Communicated by Richard Lippmann

A Neural Network for Nonlinear Bayesian Estimation in Drug Therapy

Reza Shadmehr
Department of Computer Science,
University of Southern California, Los Angeles, CA 90089 USA

David Z. D'Argenio
Department of Biomedical Engineering,
University of Southern California, Los Angeles, CA 90089 USA

The feasibility of developing a neural network to perform nonlinear Bayesian estimation from sparse data is explored using an example from clinical pharmacology. The problem involves estimating parameters of a dynamic model describing the pharmacokinetics of the bronchodilator theophylline from limited plasma concentration measurements of the drug obtained in a patient. The estimation performance of a backpropagation trained network is compared to that of the maximum likelihood estimator as well as the maximum a posteriori probability estimator. In the example considered, the estimator prediction errors (model parameters and outputs) obtained from the trained neural network were similar to those obtained using the nonlinear Bayesian estimator.

1 Introduction

The performance of the backpropagation learning algorithm in pattern classification problems has been compared to that of the nearest-neighbor classifier by a number of investigators (Gorman and Sejnowski 1988; Burr 1988; Weideman et al. 1989). The general finding has been that the algorithm results in a neural network whose performance is comparable (Burr 1988; Weideman et al. 1989) or better (Gorman and Sejnowski 1988) than the nearest-neighbor technique. Since the probability of correct classification for the nearest-neighbor technique can be used to obtain upper and lower bounds on the Bayes probability of correct classification, the performance of the network trained by Gorman and Sejnowski (1988) is said to have approached that of a Bayes decision rule.

Benchmarking the backpropagation algorithm's performance is necessary in pattern classification problems where class distributions intersect. Yet few investigators (Kohonen et al. 1988) have compared the performance of a backpropagation trained network in a statistical

Neural Computation **2, 216–225** (1990)

pattern recognition or estimation task, to the performance of a Bayesian or other statistical estimators. Since Bayesian estimators require *a priori* knowledge regarding the underlying statistical nature of the classification problem, and simplifying assumptions must be made to apply such estimators in a sparse data environment, a comparison of the neural network and Bayesian techniques would be valuable since neural networks have the advantage of requiring fewer assumptions in representing an unknown system.

In this paper we compare the performance of a backpropagation trained neural network developed to solve a nonlinear estimation problem to the performance of two traditional statistical estimation approaches: maximum likelihood estimation and Bayesian estimation. The particular problem considered arises in the field of clinical pharmacology where it is often necessary to individualize a critically ill patient's drug regimen to produce the desired therapeutic response. One approach to this dosage control problem involves using measurements of the drug's response in the patient to estimate parameters of a dynamic model describing the pharmacokinetics of the drug (i.e., its absorbtion, distribution, and elimination from the body). From this patient-specific model, an individualized therapeutic drug regimen can be calculated. A variety of techniques have been proposed for such feedback control of drug therapy, some of which are applied on a routine basis in many hospitals [see Vozeh and Steimer (1985) for a general discussion of this problem]. In the clinical patient care setting, unfortunately, only a very limited number of noisy measurements are available from which to estimate model parameters. To solve this sparse data, nonlinear estimation problem, both maximum likelihood and Bayesian estimation methods have been employed (e.g., Sheiner et al. 1975, Sawchuk et al. 1977). The a priori information required to implement the latter is generally available from clinical trials involving the drug in target patient populations.

2 The Pharmacotherapeutic Example

The example considered involves the drug theophylline, which is a potent bronchodilator that is often administered as a continuous intravenous infusion in acutely ill patients for treatment of airway obstruction. Since both the therapeutic and toxic effects of theophylline parallel its concentration in the blood, the administration of the drug is generally controlled so as to achieve a specified target plasma drug concentration. In a population study involving critically ill hospitalized patients receiving intravenous theophylline for relief of asthma or chronic bronchitis, Powell et al. (1978) found that the plasma concentration of theophylline, $y(t)$, could be related to its infusion rate, $r(t)$, by a simple one-compartment, two-parameter dynamic model [i.e., $dy(t)/dt = -(CL/V)y(t)+r(t)/V$]. In the patients studied (nonsmokers with no other

organ disfunction), significant variability was observed in the two kinetic model parameters: distribution volume V (liters/kg body weight) = 0.50 $\pm$ 0.16 (mean $\pm$ SD); elimination clearance CL (liters/kg/hr) = 0.0386 $\pm$ 0.0187. In what follows, it will be assumed that the population distribution of V and CL can be described by a bivariate log-normal density with the above moments and a correlation between parameters of 0.5. For notational convenience, α will be used to denote the vector of model parameters ($\alpha = [V\ CL]^T$) and μ and Ω used to represent the prior mean parameter vector and covariance matrix, respectively.

