural network

4.4. Marginalization of the network predictions
The Bayesian neural network technique was used for the modeling of creep rupture life. To obtain the posterior distribution and prediction which is off a high dimensional integral form, the Markov chain Monte Carlo method was applied. With a complex 13-31-1 architecture, the network shows a precise prediction of creep rupture life of superalloys. Automatic relevance determination reveals that the influences of Re and Cr on creep rupture life of single crystal superalloys are the greatest amongst the alloying elements.

2. Neural network

2.1. Conventional neural network

2.2. Bayesian neural network

2.3. Markov chain Monte Carlo methods


To minimize the influence of noisy low level coefficients, we applied the practical Bayesian method automatic relevance determination (ARD) model to choose the size of MLPs, which are then trained to provide forecasts.

2.2. Automatic relevance determination for NN input window determination

App10. (2005) Flow rate of some pharmaceutical diluents through die-orifices relevant to mini-tableting

2.5. Importance of input variables

Importance of the input variables was determined for the FBP model by the automatic relevance determination, ARD routine, implemented in the Netlab package (Neal, 1996).