PREDICTION OF ALPHABET MATRIX THROUGH SUPPORT VECTOR MACHINE (S.V.M.)

[1]Mohd Hammad, [2] Nirmal Singh, [3]Manish Kumar, [4]Sanjeev Kumar
[1]Student, [2]Student, [3]Student, [4]Assistant Professor
Department of Computer Science & Engineering,
ABES Institute of Technology, Ghaziabad, Uttar Pradesh, India.

I. ABSTRACT:

We administer the Prediction of alphabet matrix through Support Vector Machine to increase the precision of each letters. We find prediction matrix on 20,000 datasets. This fined how many mismatch of words and for this we create a prediction matrix after machine learning. For machine learning we use SVM. The aim of a SVM is to construct a flat boundary known as hyper plane, which separate the space to construct fairly homogeneous partitions on either side.Goal of the Prediction Matrix Alphabet is to increase or maximize the accuracy of the mismatch Alphabets. As for example, in word RAM the precision of the letter R is to be increased by Prediction Matrix.

II. KEYWORDS:

Support Vector Machine, Prediction Matrix Alphabet Dataset, and programing language Python.

III. INTRODUCTION

Prediction Matrix Alphabetical is used to examine how well our classifier performed, we need to compare the predicted letter to the true letter in the testing dataset.

The predict() function allows us to use the letter classification model to make predictions on the testing dataset.

- Using the head() function,
- we can see that the first six predicted letters were U, N, V, X, N, and H:
- > head(letter predictions)
- [1] U N V X N H
- Levels:
- A B C D E F G H I J K L M N O P Q R S T U V W X Y Z

letter_predictions	A	B	C	D	E
A	144	0	0	0	0
B	0	121	0	5	2
C	0	0	120	0	4
D	2	2	0	156	0
E	0	0	5	0	127

- The diagonal values of 144, 121, 120, 156, and 127 indicate the total number of records where the predicted letter matches the true value. Similarly, the number of mistakes is also listed. For example, the value of 5 in row B and column D indicates that there were five cases where the letter D was misidentified as a B.

- Looking at each type of mistake individually may reveal some interesting patterns about the specific types of letters the model has trouble with, but this is time consuming. We can simplify our evaluation instead by calculating the overall accuracy. This considers only whether the prediction was correct or incorrect, and ignores the type of error.

IV. PROPOSED SYSTEM

The methodology is based on the following steps:

Step 1: The dataset which is provided should be separated into two parts i.e., 90% training set and 10% testing set depending on the 10-fold cross validation strategy.

Step 2: SVM classifier is established one by one using the dissimilar kind of kernel functions (linear, polynomial etc).

Step 3: SVM classifier ensembles are established by trapping and encouraging to outcome linear, polynomial SVM ensembles.

Step 4: Before examining the accuracy of the classifier, the testing set is delivered into the established classifier.

Step 5: In order to evaluate the computational complexities of training dissimilar classifiers, the classifiers training times are measured.

V. RESULT AND DISCUSSION

The dataset considered is as follows:

T,2,8,3,5,1,8,13,0,6,6,10,8,0,8,0,8

I,5,12,3,7,2,10,5,5,4,13,3,9,2,8,4,10

D,4,11,6,8,6,10,6,2,6,10,3,7,3,7,3,9

N,7,11,6,6,3,5,9,4,6,4,4,10,6,10,2,8

G,2,1,3,1,1,8,6,6,6,6,5,9,1,7,5,10

S,4,11,5,8,3,8,8,6,9,5,6,6,0,8,9,7

B,4,2,5,4,4,8,7,6,6,7,6,6,2,8,7,10

A,1,1,3,2,1,8,2,2,2,8,2,8,1,6,2,7

J,2,2,4,4,2,10,6,2,6,12,4,8,1,6,1,7

M,11,15,13,9,7,13,2,6,2,12,1,9,8,1,1,8

X,3,9,5,7,4,8,7,3,8,5,6,8,2,8,6,7

O,6,13,4,7,4,6,7,6,3,10,7,9,5,9,5,8

G,4,9,6,7,6,7,8,6,2,6,5,11,4,8,7,8

M,6,9,8,6,9,7,8,6,5,7,5,8,8,9,8,6

R,5,9,5,7,6,6,11,7,3,7,3,9,2,7,5,11

F,6,9,5,4,3,10,6,3,5,10,5,7,3,9,6,9

O,3,4,4,3,2,8,7,7,5,7,6,8,2,8,3,8

C,7,10,5,5,2,6,8,6,8,11,7,11,2,8,5,9

T,6,11,6,8,5,6,11,5,6,11,9,4,3,12,2,4

J,2,2,3,3,1,10,6,3,6,12,4,9,0,7,1,7

J,1,3,2,2,1,8,8,2,5,14,5,8,0,7,0,7

H,4,5,5,4,4,7,7,6,6,7,6,8,3,8,3,8

S,3,2,3,3,2,8,8,7,5,7,5,7,2,8,9,8

O,6,11,7,8,5,7,6,9,6,7,5,9,4,8,5,5

J,3,6,4,4,2,6,6,4,4,14,8,12,1,6,1,6

C,6,11,7,8,3,7,8,7,11,4,7,14,1,7,4,8

M,7,11,11,8,9,3,8,4,5,10,11,10,10,9,5,7

W,12,14,12,8,5,9,10,4,3,5,10,7,10,12,2,6

H,6,9,8,7,6,8,6,6,7,7,7,9,6,8,4,8

G,3,6,4,4,2,6,6,5,5,6,6,9,2,8,4,8

L,2,3,3,4,1,0,1,5,6,0,0,6,0,8,0,8

L,1,3,3,1,1,6,4,1,7,8,3,10,0,7,2,9

X,8,12,8,6,4,3,10,4,7,12,11,9,3,7,3,4

B,5,9,7,7,10,9,8,4,4,6,8,6,6,11,8,7

M,6,9,9,7,6,5,6,3,5,10,9,9,8,5,2,7

G,4,7,6,5,3,6,6,6,8,6,5,9,3,10,4,8

O,4,7,5,5,3,7,7,8,6,7,6,8,3,8,3,8

P,3,6,4,4,2,4,14,8,1,11,6,3,0,10,4,8

G,4,9,5,6,6,8,5,4,3,7,5,11,6,8,5,11

E,3,4,3,6,2,3,8,6,10,7,6,15,0,8,7,8

X,5,11,8,8,4,8,8,1,8,10,5,7,3,8,4,8

The result estimated in this project is as follows:

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
A	205	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	
B	0	170	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	1	0	0