Given this a priori population information, a typical initial infusion regimen would consist of a constant loading infusion, r_1, equal to 10.0 mg/kg/hr for 0.5 hr, followed by a maintenance infusion, r_2, of 0.39 mg/kg/hr. This dosage regimen is designed to produce plasma concentrations of approximately 10 μg/ml for the patient representing the population mean (such a blood level is generally effective yet nontoxic). Because of the significant intersubject variability in the pharmacokinetics of theophylline, however, it is often necessary to adjust the maintenance infusion based on plasma concentration measurements obtained from the patient to achieve the selected target concentration. Toward this end, plasma concentration measurements are obtained at several times during the initial dosage regimen to estimate the patient's drug clearance and volume. We assume that the plasma measurements, $z(t)$, can be related to the dynamic model's prediction of plasma concentration, $y(t, \alpha)$, as follows: $z(t) = y(t, \alpha) + e(t)$. The measurement error, $e(t)$, is assumed to be an independent, Gaussian random variable with mean zero and standard deviation of $\sigma_t(\alpha) = 0.15 \times y(t, \alpha)$. A typical clinical scenario might involve only two measurements, $z(t_1)$ and $z(t_2)$, where t_1 = 1.5 hr and t_2 = 10.0 hr. The problem then involves estimating V and CL using the measurements made in the patient, the kinetic model, knowledge of the measurement error, as well as the prior distribution of model parameters.

3 Estimation Procedures

Two traditional statistical approaches have been used to solve this sparse data system estimation problem: maximum likelihood (ML) estimation and a Bayesian procedure that calculates the maximum a posteriori probability (MAP). Given the estimation problem defined above, the ML estimate, α^{ML}, of the model parameters, α, is defined as follows:

$$\alpha^{ML} = \arg\left\{\min_{\alpha \in R^2} O_{ML}(\alpha)\right\} \tag{3.1}$$

$$O_{ML}(\alpha) = ln\ \det\Sigma(\alpha) + [z - y(\alpha)]^T \Sigma(\alpha)^{-1} [z - y(\alpha)] \tag{3.2}$$

where $z = [z(t_1)\, z(t_2)]^T, y(\alpha) = [y(t_1, \alpha)\, y(t_2, \alpha)]^T$, and $\Sigma(\alpha) = \text{diag}\{\sigma_{t_1}(\alpha)\, \sigma_{t_2}(\alpha)\}$. The MAP estimator is defined as follows:

$$\alpha^{MAP} = \arg\left\{\min_{\alpha \in R^2} O_{MAP}(\alpha)\right\} \tag{3.3}$$

$$O_{MAP} = ln\ \det\Sigma(\alpha) + [z - y(\alpha)]^T \Sigma(\alpha)^{-1} [z - y(\alpha)] + [ln\ \alpha - \nu]^T \Phi^{-1} [ln\ \alpha - \nu] + 2tr A(\alpha) \tag{3.4}$$

where $\nu = \{\nu_i\}, i = 1, 2, \Phi = \{\phi_{ij}\}, i = j = 1, 2$, with $\nu_i = ln\ \mu_i - \phi_{ii}/2, i = 1, 2$, and $\phi_{ij} = ln(\omega_{ij}/\mu_{ij}\mu_{ij} + 1), i, j = 1, 2$. The mean and covariance of the prior parameter distribution, μ and Ω (see above), define the quantities μ_i and ω_{ij}. Also, $A(\alpha) = \text{diag}\{ln\alpha_1\ ln\alpha_2\}$. The corresponding estimates of the drug's concentration in the plasma can also be obtained using the above parameter estimates together with the kinetic model. To obtain the ML and MAP estimates a general purpose pharmacokinetic modeling and data analysis software package was employed, which uses the Nelder–Mead simplex algorithm to perform the required minimizations and a robust stiff/nonstiff differential equation solver to obtain the output of the kinetic model (D'Argenio and Schumitzky 1988).

As an alternate approach, a feedforward, three-layer neural network was designed and trained to function as a nonlinear estimator. The architecture of this network consisted of two input units, seven hidden units, and four output units. The number of hidden units was arrived at empirically. The inputs to this network were the patient's noisy plasma samples $z(t_1)$ and $z(t_2)$, and the outputs were the network's estimates for the patient's distribution volume and elimination clearance (α^{NN}) as well as for the theophylline plasma concentration at the two observation times $[y(t_1),\ y(t_2)]$.

