COMP 4211 - Machine Learning Programming Assignment 1 Report

Cheng Chi Fung
cfchengac@connect.ust.hk

Abstract

In this assignment, we used linear regression, logistic regression and single layer neural network to perform the regression and classification on three binary classification and regression data sets.

1 Linear Regression

1.1 Experienment Model Settings

For linear regression, we used the built-in models **LinearRegression** in skitlearn to perform the linear regression on three regression datasets .This model fits a linear model with coefficients $w = (w_1, ..., w_p)$ to minimize the residual sum of squares between the observed responses in the dataset, and the responses predicted by the linear approximation. This model has **no hyperparameter** to set.

Table 1: Experiments Datasets Settings

Dataset	Training set	Test set
Fifa	13191	4397
Finance	2754	918
Orbits	9642	3215

1.2 Experiment Results

The following are the experiment results which using the above model setting on fifa, finanace and orbits regression datasets.

Table 2: Experiments Results of Linear Regression on three Regression Datasets

Sum of Squared Error (Mean)		
Dataset	Training set	Test set
Fifa	1583.01911 (0.12000)	535.36126 (0.12175)
Finance	394.91758 (0.14339)	163.38803 (0.17798)
Orbits	917.91272 (0.09519)	311.47817 (0.09688)

From the experiment results, we found out that excepts the **finance** datasets, the mean squared error of the training set and test set are almost the same. And, for the finance datasets, the mean squared error of the test set (0.17798) are a much higher than the training set (0.14339). It means that this model may **overrfit** the training data of the finance datasets.

2 Logistic Regression

2.1 Experienmental Model Settings

To perform the logistic regression with gradient-descent algorithm by minimizing the cross-entropy loss on the three classification datasets, we used the built-in model **SGDClassifier** in skitlearn. This model implements regularized linear models with **stochastic gradient descent** (SGD) learning.

To get better training results, we performed the **hyperparameters tunning** by using **GridSearchCV** in skitlearn to obtain best hyperparameters parameters for each datasets. The following are the settings of parameters tuning.

Table 3: Parameters Tuning Settings (Tuned hyperparameters *)

Name	Parameter Setttings
max_iter	5000
tol	0.000000001
learning_rate*	$0.1, 0.01, 0.001$
number of cross folds	5

Table 4: The results of the Hyperparameters Tuning

Name	Parameter Setttings
max_iter	5000
tol	0.000000001
learning_rate*	$0.1, 0.01, 0.001$
number of cross folds	5

2.2 Experiment Results

The following are the experiment results which using the above model setting on fifa, finanace and orbits classification datasets.

3 Sinlge Layer Neural Network

3.1 Experienmental Model Settings

In this assignment, we adopted **Logistic Regression** as the prediction model. The formula we used in our prediction model were the followings.

3.2 Experienment Results