PROJECT TITLE: Predicting Opioid Addiction Likelihood Using Neural Nets

PROJECT SUMMARY: Utilize artificial and convolutional neural networks in TensorFlow to classify individuals at risk for opioid addiction based off socioeconomic, familial, and personal health data.  Datasets include US Census Data, Center for Disease Control, and National Institute on Drug Abuse.

MILESTONES:

* Data collection - DONE
* Data exploration – DONE
* Feature Exploration - DONE
* Build models: ANN, CNN - DONE
* Model Tuning – DONE
* Result visualization - DONE

PROPOSED ‘TO DO’ FROM THE LAST WEEK (Copy & Paste from your previous week’s TO DO)

*This week I want to finalize my neural net composition and begin analyzing the results. I also want to begin my final report/presentation, as it helps force me to organize my code and think about my whole process in a more comprehensive way.*

THIS WEEK’S PROGRESS

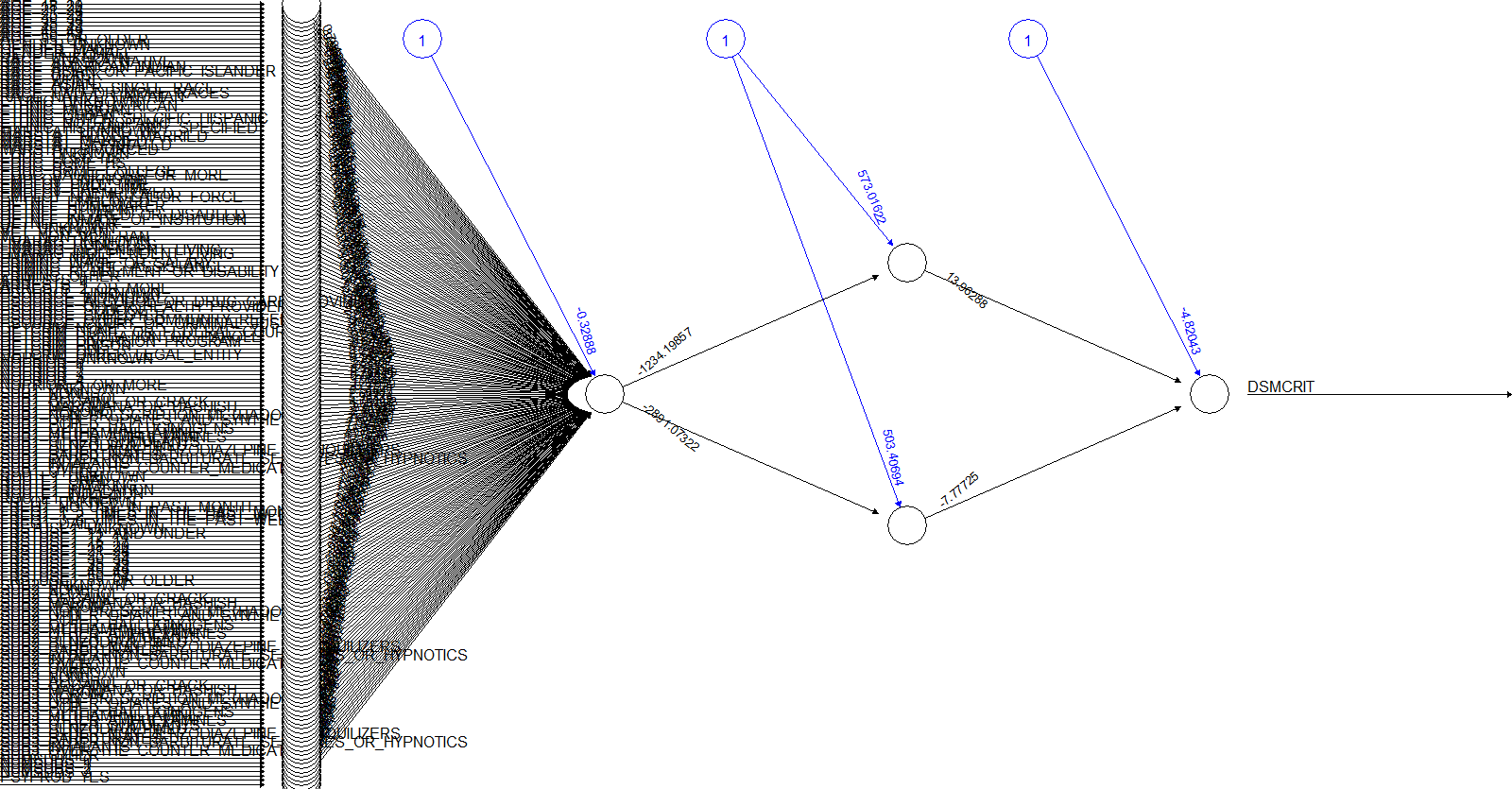
Given previous issues with data size and processing abilities, I refined my data cleansing process and limited my sample size for my models, ultimately allowing for a data set that can be processed in just a couple of hours with 4 different models. I decided to use Spark to distribute the processing of the data as well, however, that is only being distributed across the cores of my computer. In addition to finalizing the models, I have been working to analyze the results and create confusion matrices to establish the accuracy, precision, and recall.

ISSUES AND DISCUSSION:

I wanted to use AWS to run my models, but the time crunch, and getting a new job (woo!) limited the amount of time I could dedicate to standing up the instance and running the model. I was able to create the EC2 instance, load data into the S3 buckets, and begin linking R and my AWS instance, but found that it would take away the time necessary to actually create reliable, consistent models.

TO DO:

I need to finish analyzing the results and being cleaning and organizing my code. Most of my week will be dedicated to creating the final presentation and organizing my data in a github repository.



Result of most accurate neural net.