

4.11.20

Kelly Chinn

Kal Academy AI Bootcamp Assignment 3

ANN model – Heart Disease dataset

---

Model: ANN with one hidden layer and two dropout layers

Accuracy: 0.81 – 0.85

How I chose my model:

I tried many different combinations before I settled on this one. I ran several sets in which I converted some of the categorical columns (using `pd.get_dummies`), and several sets where I left those columns as is. I found the model accuracy was slightly higher using the categorical set. I also tried one hidden layer vs. two, and varied dropout layers from 0 to 2, with dropout rates from 0.1 – 0.5. Results were better with the dropout layers, and it seemed maybe slightly better with two than with one. I didn't find a measurable improvement with an additional hidden layer, so I kept just one. Additionally, I experimented with initial Dense layer units ranging from 128 down to 16; the most consistent model performance was with an input layer of 16 and a hidden layer with 8 units.

Since the ANN has some random functionalities, the accuracy and loss scoring varied on each run. Most of the models I tried were consistently around 0.80 – 0.83 accuracy, except this final one which averaged 0.83 – 0.85. I wasn't able to find any parameter tweaks to get it higher.

Other notes:

I also tried to see if I could make a working model using the original 5 value target column. I was able to make a model with an output Dense layer with 5 units, but it gave me an error when trying to run the metrics. (ValueError: Classification metrics can't handle a mix of multilabel-indicator and multiclass targets.) I did a little research and found a post on GitHub saying it has to do with the encoding not working with GridSearchCV and cross\_validation. I haven't had a chance to try the suggestions there yet, though I plan to just to see if I can make it work.