# A Neural Network Model for Predicting Nationalities when IIA Fails

*Zhaokun Zhang, Zihan Zhang*

**Abstract**

In this note, we are going to discuss the elements of Neural Network and its application when IIA assumption fails. We adopt Neural Network for predicting nationalities instead of multinomial logit regression in such case and compare the results.

## 1 Principles of Neural Network

### 1.1 Defination

Neural networks are computing systems inspired by the biological neural networks and astrocytes that constitute animal brains. The neural network itself is not an algorithm, but rather a framework for many different machine learning algorithms to work together and process complex data inputs. Such systems “learn” to perform tasks by considering examples, generally without being programmed with any task-specific rules.

Neural networks are based on a collection of connected units or nodes called artificial neurons, which loosely model the neurons in a biological brain. Each connection, like the synapses in a biological brain, can transmit a signal from one artificial neuron to another. An artificial neuron that receives a signal can process it and then signal additional artificial neurons connected to it.

Neural networks help us cluster and classify. You can think of them as a clustering and classification layer on top of the data you store and manage. They help to group unlabeled data according to similarities among the example inputs, and they classify data when they have a labeled dataset to train on. Besides, neural networks can also extract features that are fed to other algorithms for clustering and classification. So you can think of deep neural networks as components of larger machine-learning applications involving algorithms for reinforcement learning, classification and regression.

In all, simply speaking, the neural network is a computing system made up of a number of simple, highly interconnected processing elements, which process information by their dynamic state response to external inputs.

### 1.2 Elements

Neural networks are typically organized in layers. Layers are made up of a number of interconnected “nodes” which contain an “activation function”. A neural networks of L layers contains L-1 hidden layers and one output layer. Each hidden layers contains 1+d units (also called “neurons”), indexed 0,1,. . . ,d, with unit 0 being the bias unit. Patterns are presented to the network via the input layer, which communicates to one or more hidden layers where the actual processing is done via a system of weighted “connections”. The hidden layers then link to an output layer where the answer is output.

### 1.3 Advantages and Limitations

Depending on the nature of the application and the strength of the internal data patterns you can generally expect a network to train quite well. This applies to problems where the relationships may be quite dynamic

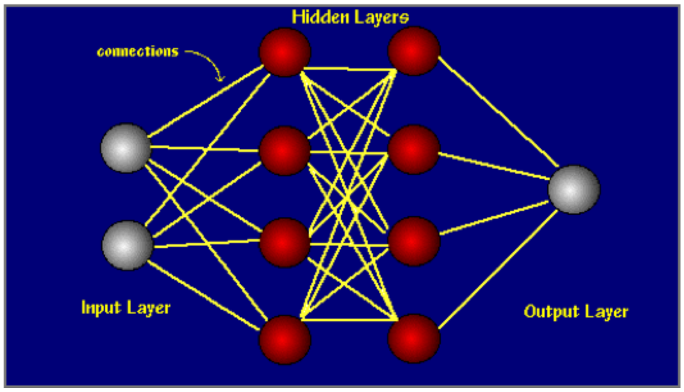


Figure 1: Neural Network

or non-linear. Artificial neural networks (ANN) provide an analytical alternative to conventional techniques which are often limited by strict assumptions of normality, linearity, variable independence etc. Because an ANN can capture many kinds of relationships, it allows the user to quickly and relatively easily model phenomena which otherwise may have been very difficult to explain.

However, there are some specific issues potential users should be aware of. One drawback is that many types of networks are in a sense the ultimate "black boxes". Apart from defining the general architecture of a network and perhaps initially seeding it with a random number, the user has no other role than to feed it input and watch it train and await the output.

# 2 Real Analysis

In our previous homework note, we have tried to predict the nationalities of citizens in various countries according to their different social recognition on same questions. In this note, we are going to use the same data set with less variables observed to do a similar task with a different method. We are going to predict the nationality of residents from the UK, France, Italy, Russia and Sweden with their attitudes towards society and politics. Multinational logit regression and Neural network would be adopted.

