CATALOGUE

1.GENERAL 1

2.DATA PROCESSING 2

2.1. processing of ICI data 2

2.2. processing of FACTORY data 2

2.2.1 process of output columns 2

2.2.2 Discard invalid data 3

2.2.3 Replace the same data 3

2.2.4 Transform data properly 3

3.CLASSIFIER CHOICE 4

3.1 KNN 4

3.2 SVM 4

3.3 MaxEnt 4

4.REGRESSION EXPERIMENT 5

4.1 results on ICI data 5

4.2 results on FACTORY data 7

5.CLASSIFICATION EXPERIMENT 18

5.1 classification experiment on ICI data 19

5.2 classification experiment on FACTORY data 21

6.PARAMETER OPTIMIZATION 31

7.CONCLUSION 32

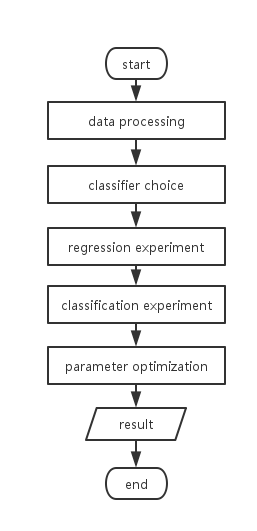
8.REFERENCE 33

# 1.GENERAL

According to the format of the data as well as the style of the result, we got following processing steps as follows.

1. Process the data properly to meet the qualification of the classifier.
2. Do the regression experiment using some kind classifier, and check whether this classifier is capable of processing the given data.
3. Using the classifier mentioned above to do the classification experiment and get the result.
4. Modify the parameters of the classifier to get a better result.

Here is the flow chart.



We cleaned and filtered the whole dataset, then used a small subset of data to test the results of three candidate classifiers. Based on these results and empirical methods, we selected the best classifier – SVM. Then we conducted regression and classification experiment to get the feature combinations, and select one with greatest impact on the result. Finally, we reproduced and optimized the result in R.

# 2.DATA PROCESSING

## 2.1. processing of ICI data

There were 34 columns and 1941 rows in the ICI data, which containing former 27 columns as input columns and latter 7 columns as output columns. The data in the input columns were complete and with proper format, so that the classifier could use them straightly. Yet there were 7 output columns, and the result of each column could be 0 or 1, so there were 2^7=128 kinds of result in total in theory. But we found that in fact there was and only one column turned out to be 1 in seven columns and the rest six columns turned out to be 0. So there were 7 kinds of result finally. After the processing, there was only 1 output column with 7 kinds, which were 1,2,3,4,5,6,7, respectively.

## 2.2. processing of FACTORY data

There were 215 columns and 1000 rows in the factory data, which containing 23rd to 26th columns as input columns and the rest as output columns. There existed problems of incompletion of data and incapability of processing. The detailed processing was as follows.

### 2.2.1 process of output columns

The 23rd column was the result (code) of first inspection and the 24th column was the description in Japanese of the 23rd column. The 25th column was the result (code) of the final inspection and the 26th column was the description in Japanese of the 25th column. We didn’t need the result of the first inspection as the final output, so we could discard column 23, 24 and 26 as the result description was useless in our experiment.

### 2.2.2 Discard invalid data

There were some invalid rows in that they lack of data of some columns, which was marked by ‘NA’. As these rows of data was of no use in our regression nor classification experiment, we just discarded these data, about 90 rows.

### 2.2.3 Replace the same data

There existed some columns in which the data is totally identical to each other. These data were useless in our processing of data because they could not help us find the separation line/hyper plane during the classification experiment. So, all of the data in these columns were replaced with ‘0’ to get rid of the disturb, about 12 columns. What’s more, these data were in the same columns where they used to be, so that we could find the position of the columns according to the initial data straightly rather than counting from the beginning clumsily.

### 2.2.4 Transform data properly

Data in some columns were hard to process during the regression/classification experiment, in that they were not real numbers. On this condition, we handled the data as follows.

1. There existed letters/words in some column, and we just regarded these columns as invalid columns. They usually were in shape of description, date, etc. They were replaced with ‘0’ or ‘1’ simply. For convenience, the serial numbers of columns will be showed in an excel. They were column X, Z, AL, AM, AR, AS, AW, AX, BM, EQ, FA, GG.
2. Data in some columns appeared with some symbols, such as ‘-’ or ‘\_’. In order to get the real numbers as the proper data, we just striped or replaced these symbols or just retain some number in given positions to transform the data into the real number. They were column G, AO, AU, AY, EP, EZ, GF.
3. Although there were some data in the shape of time or date, they were of some use as a matter of experience. Data in these columns were transformed into a integer to show the time. They were column AJ and AK.
4. In some columns, there existed only two kinds of data, but they were very different with each other and they were too big. For convenience, we transformed one of these two kind of data into ‘1’ and the other was ‘0’. They are column S, FU, FW, FX.

# 3.CLASSIFIER CHOICE

There were 3 classifiers in our plan to choose, which were SVM, MaxEnt and KNN. Each one has its own advantages and shortcomings.

## 3.1 KNN

KNN is a kind of unsupervised learning method. It just lumps the neighbor points together to form a class after iteration.