To determine the weights of the network, a *training set* was simulated using the kinetic model defined above. Model parameters (1000 pairs) were randomly selected according to the log–normal prior distribution defining the population $(\alpha_i, i = 1, \ldots, 1000)$, and the resulting model outputs determined at the two observation times $[y(t_1, \alpha_i), y(t_2, \alpha_i), i = 1, \ldots, 1000]$. Noisy plasma concentration measurements were then simulated $[z(t_1)_i, z(t_2)_i, i = 1, \ldots, 1000]$ according to the output error model defined previously. From this set of inputs and outputs, the backpropagation algorithm (Rumelhart et al. 1986) was used to train the network as follows. A set of 50 vectors was selected from the full *training set*, which included the vectors containing the five smallest and five largest values of V and CL. After the vectors had been learned, the performance of the network was evaluated on the full *training set*. Next, 20 more vectors were added to the original 50 vectors and the network was retrained. This procedure was repeated until addition of 20 new training vectors did not produce appreciable improvement in the ability of the network to estimate parameters in the full *training set*. The final network was the result of training on a set of 170 vectors, each vector being presented

to the network approximately 32,000 times. As trained, the network approximates the minimum expected (over the space of parameters and observations) mean squared error estimate for $\alpha, y(t_1)$ and $y(t_2)$. [See Asoh and Otsu (1989) for discussion of the relation between nonlinear data analysis problems and neural networks.]

4 Results

The performance of the three estimators (ML, MAP, NN) was evaluated using a *test set* (1000 elements) simulated in the same manner as the training set. Figures 1 and 2 show plots of the estimates of V and CL, respectively, versus their true values from the test set data, using each of the three estimators. Also shown in each graph are the lines of regression (solid line) and identity (dashed line).

To better quantify the performance of each estimator, the mean and root mean squared prediction error (Mpe and $RMSpe$, respectively) were determined for each of the two parameters and each of the two plasma concentrations. For example, the prediction error (percent) for the NN volume estimate was calculated as $pe_i = (V_i^{NN} - V_i)100/V_i$, where V_i is the true value of volume for the ith sample from the test set and V_i^{NN} is the corresponding NN estimate.

Table 1 summarizes the resulting values of the Mpe for each of the three estimators. From inspection of Table 1 we conclude that the biases associated with each estimator, as measured by the Mpe for each quantity, are relatively small, and comparable. As a single measure of both the bias and variability of the estimators, the $RMSpe$ given in Table 2 indicate that, with respect to the parameters V and CL, the precision of the NN and MAP estimators is similar and significantly better than that of the ML estimator in the example considered here.

For both the nonlinear maximum likelihood and Bayesian estimators, an asymptotic error analysis could be employed to provide approximate errors for given parameter estimates. In an effort to supply some type of

	Mpe (%)			
Estimator	V	CL	$y(t_1)$	$y(t_2)$
ML	2.5	3.4	-1.1	-3.0
MAP	1.0	6.1	0.8	1.5
NN	4.7	3.8	0.6	7.3

Table 1: Mean Prediction Errors (Mpe) for the Parameters (V and CL) and Plasma Concentrations [$y(t_1)$ and $y(t_2)$] as Calculated, for Each of the Three Estimators, from the Simulated Test Set.