The data comes from the ESS survey 2016 data with 534 variables and 44,317 observations included initially. We did data cleaning, removing some individuals with omitted values in some variables and keeping only variables related to politics and society in 5 countries mentioned above. Finally our data set contains 2,882 observations and 25 variables inside.

```
## import data
rm(list = ls())
library(AER)
library(mlogit)
library(glmnet)
library(nnet)


ESS8_c5 <- read.csv("ESS8_c5.csv",header = T)
ESS8_c5 <- ESS8_c5[,-1]
ess <- na.omit(ESS8_c5)
n=nrow(ess)
train <- sample(n,0.8*n)
ess.tr <- ess[train,] ## Training set
ess.test <- ess[-train,] ## Test set
```

```r
x <- data.frame(ess.tr[,3:26])
x2 <- data.frame(ess.test[,3:26])
y <- ess.tr$cntry
x1 <- as.matrix(x)
```

## 2.1 Predicting with Multinomial Logit Regression

First we adopted the multinational logit regression for the prediction.

```r
attach(ess.tr)
mlfit <- multinom(cntry ~ . ,data = data.frame(ess.tr[,2:26]))
mlfit.yhat <- predict(mlfit, ess.test)
t1 <- table(mlfit.yhat,ess.test$cntry,dnn = c("predicted","true"))
pred_accuracy_multi <- 1 - sum(diag(t1))/sum(t1)
detach(ess.tr)
```

```r
t1
```

```
##          true
## predicted  FR  GB  IT  RU  SE
##        FR 132  26  25   3  12
##        GB  30  78  14   7  28
##        IT  10   4  27   1   3
##        RU   8   8   3  10   1
##        SE  11  38   3   0  95
```

```r
pred_accuracy_multi
```

```
## [1] 0.407279
```

We could see from the result that the error rate was almost 40 percent.

## 2.2 IIA assumption and Test

Generally speaking, Independence of Irrevelant Alternatives assumption suggests that the preference of an individual between item A and item B does not change when another item X alters if IIA assumption holds, which means the outcome of a choice between two items only depend on the features of these two items and it is not affected by the features or characteristics of other items or variables.

Our dataset contains five countries located in the European continent, including The UK, France, Italy, Russia and Sweden. From the map we can discover that the England, France and Italy locate in a close region, while the other countries are separated by a relatively long distance. Therefore, it is reasonable to infer that citizens in Britain, France and Italy may be correlated in some aspects because they are neighbors. In other word, the IIA assumption could be violated. Hence, it's better to check the assumption and improve our method.

Hausman and McFadden (1984) proposed a generalized method with underlying principle of removing one alternatives, estimating the model again and comparing the results. This method is a wide used way to check the validity of IIA assumption.

The statistic generated by Hausman and McFadden could be represented as:

$$H_1 = [\widehat{\beta_\mathbf{J}} - \widehat{\beta_\mathbf{J-1}}][\widehat{\mathbf{V_{J-1}}} - \widehat{\mathbf{V_J}}][\widehat{\beta_\mathbf{J}} - \widehat{\beta_\mathbf{J-1}}]$$

which follows a Chi-squire distribution.

Now we would like to apply this test to our model.

```r
# HMF test
long<- mlogit.data(ess.tr[,2:26], shape="wide",  choice = "cntry")
m1 <- mFormula(cntry~ 0|netusoft+ppltrst+pplfair+pplhlp+polintr+psppsgva+actrolga+psppipla+cptppola+trst
mlfit2 <- mlogit(m1,long,reflevel = "FR",seed = 100)

mlfit3 <- mlogit(m1,long,alt.subset = c("FR","GB","IT"))
hmftest(mlfit2,mlfit3)
```

```
##
##  Hausman-McFadden test
##
## data:  long
## chisq = -39.942, df = 48, p-value = 1
## alternative hypothesis: IIA is rejected
```

The result shows the P-value was close to 1, which means the alternative assumption, IIA holds, was rejected, indicating that there might be some correlations among these countries and the result from multinomial logit regression could be imprecise.