1. simple and effICIent
2. low cost to re-train
3. the time and space complexity is linear with the scale of the corpus
4. if there exist many overlaps and intersections among the classes, KNN will function better, in that KNN classifies the data according to the position of each data point in the hyper space rather than the kinds of classes.
5. KNN is good at processing big corpus
6. KNN is a lazy learning method, thus it’s slower than most methods
7. category score is not standardized
8. hard to interpret the output
9. large calculation

## 3.2 SVM

SVM (reference [2]) is a mature classifier extensively applied on industries.

1. SVM is good at processing small corpus
2. improve the generalization performance
3. solve the high-dimensional problem
4. solve the non-linear problem
5. avoid the problem of the neural network structure choice and local minima
6. sensitive to the loss of data
7. no unified solution to the non-linear problem

## 3.3 MaxEnt

MaxEnt is a classifier which has good performance in some fields, especially NLP

1. just focus on the choice of features, rather than the combination of features.
2. the parameters can be smoothed without regular smoothing algorithms
3. fast, effICIent, result is of good persuasion

After conducting some experiments with small corpus, we decided to use the SVM to do the whole job. The reasons are as follows.

MaxEnt is good at processing linguistic segments such as sentences, words, rather than real numbers. KNN is a unsupervised classifying method, so it could be disturbed by the large scale of the features. The number of rows of data was about 1000 and 2000, respectively, which were quite small, while the number of features are fairly large. And empirically speaking, this kind of classifying work was usually done by SVM, so we pick it.

# 4.REGRESSION EXPERIMENT

There are different kinds of kernels of SVM, and we used epsilon-SVR and nu-SVR respectively to conduct the regression experiment. The basic feature choosing strategy is a kind of greedy algorithm. On one hand we train the corpus with merely one input column (which is feature, the same below) and get the result (training #1), thus the better the result was, the feature the more important was. On the other hand we train the corpus using the whole feature combination but the chosen one below (training #2), thus the worse the result was, the feature the more important was.

## 4.1 results on ICI data

Here are two tables using different type of kernel, showing the result of regression experiment which took the raw dataset as input. They illustrated the capability of being separated correctly by the classifier of each feature.

Attention: epsilon stood for the kernel type, ‘single’ stood for training #1 while ‘reverse’ stood for training #2, which are illustrated in the paragraph below.

MSR = Mean squared error, the smaller the better.

SCC = Squared correlation coefficient, the closer to 1 the better.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | ICI |  |  |  |
| kernel | epsilon-SVR |  |  |  |
|  | single |  | reverse |  |
| features | MSR | SCC | MSR | SCC |
| A | 4.56917 | 0.126794 | 5.03551 | 0.000500628 |
| B | 4.40277 | 0.130173 | 5.03551 | 0.000500628 |
| C | 5.0354 | 2.53E-05 | 5.03596 | 0.000817621 |
| D | 5.03165 | 4.84E-05 | 5.03596 | 0.000817621 |
| E | 4.63477 | 0.0600243 | 5.03596 | 0.000817621 |
| F | 4.68938 | 0.0893899 | 5.03596 | 0.000817621 |
| G | 4.67362 | 0.0968772 | 5.03596 | 0.000817621 |
| H | 5.01336 | 0.00323708 | 5.03596 | 0.000817621 |
| I | 4.77613 | 0.0960175 | 5.03596 | 0.000817621 |
| J | 5.73515 | 0.00275156 | 5.03596 | 0.000817621 |
| K | 4.14202 | 0.216261 | 5.03593 | 0.000798755 |
| L | 5.10902 | 0.0071919 | 5.03596 | 0.000817621 |
| M | 5.10902 | 0.0071919 | 5.03596 | 0.000817621 |
| N | 5.2832 | 0.0926946 | 5.03596 | 0.000817621 |
| O | 4.7446 | 0.10349 | 5.03596 | 0.000817621 |
| P | 5.68306 | 0.00391798 | 5.03596 | 0.000817621 |
| Q | 5.21392 | 0.0121271 | 5.03596 | 0.000817621 |
| R | 5.02417 | 0.0663145 | 5.03596 | 0.000817621 |
| S | 4.97051 | 0.0380861 | 5.03596 | 0.000817621 |
| T | 5.30938 | 0.0242209 | 5.03596 | 0.000817621 |
| U | 5.59027 | 0.00051342 | 5.03596 | 0.000817621 |
| V | 4.43212 | 0.131264 | 5.03596 | 0.000817621 |
| W | 4.61665 | 0.114962 | 5.03596 | 0.000817621 |
| X | 4.37359 | 0.113722 | 5.03596 | 0.000817621 |
| Y | 5.06599 | 0.011091 | 5.03596 | 0.000817621 |
| Z | 5.73728 | 0.000891975 | 5.03596 | 0.000817621 |
| AB | 4.95343 | 0.0679797 | 5.03596 | 0.000817621 |

