

Multi-Layer Perceptron

ZIH-SYUAN CHOU

Department of Applied Mathematics

National Chung Hsing University)

Taichung, Taiwan

apple7102@gmail.com

Abstract—Derive the forward and backward algorithm for 3-layers multilayers perceptron (MLP) in the Regression Problem. And implement 3-layers MLP for Boston housing price regression problem, moreover implement dropout, Xavier initialization on 3-layer MLP and compare the performance of the above models.

Index Terms—3-layers MLP, dropout, Xavier initialization, Boston housing price dataset

I. DERIVE FORWARD AND BACKWARD ALGORITHM

Multilayer perceptron is a kind of forward propagation neural network, which contains at least three layers (input layer, hidden layer and output layer), and uses the technology of "backward propagation" to achieve supervised learning of learning (model learning).

A. Forward propagation

B. Backward propagation

II. EXPERIMENTATION

In this report, we wil use exploratory data anlysis first to determine which features of training data are important and using them to train the model.

A. Data processing

This is project is to predict Boston housing price,this data set just have 504 records.Splitting this data set with six percent of training data and forty percent of testing data.we will standardize training data to let the value of training data are between -1 and 1, this step can avoid model focus on some features which value is bigger than other.

The following introducing the meaning of each features for train data:

- CRIM: per capita crime rate by town
- ZN: proportion of residential land zoned for lots over 25,000 sq.ft.
- INDUS: proportion of non-retail business acres per town
- CHAS: Charles River dummy variable (= 1 if tract bounds river; 0 otherwise)
- NOX: nitric oxides concentration (parts per 10 million)
- RM: average number of rooms per dwelling
- AGE: proportion of owner-occupied units built prior to 1940
- DIS: weighted distances to five Boston employment centres
- RAD: index of accessibility to radial highways
- TAX: full-value property-tax rate per 10,000

- PTRATIO: pupil-teacher ratio by town
- B: where B_k is the proportion of blacks by town
- LSTA: percent lower status of the population
- MEDV: Median value of owner-occupied homes in 1000's

B. EDA(Exploratory Data Analysis)

EDA will help us to analyze the affect of the features for output of model. We use Pearson correlation coefficient(Figure 1)on each training data and training target, and choose the features which correlation coefficient is outside [-0.2,0.2].

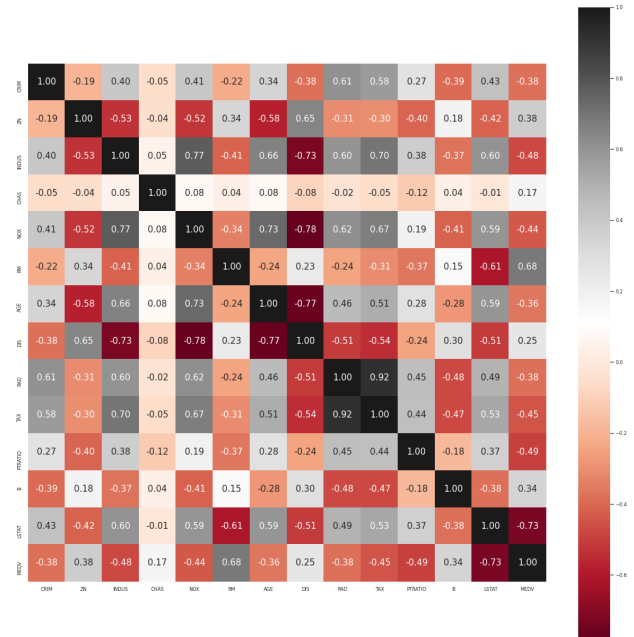


Fig. 1. features which correlation coefficient

From the above heatmap ,we can conclude that, price of MEDV greatly depends upon features RM (positively correlated) and LSTAT (negatively correlated). Also, the features AGE and DIS are negatively correlated with each other. If a house is older then Weighted distances to five Boston employment centers decreases. Similarly, other such pairs are "DIS - NOX", "DIS - INDUS" and "LSTAT - RM". And features TAX and RAD are positively correlated with each other.

C. Build model and training by class

we use python class to build essential method to do MLP, we have relu activation method, dropout method, forward method, compute_cost method, predict method and fit method, and we put backward propagation, gradient descent, and weight initialization in fit method. By the way, the cost of this model is square error, because our model is linear regression, using normal distribution to initialize neural and eta is learning rate.

You can control dropout or Xavier_initialization, by setting parameter True or False(Figure 2). If you don't use Xavier_initialization, the object will use normal distribution to initialize neural.

```
NeuralNetMLP(drop_out = True, Xavier_initialization = True)
```

Fig. 2. some parameter you can adjust

RESULT

We use 150 epochs to fit each model with different parameters(we have four condition of this model, using dropout or Xavier_initialization or not), the following figures are the cost function and scatter plot of y_{true} and y_{pred} . Figure three, without dropout or Xavier_initialization the cost curve is more smooth, but both of their cost will go down when epochs increasing. Figure four, without dropout or Xavier_initialization predict value is more close to true value, the rest of them have some outlier).

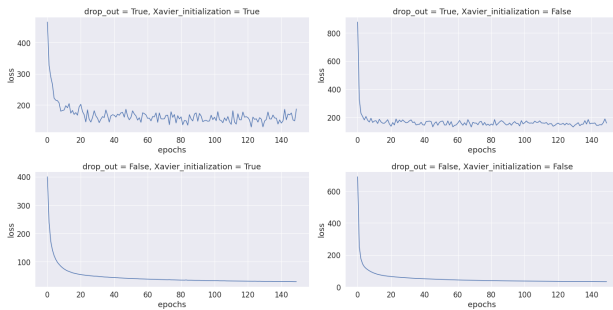


Fig. 3. cost curve of different parameters

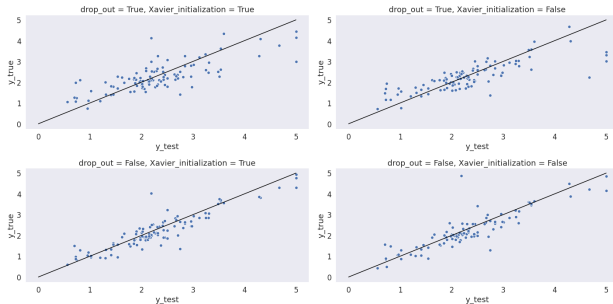


Fig. 4. scatter plot of y_{true} , y_{pred} for different parameters

We use Mean-Squared Error (MSE), R2-score and MAE to validate our accuracy of the prediction. The error is enough small, shows that we have a precise prediction to the data.

dropout or Xavier_initialization	R2 score	MSE	MAE
MLP1(True,True)	0.6	0.33	0.43
MLP2(True,False)	0.63	0.31	0.38
MLP3(False,True)	0.82	0.15	0.28
MLP4(False,False)	0.77	0.19	0.28