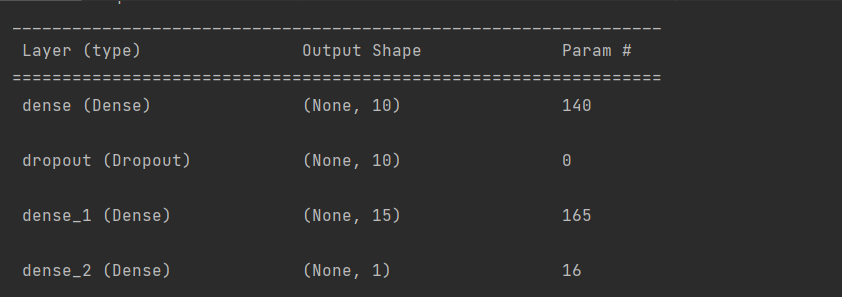
## Report

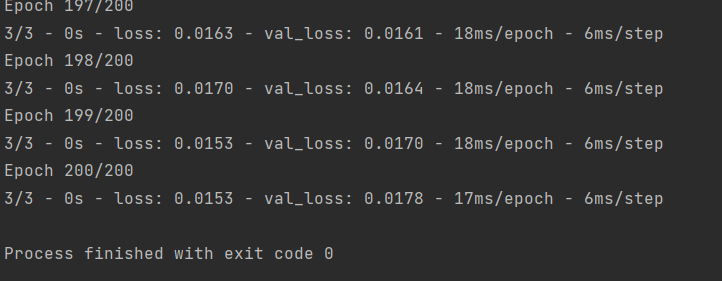
This is a group fufilled by Mei, Iva and Yiqiu.

We use the boston housing price dataset as our dataset, which is a classical dataset, we can import the dataset by using python itself. The dataset has 13 features and 1 result price. We will build model to train the model and give the model ability to predict the price according the 13 features. We build two different models, and get each result, and we will make a comparison between the two models we build.

BP Neural Network:

In the first one, we build the BP neural network model. This model use the combination of the keras package and the BP neural network. We use all the 13 features as the data and the price as the target. There are several steps we did in building this model. Firstly, we divide data into train data and test data, and we pre-process the data to normalize it and avoid the influence of big difference in dataset. Then we build the network structure, in which we have fully connected layers and one drop layer, the function of the drop layer is to prevent overfitting problem. The activation function we use is the ‘relu’ and ‘linear’. In final, we calculate the mse as the loss to evaluate the accuracy and we visualize the training process and the change of the loss function.

