

Advanced Microeconomics

Problem Set 3: Machine Learning - Neural Network

Hans Martinez

30 Mar 2021

Neural network (NN) training

For this assignment, I used **R** for wrangling the data and then adapted the code `Flux_logistic_regression.jl` to allow it to handle my transformed data. I took a data-driven approach in terms of coming up with a model to feed the NN. That is, instead of trying to figure out which variables would make most sense from a theory perspective, I just tried to use all the information possible.

After transforming the data I got a little less than 3200 variables. Initially, I used eight layers. However, I was getting low accuracy. I started decreasing the number of layers and end up with just one. Then, I tweaked the learning rate and number of repetitions. This could have been done in a more efficient way by cross-validation. Alternatively, I thought about implementing a crude random search. In short, we could sample from the 3200 variables randomly, allowing for different number of total *features*. Then pick the lowest and do a grid search around those variables. We can think of doing this also for the parameters. This would be more tractable if only one variable is tweaked at a time.

The accuracy and value of the loss function are displayed in table 1. The confusion matrix is displayed in table 2. As usual, the code for this assignment can be found in my Github repo.

Table 1: NN training results

Loss	Accuracy
0.216889	0.9214443

Table 2: Confusion matrix

1035	2	1	0	0	0
0	484	0	7	1	0
0	0	881	0	0	1
2	99	17	439	14	13
0	15	0	1	439	27
0	0	1	3	118	499