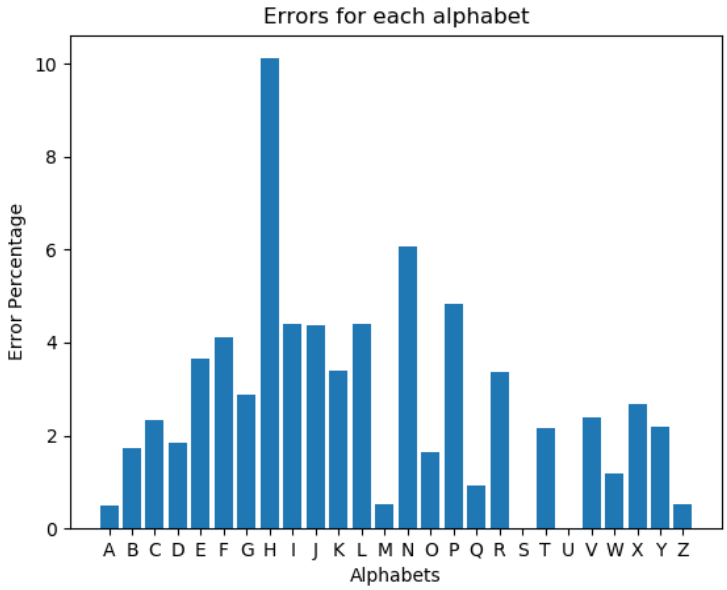
C	0	0	167	0	2	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	1	0	0	0
D	0	1	0	212	0	0	0	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
E	0	0	1	0	184	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	4
F	0	0	0	0	2	186	0	0	0	0	0	0	0	1	0	4	0	0	0	0	0	1	0	0	0	0
G	0	0	1	2	1	0	202	0	0	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0
H	0	1	0	6	1	0	2	160	0	0	2	0	1	0	0	0	1	4	0	0	0	0	0	0	0	0
I	0	0	0	0	0	0	0	0	196	8	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
J	0	0	0	0	1	0	0	0	5	175	0	0	0	0	0	0	1	0	0	0	1	0	0	0	0	0
K	0	0	0	0	0	0	0	2	0	0	171	0	0	0	0	0	0	1	0	0	0	0	0	3	0	0
L	0	1	2	0	0	0	0	2	0	1	0	196	0	0	0	0	0	0	0	0	0	0	0	3	0	0
M	0	1	0	0	0	0	0	0	0	0	0	0	186	0	0	0	0	0	0	0	0	0	0	0	0	0
N	0	0	0	0	0	0	0	4	0	0	0	0	2	186	1	0	0	1	0	0	0	3	1	0	0	0
O	0	0	1	0	0	0	0	0	0	0	0	0	0	0	178	0	1	0	0	0	0	0	1	0	0	0
P	0	1	0	0	0	4	0	3	0	0	0	0	0	0	0	197	2	0	0	0	0	0	0	0	0	0
Q	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	215	0	0	0	0	0	0	0	0	0
R	0	2	0	0	0	0	0	1	0	0	3	0	0	1	0	0	0	201	0	0	0	0	0	0	0	0
S	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	198	0	0	0	0	0	0	0
T	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	180	0	0	0	0	1	0
U	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	215	0	0	0	0	0
V	0	4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	164	0	0	0	0
W	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	165	0	0	0
X	0	1	0	1	0	0	0	0	0	0	3	0	0	0	0	0	0	0	0	0	0	0	0	181	0	0
Y	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	1	179	0
Z	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	193

In the past, the research for the article thorough check the capability of many alternatives of Holland-style adaptive classifier systems to acquire information to correctly estimate the letter categories related to vectors of 16 integer attributes drawn out from raster scan images of the letters. The ultimate accuracy acquired was a little over 80%.

Here we have tried increasing this accuracy of the mismatching alphabets with the help of SVM.

The accuracy of this model which is calculated in this project is 97.24000000000001...

GRAPH FOR THE RESULT:



VI. CONCLUSION

We have taken the dataset of alphabets in order to increase the accuracy of mismatched alphabets. In the past, the researchers estimated the accuracy of this data which was accurate but could also be increased using SVM. Now, here in this project we have tried to increase the accuracy of this data and were successful doing this. The accuracy of the mismatched data is increased as earlier it was 80% and now it is 97.24 approximately. In future, this accuracy can also be increased by using some more better technologies of machine learning.

VII. REFERENCES

1. David Crockett, Ryan Johnson and Brian Eliason, , "What is Data Mining in Healthcare?" *Health Catalyst*, 2004.
2. Brewer, W. F. Is reading a letter-by-letter process? In J. F. Kavanaugh & L G. Mattingly (Eds.).Language by ear and by eye. Cambridge, Mass: M.LT Press, 1972. pp. 359–365.
3. Cortes, C.; Vapnik, V. (1995). "Support-vector networks". Machine Learning. doi:10.1007/BF00994018.
4. Brett Lantz.; Machine Learning with R Second Edition "support vector machine". Machine Learning.
5. Andreas C. Mueller and Sarah Guido.; Machine Learning with Python "support vector machine". Machine Learning
6. DB Statlog database: http://archive.ics.uci.edu/ml/machinelearning databases/statlog/heart