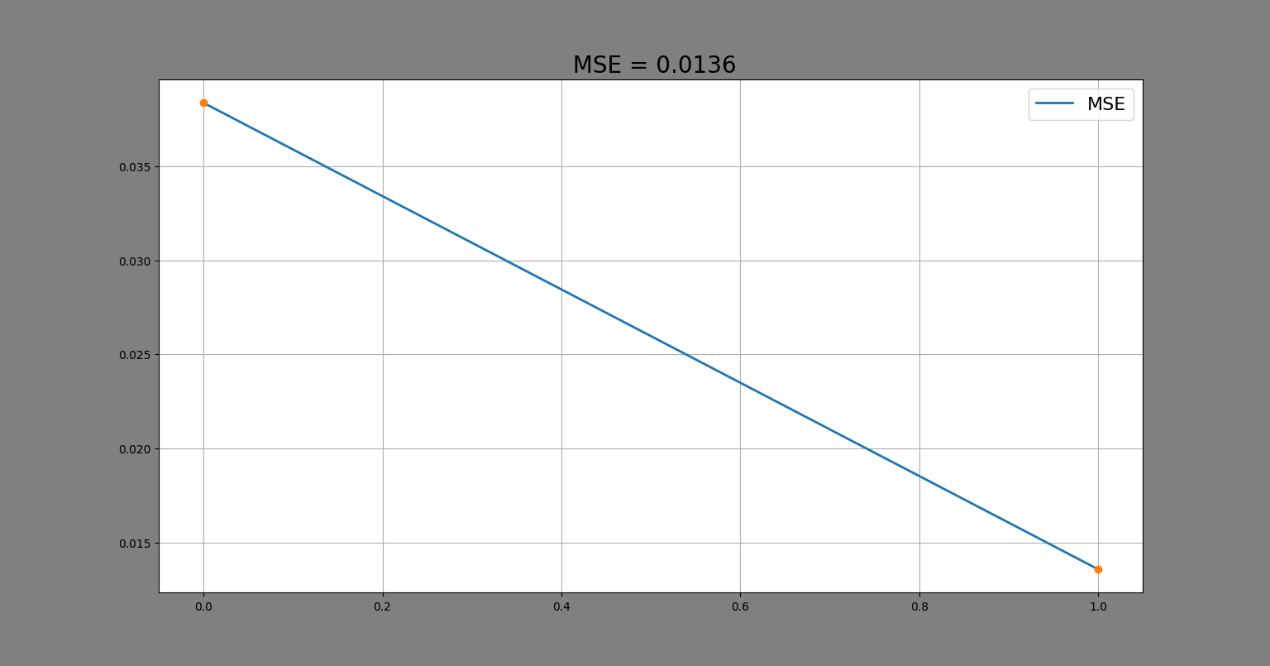


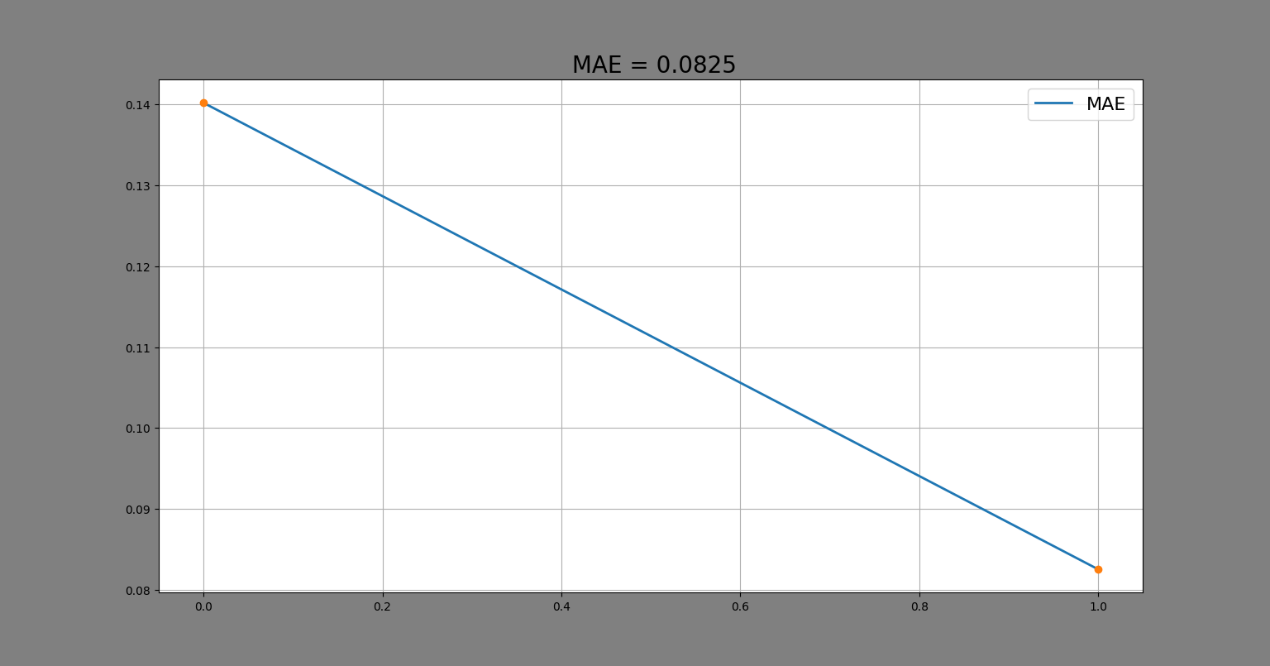


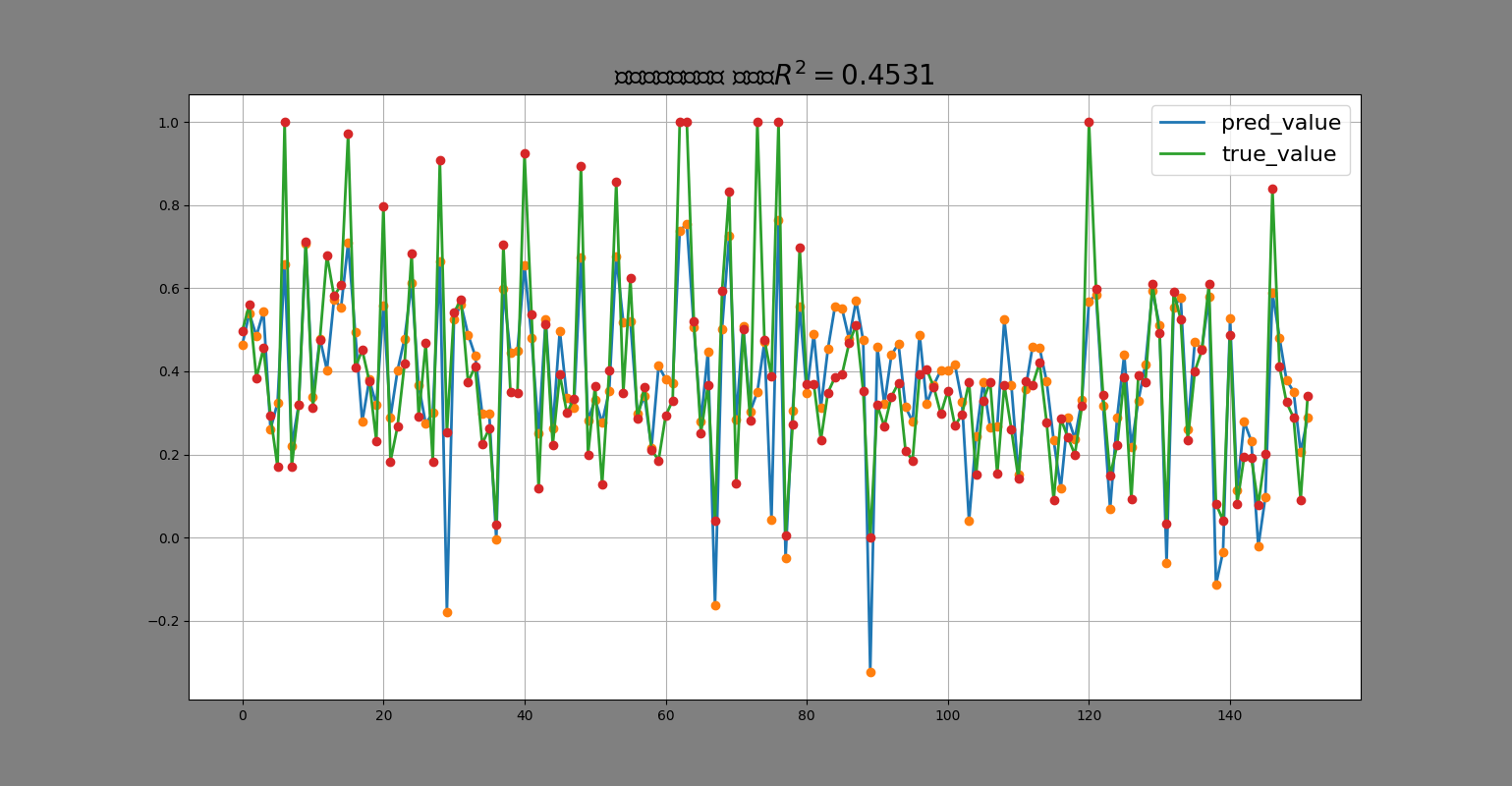
This is the final loss we got, it is around 0.0153, that is a pretty low rate, which means the effect of our model is good.

Linear Regression

For making a comparison to the NN model we build, we build another model using a simpler algorithen, we choose the linear regression model to compare our result. We also select the Boston house price dataset as our dataset. Same as what we do in the NN model, we firstly pre-process the dataset to check whether or not it has the null data, we check the mean, max, min of the data and see the correlation of the features and the price. We screen out the correlation rate features which is above 0.5, and find that the LSTAT、PTRATIO、RM three features have the high correlation with the target. So instead of using all the features to predict as we do in the NN model, this time we only select the three features having the high correlation to build the model. We plot the relation between the three features and the target. Then we use the training data to linear predict and use the test data to test. We calculate and plot the MSE and MAE to evaluate the accuracy.







As the result, we can see the final mse is 0.0136 and the final mae is 0.0825.

Comparison:

After comparing the two models, I find it is an unsuccessful experiment, because the two models have almost the same mse rate, which means they have almost the same accuracy. We expected the NN model will have higher accuracy than the linear one, because we thought the algorithem is more complex and the features it takes account in is more. We think there are two main reasons leading to the same accuracy, the first one is about the dataset, to predict the price by using the features seems not be a complex task, we assume the NN model can have a higher performance in more complex task such as the image recognition, both the two models have a very low loss rate in dealing with the regression predict task. Second is that we only have very limited layers in our NN model, which could also be the reason. In conclusion, both the two models have a high performance in predicting the linear number, and in this task the NN model have almost the same accuracy with the linear prediction model.

Task to be continued and explanation:

We zip the project file to submit. The NN model is the main.py and the linear regression model is the main2.py. We also have an unfinished model saved as main3.py. In the NN model, we use the keras and the function provided by it and python to build the neural network, but the package and the function makes the coding pretty easy. Therefore in the main3.py, we tried to build the model in BP neural network model all by ourselves. But the time is limited and we have to submit what we have done before the deadline. We will try to figure out how to build the NN model in the spare time and we would also introduce that in our presentation.