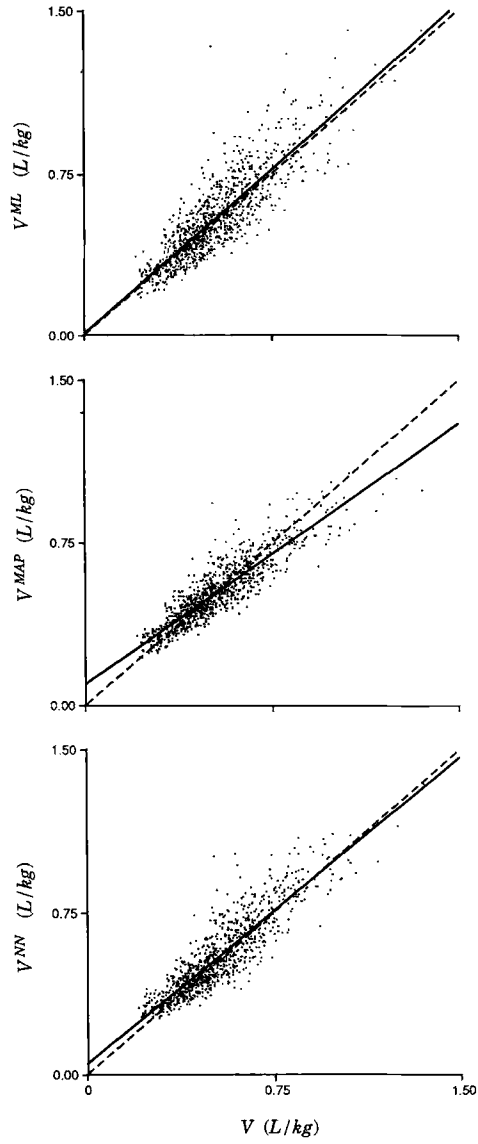

Figure 1: Estimates of V for the ML, MAP, and NN procedures (top to bottom), plotted versus the true value of V for each of the 1000 elements of the test set. The corresponding regression lines are as follows: $V^{ML} = 1.0V + 0.004$, $r^2 = 0.74$; $V^{MAP} = 0.80V + 0.094$, $r^2 = 0.81$; $V^{NN} = 0.95V + 0.044$, $r^2 = 0.80$.

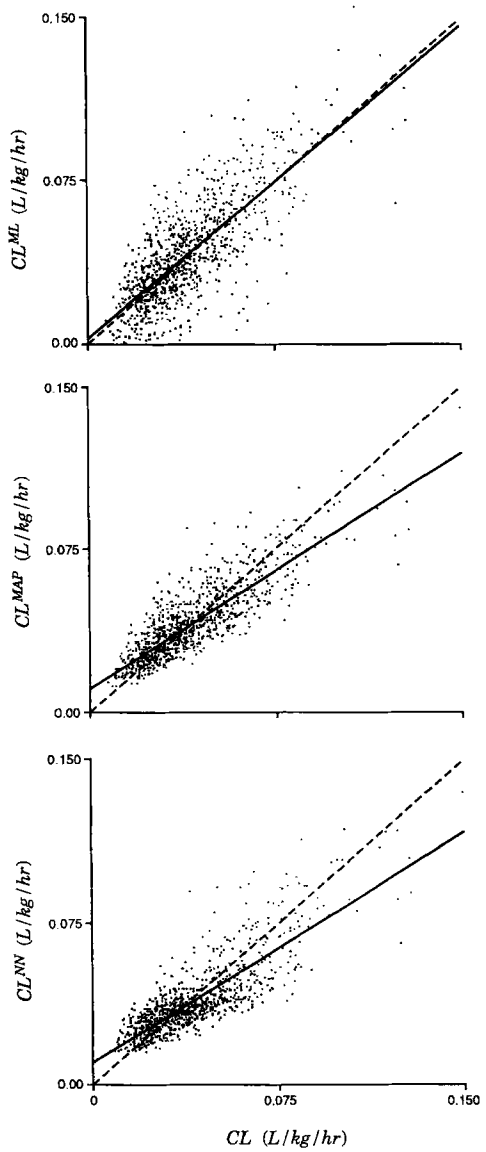

Figure 2: Estimates of CL for the ML, MAP, and NN procedures (top to bottom), versus their true values as obtained from the test set data. The corresponding regression lines are as follows: $CL^{ML} = 0.96CL + 0.002$, $r^2 = 0.61$; $CL^{MAP} = 0.73CL + 0.010$, $r^2 = 0.72$; $CL^{NN} = 0.69CL + 0.010$, $r^2 = 0.69$.

	$RMSpe$ (%)			
Estimator	V	CL	$y(t_1)$	$y(t_2)$
ML	21.	44.	16.	16.
MAP	14.	30.	12.	13.
NN	16.	31.	13.	14.

Table 2: Root Mean Square Prediction Errors ($RMSpe$) for Each Estimator.

error analysis for the NN estimator, Figure 3 was constructed from the test set data and estimation results. The upper panel shows the mean and standard deviation of the prediction error associated with the NN estimates of V in each of the indicated intervals. The corresponding results for CL are shown in the lower panel of Figure 3. These results could then be used to provide approximate error information corresponding to a particular point estimate (V^{NN} and CL^{NN}) from the neural network.

5 Discussion

These results demonstrate the feasibility of using a backpropagation trained neural network to perform nonlinear estimation from sparse data. In the example presented herein, the estimation performance of the network was shown to be similar to a Bayesian estimator (maximum a posteriori probability estimator). The performance of the trained network in this example is especially noteworthy in light of the considerable difficulty in resolving parameters due to the uncertainty in the mapping model inherent in this estimation problem, which is analogous to intersection of class distributions in classification problems.