table 1. regression on ICI data using epsilon-SVR

From the table above we could get some important features, which had lower MSR, say, less than 5.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | ICI |  |  |  |
| kernel | nu-SVR |  |  |  |
|  | single |  | reverse |  |
| features | MSR | SCC | MSR | SCC |
| A | 4.20752 | 0.0981131 | 4.59624 | 6.86E-05 |
| B | 4.13051 | 0.105006 | 4.59624 | 6.86E-05 |
| C | 4.59562 | 6.72E-05 | 4.59673 | 0.000592669 |
| D | 4.59392 | 0.000535612 | 4.59673 | 0.000592669 |
| E | 4.53098 | 0.0234467 | 4.59673 | 0.000592669 |
| F | 4.49199 | 0.0444134 | 4.59673 | 0.000592669 |
| G | 4.39694 | 0.0590242 | 4.59673 | 0.000592669 |
| H | 4.5807 | 0.00323949 | 4.59673 | 0.000592669 |
| I | 4.28665 | 0.0758236 | 4.59673 | 0.000592669 |
| J | 4.67586 | 0.00014023 | 4.59673 | 0.000592669 |
| K | 3.95161 | 0.192411 | 4.5967 | 0.000548998 |
| L | 4.6203 | 0 | 4.59673 | 0.000592669 |
| M | 4.6203 | 0 | 4.59673 | 0.000592669 |
| N | 4.1051 | 0.108153 | 4.59673 | 0.000592669 |
| O | 4.23493 | 0.102119 | 4.59673 | 0.000592669 |
| P | 4.62021 | 0.00370721 | 4.59673 | 0.000592669 |
| Q | 4.62025 | 0.00147174 | 4.59673 | 0.000592669 |
| R | 4.60808 | 0.0395117 | 4.59673 | 0.000592669 |
| S | 4.61987 | 0.00755841 | 4.59673 | 0.000592669 |
| T | 4.62001 | 0.0363257 | 4.59673 | 0.000592669 |
| U | 4.62031 | 9.57E-06 | 4.59673 | 0.000592669 |
| V | 4.39763 | 0.132478 | 4.59673 | 0.000592669 |
| W | 4.39611 | 0.102129 | 4.59673 | 0.000592669 |
| X | 4.20125 | 0.116075 | 4.59673 | 0.000592669 |
| Y | 4.61082 | 0.00408828 | 4.59673 | 0.000592669 |
| Z | 4.62025 | 0.00281365 | 4.59673 | 0.000592669 |
| AB | 4.5207 | 0.0290069 | 4.59673 | 0.000592669 |

table 2. regression on ICI data using nu-SVR

The same as above, we could get some important features via different kernel, say, have MSR closer to 4.