### 2.3 Predicting with 2-layer neural network

Since IIA assumption doesn't hold, we have to improve our method in predicting. Neural network could be an ideal alternatives for prediction.

```r
# Fit a neural network
x <- scale(x,center = T,scale = T)
x2 <- scale(x2,center = T,scale = T)
ess.tr1 <- data.frame(ess.tr$cntry,x)
ess.test1 <- data.frame(ess.test$cntry,x2)
netfit = nnet(ess.tr.cntry ~.,ess.tr1,maxit=10000, size=10,
           decay=0.25) ## the value of decay is generated from CV method.

net.hat <- predict(netfit,ess.test1,type ="class" )
t2<- table(net.hat,ess.test1$ess.test.cntry,dnn = c("predicted","true"))
pred_accuracy_net <- 1 - sum(diag(t2))/sum(t2)
```

```r
t2
```

```
##          true
## predicted  FR  GB  IT  RU  SE
##        FR 134  23  28   4  14
##        GB  19  97   6   5  29
##        IT  12   2  31   1   5
##        RU   7  10   2  11   1
##        SE  19  22   5   0  90
```

```r
pred_accuracy_net
```

```
## [1] 0.3708839
```

## Reference


[1] Artificial neural network, Wikipedia, https://en.wikipedia.org/wiki/Artificial_neural_network

[2] “A Basic Introduction to Neural Networks”, http://pages.cs.wisc.edu/~bolo/shipyard/neural/local.html

[3] Huynh T Q , Setiono R . Effective neural network pruning using cross-validation[C]// IEEE International Joint Conference on Neural Networks. IEEE, 2005.

[4] Jiaming Mao. Neural Networks.https://jiamingmao.github.io/data-analysis/assets/Lectures/Neural_Networks.pdf

[5] Zihan Zhang. Independence of Irrelevant Alternatives (IIA) Assumptions and some Extentions. https://landbuland.github.io/moments/2019/04/07/Homework-3-for-Data-Analysis.html

## Appendix

Table 1: Interpretation of variables

| Variables | Theme | Interpretation |
|---|---|---|
| cntry | | Country |
| netusoft | Society | Internet use, how often |
| ppltrst | Society | Most people can be trusted or you can't be too careful |
| pplfair | Society | Most people try to take advantage of you, or try to be fair |
| pplhlp | Society | Most of the time people helpful or mostly looking out for themselves |
| polintr | Society | How interested in politics |
| psppsgva | Society | Political system allows people to have a say in what government does |
| actrolga | Society | Able to take active role in political group |
| psppipla | Society | Political system allows people to have influence on politics |
| cptppola | Society | Confident in own ability to participate in politics |
| trstprl | Society | Trust in country's parliament |
| trstlgl | Society | Trust in the legal system |
| trstplc | Society | Trust in the police |
| trstplt | Society | Trust in politicians |
| trstprt | Society | Trust in political parties |
| trstep | Society | Trust in the European Parliament |
| trstun | Society | Trust in the United Nations |
| lrscale | Politics | Placement on left right scale |
| stflife | Politics | How satisfied with life as a whole |
| stfeco | Politics | How satisfied with present state of economy in country |
| stfgov | Politics | How satisfied with the national government |
| stfdem | Politics | How satisfied with the way democracy works in country |
| stfedu | Politics | State of education in country nowadays |
| stfhlth | Politics | State of health services in country nowadays |

