

***PROJECT REPORT***

***MACHINE LEARNING***

BY **YASH SINGLA**

***PROJECT REPORT***

***MACHINE LEARNING***

|  |  |
| --- | --- |
| **SERIAL NO** | **LIST OF CONTENTS** |
| 1 | INTRODUCTION |
| 2 | ACCURACY |
| 3 | TRAINING LOGS |

**INTRODUCTION**

This project mainly consists of machine learning library built from scratch using python and implementing same on various datasets provided by the user.

This project consists of the listed methods:

* Linear regression
* Polynomial regression
* Logistic regression
* KNN
* K means clustering
* Neural networks

Datasets that are used are:

* Linear\_train
* Linear\_test
* Polynomial\_train
* Polynomial\_test
* emnist-letters-train
* emnist-letters-test

Libraries used in this project are:

* pandas
* numpy
* matplotlib

**ACCURACY**

|  |
| --- |
| LINEAR REGRESSION |
| RMSE value is 1.2131594907485281 |

|  |
| --- |
| POLYNOMIAL REGRESSION |
| RMSE value is 1.2648596319201988 |

|  |
| --- |
| LOGISTIC REGRESSION |
| 68.0558449677016 % of the predicted alphabets are matching with the given results |

|  |
| --- |
| KNN |
| 85.6 % of the predicted alphabets are  matching with the given results |

|  |
| --- |
| K MEANS CLUSTERING |
| 61.51281516982705 % of the predicted  alphabets are matching with the given results |

|  |
| --- |
| NEURAL NETWORKS |
| 76.4242840572529 % of the predicted  alphabets are matching with the given results |

**LINEAR REGRESSION**

* Used Normalization for training data and used the same mean and deviation value to scale test data.
* For cost function I took alpha=0.1 and no of iterations =140
* Used matplotlib to plot graph between no of iterations (x axis) and cost function(J) (y axis) to show the decreasing trend.

**POLYNOMIAL REGRESSION**

* Used Normalization for training data and used the same mean and deviation value to scale test data.
* The degree of the polynomial was chosen to be 20 in order to achieve high accuracy.
* For cost function I took alpha=0.01 and no of iterations =5000
* Used matplotlib to plot the predicted values with the given values with reference to input values

**LOGISTIC REGRESSION**

* Used scaling to make all the values to be between 0 and 1.
* For cost function I took alpha=0.003 and no of iterations =1000
* Used sigmoid function for cost function.
* Calculated the percentage of correctly predicted values to the total values for accuracy check.

**KNN**

* Used scaling to make all the values to be between 0 and 1.
* Euclidian distance is used to calculate the distances between the two set of values .
* Value of k was chosen to be 350.
* Calculated the percentage of correctly predicted values to the total values for accuracy check.

**K MEANS CLUSTERING**

* Used scaling to make all the values to be between 0 and 1.
* Used the mean value to obtain centroid for cluster.
* Calculated the percentage of correctly predicted values to the total values for accuracy check.

**NEURAL NETWORK**

* Used scaling to make all the values to be between 0 and 1.
* For cost function I took alpha=0.001 and no of iterations =500
* Calculated the percentage of correctly predicted values to the total values for accuracy check.