## 4.2 results on FACTORY data

The explanation of the titles was as same as the ICI part.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | ICI |  |  |  |
| kernel | epsilon-SVR |  |  |  |
|  | single |  | reverse |  |
| features | MSR | SCC | MSR | SCC |
| A | 0.159697 | 0.469481 | 0.282314 | 0.000411919 |
| B | 0.196744 | 0.479791 | 0.282443 | 1.60E-05 |
| C | 0.284211 | 5.22E-16 | 0.28219 | 0.000537251 |
| D | 0.284211 | 5.22E-16 | 0.28219 | 0.000537251 |
| E | 0.238032 | 0.475612 | 0.282699 | 9.24E-05 |
| F | 0.158856 | 0.473275 | 0.282234 | 0.000715927 |
| G | 1568.8 | 0.0183575 | 0.282557 | 0.000161982 |
| H | 0.284211 | 5.22E-16 | 0.282781 | 6.97E-05 |
| I | 0.284211 | 5.22E-16 | 0.282786 | 9.54E-05 |
| J | 0.284211 | 2.36E-16 | 0.28219 | 0.000537251 |
| K | 0.158877 | 0.47322 | 0.282182 | 0.000676436 |
| L | 0.284211 | 5.22E-16 | 0.282607 | 0.000190181 |
| M | 0.284211 | 1.45E-16 | 0.28219 | 0.000537251 |
| N | 0.284211 | 5.22E-16 | 0.28219 | 0.000537251 |
| O | 0.284211 | 5.22E-16 | 0.28219 | 0.000537251 |
| P | 0.284211 | 5.22E-16 | 0.28219 | 0.000537251 |
| Q | 0.284211 | 5.22E-16 | 0.28219 | 0.000537251 |
| R | 0.202111 | 0.390664 | 0.282594 | 0.000215205 |
| S | 0.284211 | 4.78E-16 | 0.28219 | 0.000537251 |
| T | 54.9084 | 0.0157187 | 0.282487 | 7.84E-05 |
| U | 0.284211 | 5.22E-16 | 0.28219 | 0.000537251 |
| V | 0.284211 | 5.22E-16 | 0.28219 | 0.000537251 |
| W | 0.295223 | 0.00273253 | 0.282322 | 0.000552984 |
| X | 0.284211 | 5.22E-16 | 0.28219 | 0.000537251 |
| Y | 0.284211 | 5.22E-16 | 0.28219 | 0.000537251 |
| Z | 0.284211 | 5.22E-16 | 0.28219 | 0.000537251 |
| AA | 0.284211 | 5.22E-16 | 0.28219 | 0.000537251 |
| AB | 0.284211 | 5.22E-16 | 0.28219 | 0.000537251 |
| AC | 0.284211 | 8.11E-16 | 0.28219 | 0.000537644 |
| AD | 0.284211 | 8.11E-16 | 0.28219 | 0.000537644 |
| AE | 0.287457 | 0.000379279 | 0.282188 | 0.000582952 |
| AF | 0.237178 | 0.479198 | 0.28256 | 5.18E-05 |
| AG | 0.0110247 | 0.996149 | 0.28219 | 0.000537235 |
| AH | 0.284211 | 5.22E-16 | 0.28219 | 0.000537251 |
| AI | 0.284211 | 5.22E-16 | 0.28219 | 0.000537251 |
| AJ | 0.298676 | 0.00114905 | 0.282254 | 0.000362535 |
| AK | 0.159226 | 0.438839 | 0.282533 | 0.00024994 |
| AL | 0.284211 | 5.22E-16 | 0.28219 | 0.000537251 |
| AM | 0.284211 | 5.22E-16 | 0.28219 | 0.000537251 |
| AN | 0.284211 | 8.58E-16 | 0.282188 | 0.00054709 |
| AO | 288.022 | 0.26323 | 0.282454 | 0.000162078 |
| AP | 0.301977 | 0.000159244 | 0.282165 | 0.000881181 |
| AQ | 0.237191 | 0.479195 | 0.282678 | 0.000190934 |
| AR | 0.284211 | 5.22E-16 | 0.28219 | 0.000537251 |
| AS | 0.284211 | 5.22E-16 | 0.28219 | 0.000537251 |
| AT | 0.284211 | 1.47E-16 | 0.28219 | 0.000537251 |
| AU | 11.2039 | 0.0494116 | 0.282211 | 0.000291396 |
| AV | 0.284211 | 5.22E-16 | 0.282164 | 0.000884521 |
| AW | 0.284211 | 5.22E-16 | 0.28219 | 0.000537251 |
| AX | 0.284211 | 5.22E-16 | 0.28219 | 0.000537251 |
| AY | 1568.8 | 0.0183575 | 0.282529 | 9.18E-05 |
| AZ | 0.306307 | 0.000723236 | 0.282483 | 0.00017234 |
| BA | 0.237152 | 0.4792 | 0.282671 | 0.000116043 |
| BB | 0.284211 | 2.75E-16 | 0.282188 | 0.000546655 |
| BC | 0.284211 | 1.73E-16 | 0.282188 | 0.000546655 |
| BD | 0.284126 | 0.0145098 | 0.28219 | 0.00053785 |
| BE | 0.283932 | 0.0230088 | 0.282187 | 0.000549247 |
| BF | 0.284015 | 0.0158456 | 0.282187 | 0.000551844 |
| BG | 0.284066 | 0.0669828 | 0.282475 | 0.000268078 |
| BH | 0.234556 | 0.18169 | 0.281938 | 0.0022128 |
| BI | 0.283988 | 0.0153212 | 0.28219 | 0.000537487 |
| BJ | 0.284115 | 0.0190725 | 0.28219 | 0.00053731 |
| BK | 0.284234 | 9.36E-05 | 0.282354 | 0.00025372 |
| BL | 0.19912 | 0.294375 | 0.282719 | 0.000138813 |
| BM | 0.284211 | 5.22E-16 | 0.28219 | 0.000537251 |
| BN | 0.284211 | 5.22E-16 | 0.28219 | 0.000537251 |
| BO | 0.284211 | 5.22E-16 | 0.28219 | 0.000537251 |
| BP | 0.284211 | 5.22E-16 | 0.28219 | 0.000537251 |
| BQ | 0.284211 | 5.22E-16 | 0.28219 | 0.000537251 |
