

The Experiment Report of Machine Learning

SCHOOL: SHIEN-MING WU SCHOOL OF

INTELLIGENT ENGINEERING

SUBJECT: The super robot Everest class

|  |  |
| --- | --- |
| *Author:*  Xilang Zeng | *Supervisor:*  Mingkui Tan |
|  |  |
| *Student ID:*  202130461984 | *Grade:*  Undergraduate |

2024-3-28

# Experiment 1: Linear Regression and Stochastic Gradient Descent

***Abstract—This experiment involves using linear regression with the Housing dataset from LIBSVM Data. The parameters of the linear regression are optimized using both the closed-form solution and the stochastic gradient descent method for training. Then, the trained linear regression model's prediction accuracy is validated using a validation set.*** ***Additionally, an application experiment was conducted using a student grades dataset, where the grades from all courses were used to predict the GPA in a machine learning course.***

## INTRODUCTION

To predict the Housing data from LIBSVM Data, this experiment utilizes a linear regression model, experimenting with and comparing two methods: the closed-form solution and stochastic gradient descent. The effectiveness of the two methods and the impact of different parameters are compared.

Since linear regression has a closed-form solution, it's clear that the stochastic gradient descent method does not hold an advantage in terms of computational speed.

To test the predictive capabilities of linear regression, an experiment was also conducted using a student grades dataset. This experiment aimed at predicting the GPA in a machine learning course based on the grades from various courses. Due to the lack of nonlinearity in linear regression, there are substantial differences in its predictive performance across different datasets.

## METHODS AND THEORY

*In this section, you are asked to give a complete introduction to the experiment. For instance, the chosen methods, the related theories, the related equations (loss function), the derivation process (taking the gradient) and so on.*

1. Reading the experimental data

The Housing dataset stored in LIBSVM format is loaded using the Python library sklearn, and it is divided into training and validation sets through the sklearn library. The validation set accounts for 10% of the original dataset, and the data is randomized before the split. After reading the data, preprocessing is performed to convert the sparse matrix into a dense array format, and the target vectors y\_train and y\_valid are transformed from one-dimensional arrays into two-dimensional arrays.

1. Splitting into training and validation sets

The training set is the dataset used to train the model. The model learns and establishes rules for prediction or classification from this part of the data. The training set constitutes a major portion of the entire dataset. The validation set helps us understand the model's performance on unknown data and is used to avoid overfitting of the model. It usually constitutes a smaller proportion of the entire dataset. In this experiment, we divided the dataset into 90% for the training set and 10% for the validation set. Before the division, we also randomized the order of the dataset's data.

1. Choosing the loss function
2. Computing model parameters
3. Validating with the validation set

### Printing the loss curve

## EXPERIMENT

### Dataset

This section represents the related information of datasets, such as the content, the number of data, the training set, the validation set and so on.

### Implementation

All detailed implementation in your experiment: initialization, process, results, all kinds of parameters. In a word, describe clearly What you do and how you do. Figures and tables should be labeled and numbered, such as in Table I and Fig. 1.

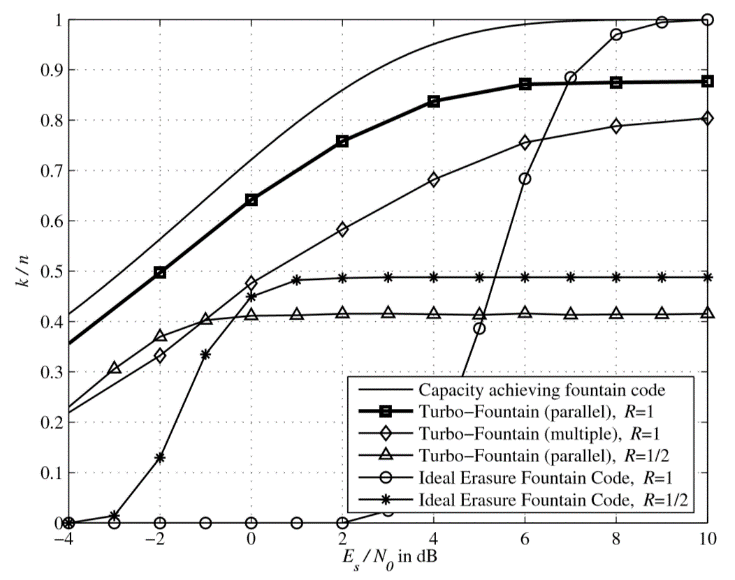
TABLE I

SIMULATION PARAMETERS

|  |  |
| --- | --- |
| Information message length | *k* = 16000 bit |
| Radio segment size | *b* = 160 bit |
| Rate of component codes | R = 1/3 |

## CONCLUSION

This section summarizes the paper. In our experiments, you can also write your gains and inspirations in here.

Figure. 1. Simulation results on the AWGN channel.