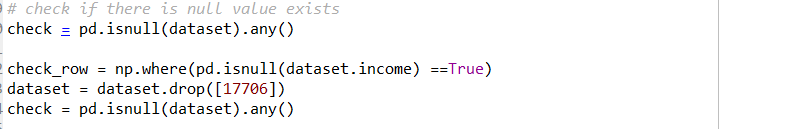
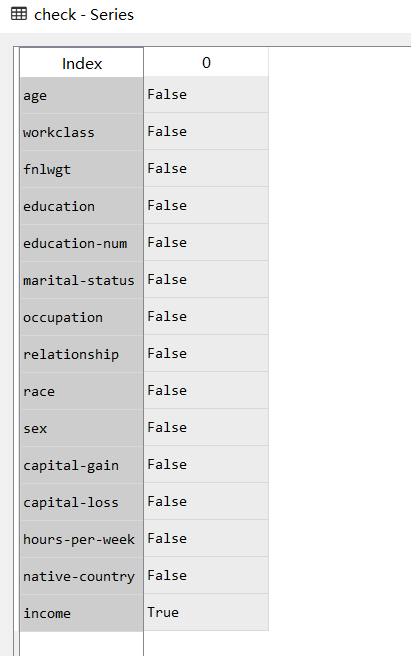
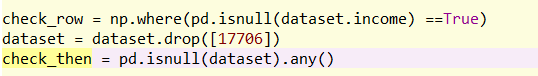
1. Description of the variables
2. Preprocessing of variables
3. deal with NULL value



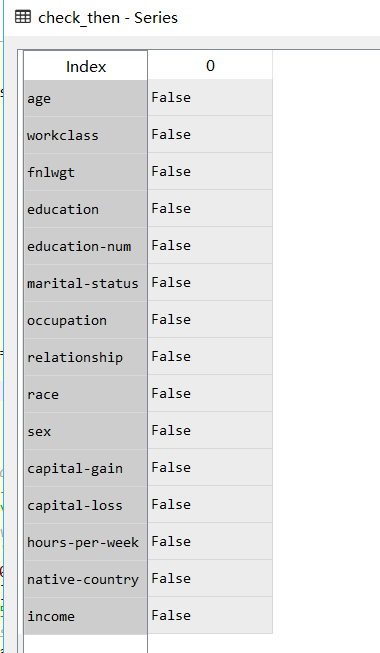
First import the data into spider, then we use the isnull()function to check if there is null value in the data set, then we can see :



It shows that there is one variable called ‘ income’ that has NULL value, ‘income’ variable is our target variable, the class of this dataset, so I drop the rows that has NULL values in ‘income’ variable.

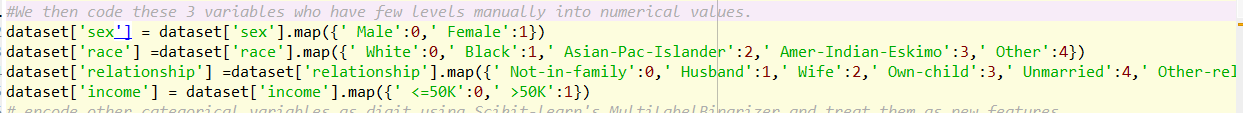


Then there is no NULL value in this data set:

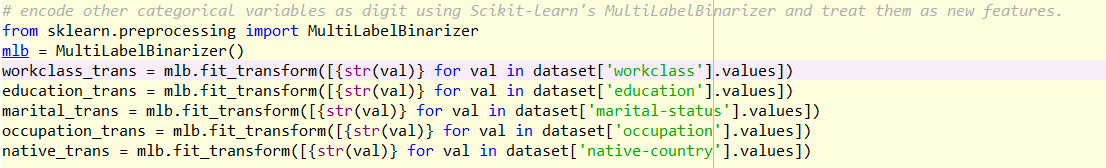


1. Use MultiLabelBinarizer in sklearn library to encode the categorical variables: in this dataset we have 8 categorical variables, which are: ‘workclass’, ‘education’, ‘marital-status’, ‘occupation’, ‘relationship’, ‘race’, ’sex’, ‘native-country’

