Sections and local group meeting reports.

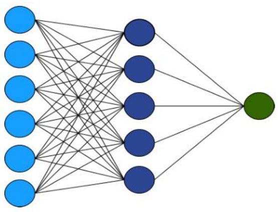
FeedForward Neural Nets as Statistical Models

Written by **Gilbert MacKenzie** on March 31st., 2023. Posted in Section and local group meeting reports.

The Northern Ireland local group of the RSS held an online meeting on Wednesday, October 26th., 2022, at 1pm (GMT), using MS Teams.

The speaker was Dr. Andrew McInerney, Department of Mathematics and Statistics, Limerick University, Ireland.

Dr. McInerney opened by reporting a resurgence of interest in merging statistical models and neural networks. The ultimate objective was to develop neural nets which could be interpreted. A key part of this goal was to have principled variable selection and Andrew argued for the use of information criteria, as in statistical models. The figure shows the structure of a simple feed-forward



neural net comprising an input layer (p x_is), a hidden layer (q nodes) and output $(y_i)s$ for $i = 1, ..., n$ units. It is convenient to represent the neural net mathematically as:

$$NN(x_i) = \gamma_0 + \sum_{k=1}^{q} \gamma_k \, \phi\Big(\sum_{j=0}^{p} \omega_{jk} x_{ji}\Big)$$

Converting this to a statistical model we have:

$$y_i = NN(x_i, \theta) + \epsilon_i \qquad \text{where} \quad \epsilon_{\rm i} \sim \mathrm{N}(0, \sigma^2)$$

and a standard Gaussian $\log_e$ likelihood, $\ell(\theta)$, for inference (here θ represents the collection of all parameters to be estimated). Standard theory leads to $\hat{\theta} \sim N(\theta, \Sigma)$ where $\hat{\Sigma} = I_o(\hat{\theta})^{-1}$, where $I_o(\hat{\theta})$ is the observed, positive definite, information matrix, whence the uncertainty in θ can be assessed.

However in Neural Nets the information matrix has frequently been found to be singular, due to presence of *redundant* hidden nodes (over-fitting), whence $I_o(\hat{\theta})$ becomes less than full rank and some parameters unidentifiable. Andrew argued cogently for principled *model selection* methods which would eliminate such redundancies. He adopted the BIC criterion for model selection

$$BIC(\theta) = -\ell(\hat{\theta}) + \log_e(n)(K+1) \qquad \text{where} \quad K = (p+2)q + 1$$

rather than the out-of-sample MSE used in Machine Learning (ML).

A three phase (backwards) search strategy of the Model space was proposed, comprising: hidden-node selection, input-node selection followed by fine tuning. Andrew showed the results of a small simulation of a Neural Net with true structure $p = 3$, $q = 3$, to which 10 redundent input nodes had been added and in which there was a maximum of 10 hidden nodes. The primary comparison between the Statistical and ML approaches is shown in the table below for a sample of $n = 500$ units: The statistical approach is better fitting in terms of

Table: Results comparing the Statistical approach with ML

		Performance metrics			
n	Method	BIC	MSE	MSE (Test)	K (16)
500	BIC-based	**1152.9**	0.51	**0.53**	**16**
	MSE-based	1267.2	**0.50**	0.57	37

BIC, has the same MSE, and uses half the number of parameters.

With the availability of the estimated variance-covariance matrix, Andrew went on to show how to use Wald tests to explore the structure of the Neural Net and, to aid interpretation further, he defined *covariate effects* and their standard errors and showed how to estimate and plot these new quantities.

He concluded with a comprehensive Boston Housing example, analysed in his new R package `statnnet`, which illustrated all of the newly developed methods.

This was simply an excellent and very timely talk showing the advantages of Statistical Modelling in a ML context. It was very well received by the meeting and the speaker was congratulated in the usual way. A short discussion ensued and the speaker responded to several questions about the model selection strategy, methods for deeper learning nets and pruning to natural structures.

The Chair thanked the speaker for a very stimulating talk and concluded the meeting by thanking everyone for their attendance and support.

References
McInerney, A. and Burke, K. (2022). A Statistically-Based Approach to Feedforward Neural Network Model Selection. arXiv preprint arXiv:2207.04248.

```
R package
devtools::install_github("andrew-mcinerney/statnnet")
library(statnnet)
nn <- selectnn(medv~ ., data = Boston, Q = 10, n_init = 10, maxit = 5000)
```