Table 2: Descriptive statistics of variables

| | France | | | Britain | | | Italy | | | Sweden | | | Russia | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Obv. | Mean | ST.D | Obv. | Mean | ST.D | Obv. | Mean | ST.D | Obv. | Mean | ST.D | Obv. | Mean | ST.D |
| netusoft | 951 | 4.13 | 1.45 | 773 | 4.29 | 1.27 | 304 | 4.19 | 1.34 | 686 | 4.63 | 1.00 | 168 | 3.78 | 1.65 |
| ppltrst | 951 | 4.66 | 2.08 | 773 | 5.49 | 2.14 | 304 | 5.01 | 2.15 | 686 | 6.42 | 1.96 | 168 | 4.23 | 2.37 |
| pplfair | 951 | 6.00 | 1.94 | 773 | 5.85 | 1.98 | 304 | 4.89 | 2.10 | 686 | 6.81 | 1.74 | 168 | 5.14 | 2.30 |
| pplhlp | 951 | 4.81 | 2.08 | 773 | 5.79 | 1.86 | 304 | 4.20 | 2.05 | 686 | 6.17 | 1.80 | 168 | 4.76 | 2.28 |
| polintr | 951 | 2.49 | 0.95 | 773 | 2.16 | 0.85 | 304 | 2.69 | 0.85 | 686 | 2.15 | 0.74 | 168 | 2.25 | 0.79 |
| psppsgva | 951 | 2.13 | 0.89 | 773 | 2.54 | 0.82 | 304 | 1.74 | 0.71 | 686 | 2.58 | 0.83 | 168 | 2.07 | 0.91 |
| actrolga | 951 | 2.05 | 1.04 | 773 | 2.44 | 1.12 | 304 | 2.16 | 0.98 | 686 | 2.63 | 1.02 | 168 | 1.90 | 0.96 |
| psppipla | 951 | 2.02 | 0.85 | 773 | 2.44 | 0.80 | 304 | 1.78 | 0.69 | 686 | 2.73 | 0.86 | 168 | 1.87 | 0.82 |
| cptppola | 951 | 2.27 | 0.90 | 773 | 2.47 | 1.09 | 304 | 2.37 | 0.86 | 686 | 2.62 | 1.00 | 168 | 1.97 | 1.06 |
| trstprl | 951 | 4.17 | 2.22 | 773 | 4.78 | 2.33 | 304 | 3.61 | 2.43 | 686 | 6.23 | 2.15 | 168 | 3.99 | 2.62 |
| trstlgl | 951 | 5.00 | 2.31 | 773 | 6.01 | 2.22 | 304 | 4.73 | 2.44 | 686 | 6.43 | 2.02 | 168 | 4.05 | 2.77 |
| trstplc | 951 | 6.47 | 2.06 | 773 | 6.66 | 2.09 | 304 | 6.35 | 1.88 | 686 | 6.81 | 1.94 | 168 | 4.18 | 2.82 |
| trstplt | 951 | 2.91 | 2.00 | 773 | 3.80 | 2.20 | 304 | 2.47 | 2.17 | 686 | 4.88 | 1.90 | 168 | 3.48 | 2.54 |
| trstprt | 951 | 2.84 | 1.96 | 773 | 3.93 | 2.06 | 304 | 2.52 | 2.22 | 686 | 4.96 | 1.86 | 168 | 3.60 | 2.58 |
| trstep | 951 | 3.87 | 2.23 | 773 | 3.70 | 2.36 | 304 | 4.45 | 2.34 | 686 | 4.87 | 1.99 | 168 | 2.88 | 2.56 |
| trstun | 951 | 4.95 | 2.35 | 773 | 5.59 | 2.23 | 304 | 5.02 | 2.34 | 686 | 6.31 | 1.93 | 168 | 2.95 | 2.56 |
| lrscale | 951 | 4.94 | 2.40 | 773 | 4.88 | 1.88 | 304 | 4.92 | 2.25 | 686 | 5.11 | 2.28 | 168 | 5.15 | 1.98 |
| stflife | 951 | 6.52 | 2.19 | 773 | 7.37 | 1.91 | 304 | 7.25 | 1.77 | 686 | 8.07 | 1.53 | 168 | 5.65 | 2.27 |
| stfeco | 951 | 3.48 | 1.78 | 773 | 5.00 | 2.01 | 304 | 3.77 | 2.04 | 686 | 6.12 | 1.91 | 168 | 3.91 | 2.17 |
| stfgov | 951 | 3.16 | 1.98 | 773 | 4.66 | 2.19 | 304 | 3.37 | 2.16 | 686 | 4.84 | 1.96 | 168 | 4.63 | 2.31 |
| stfdem | 951 | 4.27 | 2.39 | 773 | 5.32 | 2.35 | 304 | 4.08 | 2.25 | 686 | 6.45 | 2.09 | 168 | 4.38 | 2.24 |
| stfedu | 951 | 4.69 | 2.01 | 773 | 5.54 | 2.00 | 304 | 5.21 | 2.15 | 686 | 5.15 | 1.91 | 168 | 4.67 | 2.03 |
| stfhlth | 951 | 6.19 | 1.94 | 773 | 5.52 | 2.21 | 304 | 5.56 | 2.16 | 686 | 5.63 | 2.01 | 168 | 3.89 | 2.02 |