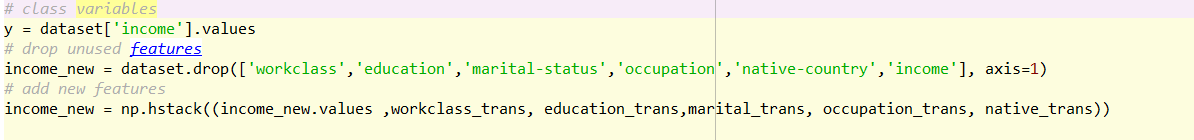
There are some attributes that has few levels, so we just manually encode them



For other attributes, use MultibLabelBinarizer to encode them:



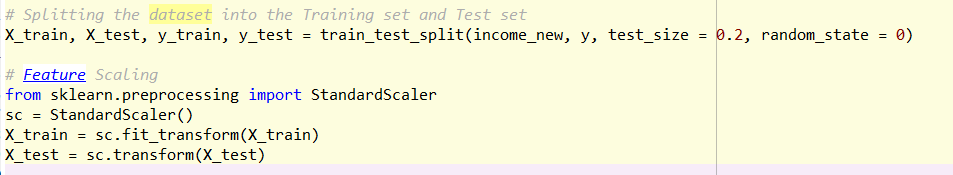
1. Then form the new attributes into a whole data set



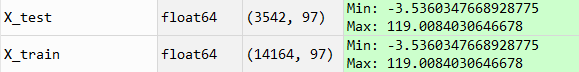


For the new data set, there are 97 columns

1. Then we split the dataset into the Training set and Test set: Test set take 20% of total data set, and do features scaling.

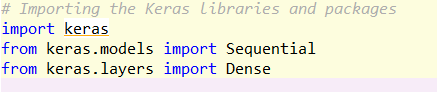


Then we can get:





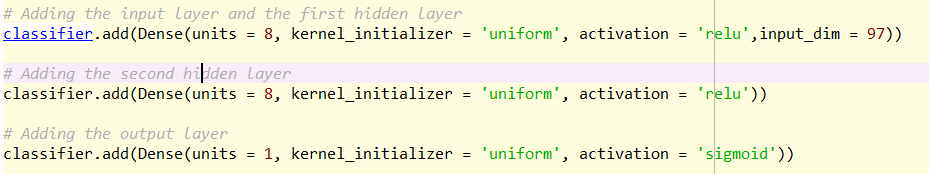
1. ANN-making
2. Importing the Keras libraries and packages



1. Initializing the ANN



1. Adding layers of this model



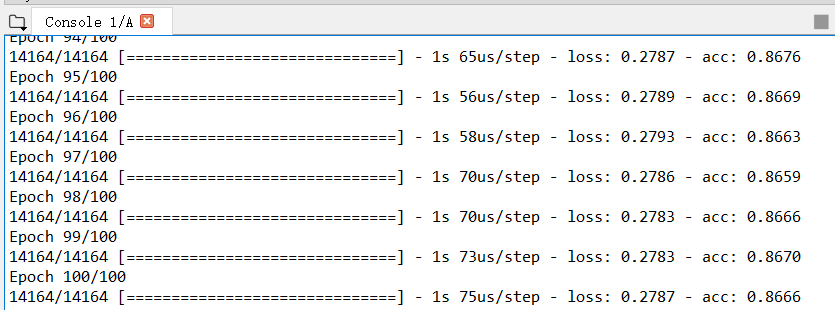
1. Compile the ANN



1. Fit the ANN to the Training set

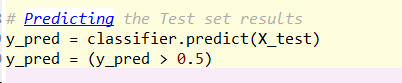


We can see from the console:



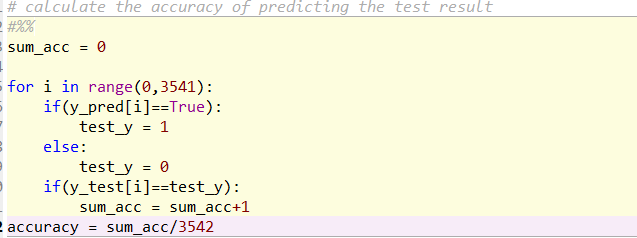
The accuracy is 86.66%

1. Predicting the Test set results



1. Evaluation

After the predicting made by ANN model, we can get the predicted results of test set:



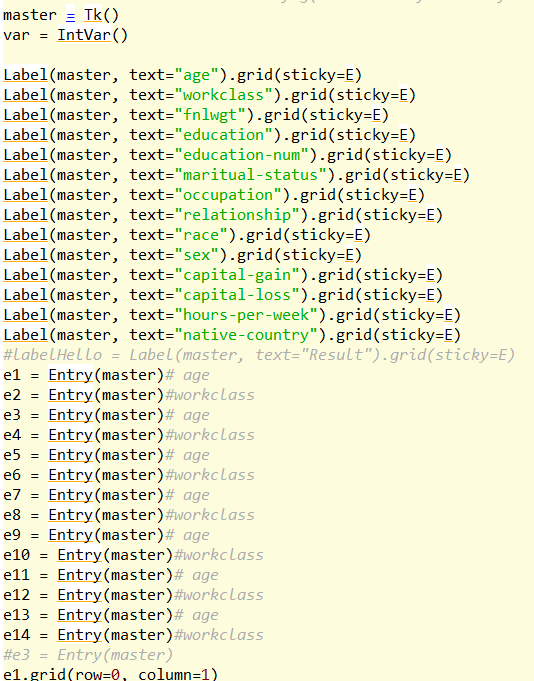
According to the console:



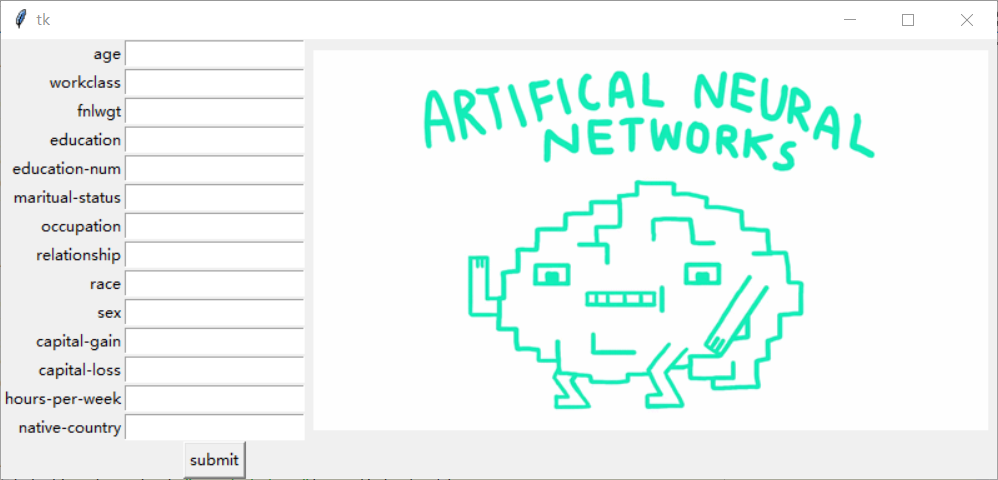
We can find at last, the accuracy of this model is 84.13%, which means there is 84.13% possibility for prediction is correct

1. User friendly interface

I use the tkinter to develop the interface of parameter input:



The result is:



When input the parameters, we can get the prediction for if this person has income >50K dollars per year.