While the particular example examined in this paper represents a realistic scenario involving the drug theophylline, to have practical utility the resulting network would need to be generalized to accommodate different dose infusion rates, dose times, observation times, and number of observations. Using an appropriately constructed training set, simulated to reflect the above, it may be possible to produce such a sufficiently generalized neural network estimator that could be applied to drug therapy problems in the clinical environment. It is of further interest to note that the network can be trained on simulations from a more complete model for the underlying process (e.g., physiologically based model as opposed to the compartment type model used herein), while still producing estimates of parameters that will be of primary clinical interest (e.g., systemic drug clearance, volume of distribution). Such an approach has the important advantage over traditional statistical estimators of building into the estimation procedure robustness to model simplification errors.

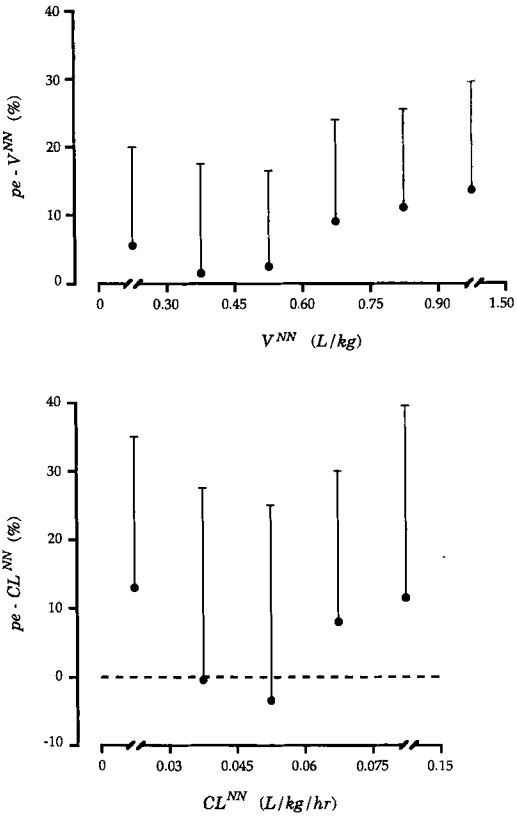

Figure 3: Distribution of prediction errors of volume (upper) and clearance (lower) for the NN estimator as obtained from the test set data. Prediction errors are displayed as mean (•) plus one standard deviation above the mean.

Acknowledgments

This work was supported in part by NIH Grant P41-RR01861. R.S. was supported by an IBM fellowship in Computer Science.

References

Asoh, H., and Otsu, N. 1989. Nonlinear data analysis and multilayer perceptrons. *IEEE Int. Joint Conf. Neural Networks* **II**, 411–415.

Burr, D. J. 1988. Experiments on neural net recognition of spoken and written text. *IEEE Trans. Acoustics Speech, Signal Processing* **36**, 1162–1165.

D'Argenio, D. Z., and Schumitzky, A. 1988. *ADAPT II User's Guide.* Biomedical Simulations Resource, University of Southern California, Los Angeles.

Gorman, R. P., and Sejnowski, T. J. 1988. Analysis of hidden units in a layered network trained to classify sonar targets. *Neural Networks* **1**, 75–89.

Kohonen, T., Barna, G., and Chrisley, R. 1988. Statistical pattern recognition with neural networks: Benchmarking studies. *IEEE Int. Conf. Neural Networks* **1**, 61–68.

Powell, J. R., Vozeh, S., Hopewell, P., Costello, J., Sheiner, L. B., and Riegelman, S. 1978. Theophylline disposition in acutely ill hospitalized patients: The effect of smoking, heart failure, severe airway obstruction, and pneumonia. *Am. Rev. Resp. Dis.* **118**, 229–238.

Rumelhart, D. E., Hinton, G. E., and Williams, R. J. 1986. Learning internal representations by backpropagation errors. *Nature* **323**, 533–536.

Sawchuk, R. J., Zaske, D. E., Cipolle, R. J., Wargin, W. A., and Strate, R. G. 1977. Kinetic model for gentamicin dosing with the use of individual patient parameters. *Clin. Pharmacol. Thera.* **21**, 362–369.

Sheiner, L. B., Halkin, H., Peck, C., Rosenberg, B., and Melmon, K. L. 1975. Improved computer-assisted digoxin therapy: A method using feedback of measured serum digoxin concentrations. *Ann. Intern. Med.* **82**, 619–627.

Vozeh, S., and Steimer, J. -L. 1985. Feedback control methods for drug dosage optimisation; Concepts, classification and clinical application. *Clin. Pharmacokinet.* **10**, 457–476.

Weideman, W. E., Manry, M. T., and Yau, H. C. 1989. A comparison of a nearest neighbor classifier and a neural network for numeric hand print character recognition. *IEEE Int. Joint Conf. Neural Networks* **1**, 117–120.

Received 16 November 1989; accepted 6 February 1990.