| BR | 0.284211 | 5.22E-16 | 0.28219 | 0.000537251 |
| BS | 0.284211 | 5.22E-16 | 0.28219 | 0.000537251 |
| BT | 0.284211 | 5.22E-16 | 0.28219 | 0.000537251 |
| BU | 0.284211 | 5.22E-16 | 0.28219 | 0.000537251 |
| BV | 0.284211 | 5.22E-16 | 0.28219 | 0.000537251 |
| BW | 0.284211 | 5.22E-16 | 0.28219 | 0.000537251 |
| BX | 0.284211 | 5.22E-16 | 0.28219 | 0.000537251 |
| BY | 0.284211 | 5.22E-16 | 0.28219 | 0.000537251 |
| BZ | 0.284211 | 5.22E-16 | 0.28219 | 0.000537251 |
| CA | 0.284211 | 5.22E-16 | 0.28219 | 0.000537251 |
| CB | 0.284211 | 5.22E-16 | 0.28219 | 0.000537251 |
| CC | 0.284211 | 5.22E-16 | 0.28219 | 0.000537251 |
| CD | 0.284211 | 5.22E-16 | 0.28219 | 0.000537251 |
| CE | 0.284211 | 5.22E-16 | 0.28219 | 0.000537251 |
| CF | 0.284211 | 5.22E-16 | 0.28219 | 0.000537251 |
| CG | 0.284211 | 5.22E-16 | 0.28219 | 0.000537251 |
| CH | 0.284211 | 1.71E-16 | 0.28219 | 0.000537251 |
| CI | 0.284211 | 5.22E-16 | 0.282573 | 4.08E-05 |
| CJ | 0.284211 | 1.71E-16 | 0.282398 | 0.000403813 |
| CK | 0.284162 | 0.00262834 | 0.282535 | 0.00021814 |
| CL | 0.284196 | 0.00177106 | 0.282536 | 0.000216376 |
| CM | 0.28424 | 0.000114465 | 0.282461 | 0.00028666 |
| CN | 0.28424 | 0.00260126 | 0.282461 | 0.00028666 |
| CO | 0.284135 | 0.00537217 | 0.282179 | 0.000538933 |
| CP | 0.28418 | 0.000147051 | 0.282178 | 0.000546373 |
| CQ | 0.284214 | 0.00754563 | 0.28219 | 0.000517044 |
| CR | 0.284158 | 0.0353891 | 0.282179 | 0.000538933 |
| CS | 0.284154 | 0.0565272 | 0.282178 | 0.000546372 |
| CT | 0.284222 | 0.00284029 | 0.282182 | 0.000542879 |
| CU | 0.284184 | 0.00441765 | 0.282328 | 0.000247203 |
| CV | 0.284205 | 0.000515637 | 0.282328 | 0.000247203 |
| CW | 0.284215 | 0.00147177 | 0.282184 | 0.000544282 |
| CX | 0.284218 | 0.000428997 | 0.28219 | 0.000521966 |
| CY | 0.284262 | 0.0051356 | 0.282676 | 0.00010177 |
| CZ | 0.284226 | 0.00821777 | 0.282667 | 0.000104441 |
| DA | 0.124036 | 0.572273 | 0.282547 | 0.000217877 |
| DB | 0.284207 | 0.00202072 | 0.28219 | 0.000537251 |
| DC | 0.261313 | 0.0728683 | 0.282722 | 0.000214656 |
| DD | 0.284211 | 7.75E-17 | 0.28219 | 0.000537251 |
| DE | 0.261037 | 0.0750526 | 0.282482 | 0.000166835 |
| DF | 0.284211 | 5.22E-16 | 0.28258 | 4.00E-05 |
| DG | 0.284211 | 5.22E-16 | 0.282185 | 0.000532148 |
| DH | 0.283521 | 0.0104563 | 0.282465 | 0.000131239 |
| DI | 0.290794 | 0.000893861 | 0.28215 | 0.000354962 |
| DJ | 0.292704 | 0.00719435 | 0.282329 | 0.000332117 |
| DK | 0.283248 | 0.0141901 | 0.282231 | 0.000381028 |
| DL | 0.283635 | 0.00583833 | 0.2824 | 0.000108631 |
| DM | 0.272293 | 0.0373239 | 0.28232 | 0.000126639 |
| DN | 0.292157 | 0.00468643 | 0.282592 | 0.00014061 |
| DO | 0.279288 | 0.0204962 | 0.28236 | 0.000199014 |
| DP | 0.274936 | 0.0353964 | 0.282452 | 0.000255112 |
| DQ | 0.235689 | 0.175031 | 0.282503 | 0.000215933 |
| DR | 0.286413 | 0.00863141 | 0.282547 | 0.000312237 |
| DS | 0.284031 | 0.00104994 | 0.282416 | 0.000291193 |
| DT | 0.297314 | 0.000873449 | 0.282286 | 0.00043958 |
| DU | 0.275672 | 0.0336833 | 0.282538 | 0.00035164 |
| DV | 0.288242 | 0.0255946 | 0.282467 | 0.000288507 |
| DW | 0.266573 | 0.061675 | 0.28244 | 0.000185309 |
| DX | 0.281473 | 0.0259592 | 0.282535 | 0.000191121 |
| DY | 0.284211 | 5.22E-16 | 0.28219 | 0.000537251 |
| DZ | 0.284211 | 5.22E-16 | 0.28219 | 0.000537251 |
| EA | 0.284211 | 5.22E-16 | 0.28219 | 0.000537251 |
| EB | 0.284211 | 5.22E-16 | 0.28219 | 0.000537251 |
| EC | 0.284211 | 5.22E-16 | 0.28219 | 0.000537251 |
| ED | 0.284211 | 5.22E-16 | 0.28219 | 0.000537251 |
| EE | 0.284211 | 5.22E-16 | 0.28219 | 0.000537251 |
| EF | 0.284211 | 5.22E-16 | 0.28219 | 0.000537251 |
| EG | 0.284211 | 5.22E-16 | 0.28219 | 0.000537251 |
| EH | 0.284211 | 5.22E-16 | 0.28219 | 0.000537251 |
| EI | 0.284211 | 5.22E-16 | 0.28219 | 0.000537251 |
| EJ | 0.284211 | 5.22E-16 | 0.28219 | 0.000537251 |
| EK | 0.284211 | 5.22E-16 | 0.28219 | 0.000537251 |
| EL | 0.284211 | 5.22E-16 | 0.28219 | 0.000537251 |
