
PRML Assignment1

Mengyi Chen
CS
ID: 19307110382
19307110382@fudan.edu.cn

Abstract

A brief report for PRML assignment 1

1 Introduction

In this assignment, 4 methods including KNN, Decision-Tree, Softmax and Feedforward Neural Network have been implemented to process the MNIST dataset. The highest accuracy 70% on the test set is achieved by using Softmax.

2 Dataset

The experiment used part of the Fashion MNIST-dataset. There are 14600 samples in the train set and 10000 samples in the test set. The number of sample categories is 10.

3 Methodology

1. KNN

By choosing k-nearest neighbors to decide the final category a sample belongs. Use cross validation for finding best parameters.

2. Decision Tree

Split the dataset when passing the tree node until the data belong to the node has only one category. Parameters are tried by human to find the best ones instead of using cross validation, due to the long running time.

3. Softmax

Labels are preprocessed to one-hot vectors for later utility. Use multi-cross-entropy as loss function and gradient descent to update weights. Stop in advance to avoid overfitting.

4. Feedforward Neural Network

A two layer network. The first layer consists of linear function and LeakyReLU while the second layer consists of linear function and Softmax. Substituting Logistic for LeakyReLU shows no great difference. Results of linear function are normalized to avoid great numbers.

4 Results

Graphics can be seen in assignment1.ipynb

1. KNN

Preprint. Under review.

Select $k = 3$ can achieve an accuracy of 46%. The bigger k is, the lower the accuracy. It takes about half an hour to run KNN.

2. Decision Tree

Set $\text{max_depth} = 100$, $\text{max_features} = 400$ gets an accuracy of 10%. Adjust parameters does not affect accuracy greatly. It takes about one hour to run Decision Tree once.

3. Softmax

Set $\text{epoch} = 3000$ can achieve an accuracy of 70%. Adding more epochs does not increase accuracy.

4. Feedforward Neural Network

It can only achieve 30% accuracy on train set and 10% accuracy on test set. Changing parameters does not affect a lot. There seems to exist some problems with this model but I can't find out.

5 Conclusion

Best performance is achieved by Softmax with an accuracy of 70%. The other three shows poor performance in processing this dataset, probably due to some error in the model.