2017-11-1

Homework 4

By yangfan sun

1. Introduction

Implement logistic regression structure to train a dataset downloaded from internet.

1. Objectives

Introduce the dataset, attach in file box, named as “datatest.txt”, about the condition of houses and its occupancy. Dataset includes six parameters totally, temperature, humidity, light, Co2, the ratio of humidity and occupancy (if “1” is occupied, “0” is not).

1. Approaches

Cross validation

MSE

Logistic regression

Sigmoid activate function

Adam optimizer with 0.001 learning rate

1. Workflow

Apply cross validation to split dataset into training and testing sets, provide placeholder for features and labels, size is [5] and [1] respectively. Then setup initial variable values for each, to product a predict label, activated by sigmoid function. The mean square error will be optimized by adam optimizer (learning rate is 0.001) to converge by iterating 100 times, according to the result.

1. Datasets

Totally 2665 samples splits into training sets (1998 samples) and test sets (667 samples) by cross validation. Features include five parameters, "Temperature", "Humidity", "Light", "CO2", "HumidityRatio". And label is "Occupancy". Assume the rate of occupancy impacts by the several condition provided of rooms.

1. Parameters

Features = 5

Labels = 1

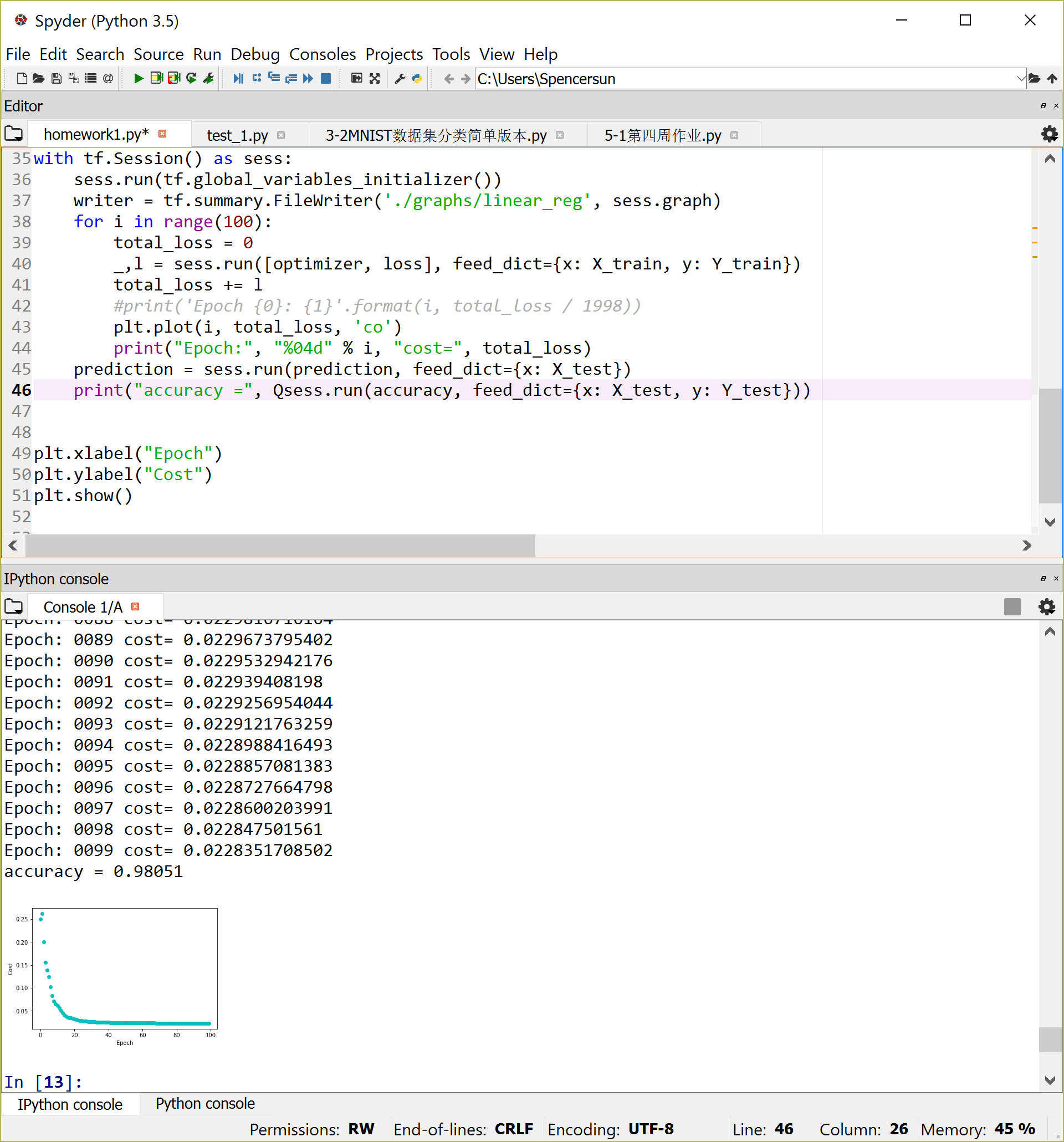
Layer = 1

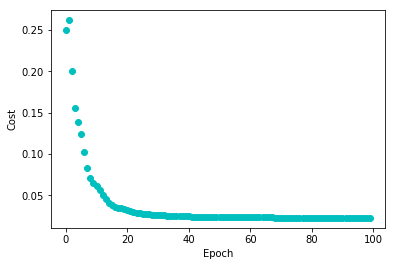
Training set: test set = 0.75

Learning rate = 0.001

Iterarion = 100

1. Evaluation & Discussion





1. Conclusion

According to the result, the accuracy of predicted labels is 98%, and it does not affect to the increasing of iteration due to its training structure. The result will be slightly worse, since apply tanh activation function instead of sigmoid.