| EM | 0.284211 | 5.22E-16 | 0.28219 | 0.000537251 |
| EN | 0.284211 | 5.22E-16 | 0.28219 | 0.000537251 |
| EO | 0.284211 | 5.22E-16 | 0.28219 | 0.000537251 |
| EP | 283.31 | 0.259966 | 0.282382 | 0.000458568 |
| EQ | 0.284211 | 5.22E-16 | 0.28219 | 0.000537251 |
| ER | 0.284211 | 5.22E-16 | 0.282337 | 0.000260717 |
| ES | 0.284211 | 5.22E-16 | 0.282342 | 0.000260793 |
| ET | 0.284211 | 2.92E-16 | 0.28219 | 0.000537251 |
| EU | 0.229414 | 0.481249 | 0.282133 | 0.000403955 |
| EV | 0.284211 | 5.22E-16 | 0.282257 | 0.000709353 |
| EW | 0.229363 | 0.481258 | 0.282368 | 0.000219113 |
| EX | 0.284211 | 5.22E-16 | 0.28219 | 0.000537153 |
| EY | 0.284211 | 4.61E-16 | 0.28219 | 0.000537251 |
| EZ | 1571.1 | 0.0190979 | 0.282313 | 0.000189961 |
| FA | 0.284211 | 5.22E-16 | 0.28219 | 0.000537251 |
| FB | 0.284211 | 1.30E-16 | 0.28219 | 0.000537251 |
| FC | 0.284211 | 1.30E-16 | 0.28219 | 0.000537251 |
| FD | 0.230417 | 0.487604 | 0.282494 | 0.000165675 |
| FE | 0.230417 | 0.487604 | 0.282494 | 0.000165675 |
| FF | 0.284211 | 5.22E-16 | 0.282279 | 0.00066047 |
| FG | 0.284211 | 2.13E-16 | 0.282185 | 0.00053603 |
| FH | 0.181913 | 0.365785 | 0.282494 | 0.000165675 |
| FI | 0.284211 | 5.22E-16 | 0.282488 | 0.00016573 |
| FJ | 0.284085 | 0.186497 | 0.28219 | 0.000537251 |
| FK | 0.284211 | 3.81E-16 | 0.28252 | 0.000124944 |
| FL | 0.229757 | 0.488147 | 0.282275 | 0.000773295 |
| FM | 0.229569 | 0.488555 | 0.282175 | 0.000641749 |
| FN | 0.22954 | 0.488321 | 0.282171 | 0.000542118 |
| FO | 0.284211 | 5.22E-16 | 0.28252 | 0.000124944 |
| FP | 0.229862 | 0.487965 | 0.282275 | 0.000773295 |
| FQ | 0.229569 | 0.488555 | 0.282175 | 0.000641749 |
| FR | 0.22954 | 0.488321 | 0.282171 | 0.000542118 |
| FS | 0.284211 | 2.24E-16 | 0.28219 | 0.000537251 |
| FT | 0.284211 | 5.22E-16 | 0.28219 | 0.000537251 |
| FU | 0.284211 | 5.22E-16 | 0.28219 | 0.000537251 |
| FV | 0.284211 | 5.22E-16 | 0.282327 | 0.000218834 |
| FW | 0.284211 | 5.22E-16 | 0.28219 | 0.000537251 |
| FX | 0.284211 | 5.22E-16 | 0.28219 | 0.000537251 |
| FY | 0.284211 | 5.22E-16 | 0.28219 | 0.000537251 |
| FZ | 0.284211 | 5.22E-16 | 0.28219 | 0.000537251 |
| GA | 0.284211 | 5.22E-16 | 0.28219 | 0.000537251 |
| GB | 0.284211 | 5.22E-16 | 0.28219 | 0.000537251 |
| GC | 0.284211 | 5.22E-16 | 0.28219 | 0.000537251 |
| GD | 0.284211 | 5.22E-16 | 0.28219 | 0.000537251 |
| GE | 0.284211 | 5.22E-16 | 0.28219 | 0.000537251 |
| GF | 11.2039 | 0.0494116 | 0.282228 | 0.000517236 |
| GG | 0.284211 | 5.22E-16 | 0.28219 | 0.000537251 |
| GH | 0.284211 | 5.22E-16 | 0.282456 | 0.000189013 |
| GI | 0.284211 | 5.22E-16 | 0.282463 | 0.000188829 |
| GJ | 0.284211 | 1.75E-15 | 0.28219 | 0.000537251 |
| GK | 0.284211 | 5.22E-16 | 0.282243 | 0.000554943 |
| GL | 0.284211 | 5.22E-16 | 0.282169 | 0.000606276 |
| GM | 0.284211 | 5.22E-16 | 0.282359 | 0.000224482 |
| GN | 0.284211 | 5.22E-16 | 0.28219 | 0.000537251 |
| GO | 0.284211 | 5.22E-16 | 0.28216 | 0.000631482 |
| GP | 0.237262 | 0.479168 | 0.282189 | 0.000541914 |
| GQ | 0.263397 | 0.066871 | 0.281733 | 0.00219577 |
| GR | 0.224253 | 0.210014 | 0.282926 | 0.000458821 |
| GS | 0.245332 | 0.137544 | 0.282221 | 0.000718864 |
| GT | 0.287436 | 0.000304578 | 0.282517 | 0.000240093 |
| GU | 0.164068 | 0.437177 | 0.282197 | 0.000501678 |
| GV | 0.284211 | 4.82E-16 | 0.28219 | 0.000537629 |
| GW | 0.284211 | 3.35E-16 | 0.28219 | 0.000538066 |
| GX | 0.284211 | 2.46E-16 | 0.282182 | 0.000566807 |
| GY | 0.284211 | 2.28E-16 | 0.28219 | 0.00053797 |
| GZ | 0.284211 | 2.02E-16 | 0.282187 | 0.000572776 |
| HA | 0.284211 | 4.58E-16 | 0.282192 | 0.000572159 |
| HB | 0.284211 | 2.72E-16 | 0.28219 | 0.000537938 |
| HC | 0.284211 | 2.18E-16 | 0.28219 | 0.000537568 |
| HD | 0.284211 | 6.40E-16 | 0.28219 | 0.000537268 |
| HE | 0.284211 | 5.22E-16 | 0.28219 | 0.000537251 |
| HF | 0.284211 | 5.22E-16 | 0.28219 | 0.000537251 |
| HG | 0.284211 | 5.22E-16 | 0.28219 | 0.000537251 |

