Report

2014313303, Taeha Hong

**1. Describe the input and functional structure of the machine learning model for the credit approval data set. (10 points)**

Input data is credit approval data set from UCI data repository. Number of instances is 690. 307 positive data(+), 383 negative data(-). Number of attributes is 15.

I used kears open source to make CNN, RNN, MLP models.

CNN

1 dimensional convolution layer -> max pooling -> batch normalization -> 1-d convolution layer -> max pooling -> batch normalization -> fully connected layer -> batch normalization -> fully connected layer -> batch normalization -> classify(sigmoid)

텍스트이(가) 표시된 사진

자동 생성된 설명

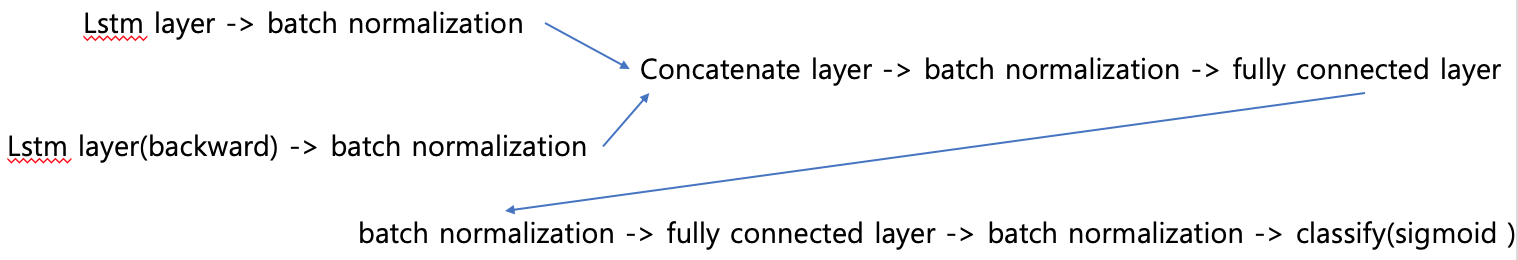
MLP

I stacked fully connected layer and batch normalization 3 times.

스크린샷이(가) 표시된 사진

자동 생성된 설명

RNN(bidirectional)



텍스트이(가) 표시된 사진

자동 생성된 설명

**2. Describe how and why learning parameters were determined for the selected learning model. (10 points)**

Decision Tree :

Max depth : 5 (I test None, 2, 3, 7, 10…. Then max depth 5 has best result.)

MLP:

Activation function : ReLu

Optimizer : Adam (lr=0.001, beta\_1=0.9, beta\_2=0.999, epsilon=1e-08)

Epoch : 30

Random Forest :

Max depth : 10

N\_estimator : 100

AdaBoost (Decision Tree) :

Max depth : 5

Algorithm : SAMME

N\_estimator : 100

CNN :

Activation function : ReLu

Optimizer : Adam (lr=0.001, beta\_1=0.9, beta\_2=0.999, epsilon=1e-08)

Epoch : 30

RNN :

Activation function : ReLu

Optimizer : Adam (lr=0.001, beta\_1=0.9, beta\_2=0.999, epsilon=1e-08)

Epoch : 50

**3. Using the methods of evaluation of the learning models described above, the measures of evaluation for each learning model are summarized. (20 points)**

10-fold evaluation (8 models)







Summary



**4. Compare and analyze the evaluation results of the summarized learning models and explain the reasons for experiment results. (30 points)**

RandomForest > AdaBoost(Decision Tree) > SVM > Decision Tree > RNN > CNN > MLP

AdaBoost boost decision tree 3% accuracy.

Deep Learning model have bad results. I think learning parameter or model structure is not suitable. If I make proper model structure, deep learning models have better results.

**5. Summarize references to the learning models and sources used in this assignment. (10 points)**

I used scikit learn and keras open source.

Data preprocessing

* Encode categorical attribute data to number.
* Split attribute and label.

텍스트이(가) 표시된 사진

자동 생성된 설명

Canonical models

* Decision Tree
* SVM
* MLP

텍스트, 스크린샷이(가) 표시된 사진

자동 생성된 설명

Committee Machines

스크린샷이(가) 표시된 사진

자동 생성된 설명

Deep Learning models

스크린샷, 텍스트이(가) 표시된 사진

자동 생성된 설명

스크린샷, 텍스트이(가) 표시된 사진

자동 생성된 설명

I made RBM models, but it predict all negative labels.

Evaluation methods

텍스트이(가) 표시된 사진

자동 생성된 설명