table 3. regression on FACTORY data using epsilon-SVR

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  |  |  |  |  |
| kernel | nu-SVR |  |  |  |
|  | single |  | reverse |  |
| features | MSR | SCC | MSR | SCC |
| A | 0.165371 | 0.473318 | 0.286226 | 0.000879012 |
| B | 0.197806 | 0.479708 | 0.28619 | 0.00028138 |
| C | 0.304679 | nan | 0.286396 | 0.000849354 |
| D | 0.304679 | nan | 0.286396 | 0.000849354 |
| E | 0.252437 | 0.479201 | 0.286706 | 5.84E-07 |
| F | 0.165399 | 0.473244 | 0.286166 | 0.000891275 |
| G | 3708.43 | 0.00162672 | 0.28587 | 0.000242886 |
| H | 0.304679 | nan | 0.286167 | 1.39E-05 |
| I | 0.304679 | nan | 0.286045 | 1.51E-05 |
| J | 0.304679 | nan | 0.286396 | 0.000849354 |
| K | 0.165373 | 0.47332 | 0.286381 | 0.00073947 |
| L | 0.304679 | nan | 0.286009 | 8.05E-07 |
| M | 0.304679 | 0 | 0.286396 | 0.000849354 |
| N | 0.304679 | nan | 0.286396 | 0.000849354 |
| O | 0.304679 | nan | 0.286396 | 0.000849354 |
| P | 0.304679 | nan | 0.286396 | 0.000849354 |
| Q | 0.304679 | nan | 0.286396 | 0.000849354 |
| R | 0.203134 | 0.390821 | 0.287095 | 1.54E-05 |
| S | 0.304679 | 0 | 0.286396 | 0.000849354 |
| T | nan | nan | 0.286616 | 0.000233359 |
| U | 0.304679 | nan | 0.286396 | 0.000849354 |
| V | 0.304679 | nan | 0.286396 | 0.000849354 |
| W | 0.316452 | 0.00257608 | 0.285969 | 0.000749802 |
| X | 0.304679 | nan | 0.286396 | 0.000849354 |
| Y | 0.304679 | nan | 0.286396 | 0.000849354 |
| Z | 0.304679 | nan | 0.286396 | 0.000849354 |
| AA | 0.304679 | nan | 0.286396 | 0.000849354 |
| AB | 0.304679 | nan | 0.286396 | 0.000849354 |
| AC | 0.304679 | 4.35E-17 | 0.286396 | 0.000849843 |
| AD | 0.304679 | 4.35E-17 | 0.286396 | 0.000849843 |
| AE | 0.30711 | 0.000479264 | 0.286369 | 0.00090735 |
| AF | 0.252431 | 0.479181 | 0.286297 | 0.00011726 |
| AG | 0.00108817 | 0.99615 | 0.286396 | 0.000849335 |
| AH | 0.304679 | nan | 0.286396 | 0.000849354 |
| AI | 0.304679 | nan | 0.286396 | 0.000849354 |
| AJ | 0.305149 | 0.00117467 | 0.286423 | 0.000483658 |
| AK | 0.16565 | 0.434831 | 0.286013 | 4.15E-05 |
| AL | 0.304679 | nan | 0.286396 | 0.000849354 |
| AM | 0.304679 | nan | 0.286396 | 0.000849354 |
| AN | 0.304679 | nan | 0.286395 | 0.000860945 |
| AO | 285.861 | 0.263531 | 0.28662 | 0.000322806 |
| AP | 0.31961 | 0.000156672 | 0.286159 | 0.00129746 |
| AQ | 0.252337 | 0.479196 | 0.286067 | 2.06E-07 |
| AR | 0.304679 | nan | 0.286396 | 0.000849354 |
| AS | 0.304679 | nan | 0.286396 | 0.000849354 |
| AT | 0.304679 | 0 | 0.286396 | 0.000849354 |
| AU | 8.14146 | 0.0419774 | 0.286543 | 0.000386563 |
| AV | 0.304678 | 0.000924599 | 0.286272 | 0.0013201 |
| AW | 0.304679 | nan | 0.286396 | 0.000849354 |
| AX | 0.304679 | nan | 0.286396 | 0.000849354 |
| AY | 3708.43 | 0.00162672 | 0.286375 | 0.000155033 |
| AZ | 0.322564 | 0.000604869 | 0.285853 | 0.000711179 |
| BA | 0.252478 | 0.479199 | 0.286164 | 3.36E-06 |
| BB | 0.304679 | 0 | 0.286395 | 0.000860409 |
| BC | 0.304679 | 0 | 0.286395 | 0.000860409 |
| BD | 0.304681 | 0.000364623 | 0.286396 | 0.000850063 |
| BE | 0.304714 | 0.0108046 | 0.286395 | 0.000863444 |
| BF | 0.304608 | 0.0635045 | 0.286395 | 0.000866475 |
| BG | 0.304526 | 0.166441 | 0.286066 | 0.000227037 |
| BH | 0.242416 | 0.182722 | 0.286115 | 0.00209626 |
| BI | 0.304673 | 0.0038303 | 0.286396 | 0.000849642 |
| BJ | 0.304638 | 0.0188206 | 0.286396 | 0.000849428 |
| BK | 0.304649 | 3.13E-05 | 0.286319 | 0.000190063 |
| BL | 0.203192 | 0.296403 | 0.286473 | 1.23E-06 |
| BM | 0.304679 | nan | 0.286396 | 0.000849354 |
| BN | 0.304679 | nan | 0.286396 | 0.000849354 |
| BO | 0.304679 | nan | 0.286396 | 0.000849354 |
| BP | 0.304679 | nan | 0.286396 | 0.000849354 |
| BQ | 0.304679 | nan | 0.286396 | 0.000849354 |
| BR | 0.304679 | nan | 0.286396 | 0.000849354 |
| BS | 0.304679 | nan | 0.286396 | 0.000849354 |
| BT | 0.304679 | nan | 0.286396 | 0.000849354 |
| BU | 0.304679 | nan | 0.286396 | 0.000849354 |
| BV | 0.304679 | nan | 0.286396 | 0.000849354 |
| BW | 0.304679 | nan | 0.286396 | 0.000849354 |
| BX | 0.304679 | nan | 0.286396 | 0.000849354 |
| BY | 0.304679 | nan | 0.286396 | 0.000849354 |
| BZ | 0.304679 | nan | 0.286396 | 0.000849354 |
| CA | 0.304679 | nan | 0.286396 | 0.000849354 |
| CB | 0.304679 | nan | 0.286396 | 0.000849354 |
| CC | 0.304679 | nan | 0.286396 | 0.000849354 |
| CD | 0.304679 | nan | 0.286396 | 0.000849354 |
| CE | 0.304679 | nan | 0.286396 | 0.000849354 |
| CF | 0.304679 | nan | 0.286396 | 0.000849354 |
| CG | 0.304679 | nan | 0.286396 | 0.000849354 |
| CH | 0.304679 | nan | 0.286396 | 0.000849354 |
| CI | 0.304659 | 0.0218827 | 0.286336 | 2.68E-05 |
| CJ | 0.304679 | nan | 0.286056 | 0.000627196 |
| CK | 0.304559 | 0.00828182 | 0.286586 | 2.41E-06 |
| CL | 0.304611 | 0.0298837 | 0.286566 | 1.52E-06 |
| CM | 0.304547 | 0.00359655 | 0.286052 | 0.000211497 |
| CN | 0.304653 | 0.00432912 | 0.286052 | 0.000211497 |
| CO | 0.304495 | 0.0138213 | 0.286425 | 0.000830432 |
| CP | 0.304601 | 0.00466723 | 0.286423 | 0.000838386 |
| CQ | 0.304647 | 0.00155819 | 0.286408 | 0.000827685 |
| CR | 0.304643 | 0.0487349 | 0.286425 | 0.000830432 |
| CS | 0.304712 | 0.0289995 | 0.286423 | 0.000838802 |
| CT | 0.304602 | 2.46E-06 | 0.286414 | 0.00086407 |
| CU | 0.304662 | 0.00217062 | 0.28696 | 7.71E-06 |
| CV | 0.304588 | 0.00656636 | 0.28696 | 7.71E-06 |
| CW | 0.304651 | 0.0152041 | 0.286439 | 0.000809818 |
| CX | 0.304659 | 0.000359745 | 0.286436 | 0.000802576 |
| CY | 0.304671 | 0.00303223 | 0.286187 | 2.64E-06 |
| CZ | 0.30467 | 0.00114534 | 0.286195 | 2.47E-06 |
| DA | 0.120567 | 0.578965 | 0.286564 | 2.45E-06 |
| DB | 0.304668 | 0.000634448 | 0.286396 | 0.000849354 |
| DC | 0.274283 | 0.0742529 | 0.28594 | 1.14E-05 |
| DD | 0.304681 | 0.00109191 | 0.286396 | 0.000849354 |
| DE | 0.273644 | 0.0795767 | 0.2868 | 1.13E-05 |
| DF | 0.304659 | 0.0218827 | 0.286351 | 3.63E-05 |
| DG | 0.304679 | nan | 0.28641 | 0.000845613 |
| DH | 0.301215 | 0.01071 | 0.28668 | 1.90E-05 |
| DI | 0.306686 | 0.00107786 | 0.286593 | 0.000450264 |
| DJ | 0.309107 | 0.00643972 | 0.286516 | 3.17E-05 |
| DK | 0.300544 | 0.0127994 | 0.286254 | 0.000440996 |
| DL | 0.301253 | 0.00657677 | 0.286615 | 1.35E-05 |
| DM | 0.288498 | 0.0368387 | 0.286808 | 3.72E-06 |
| DN | 0.307391 | 0.00438315 | 0.285948 | 0.000157329 |
| DO | 0.295483 | 0.0216352 | 0.286466 | 0.00022321 |
| DP | 0.291861 | 0.0338983 | 0.286197 | 3.92E-05 |
| DQ | 0.251957 | 0.166534 | 0.286316 | 6.20E-07 |
| DR | 0.301833 | 0.00923185 | 0.286098 | 4.02E-05 |
| DS | 0.303821 | 0.000644561 | 0.28631 | 3.59E-05 |
| DT | 0.313807 | 0.000512514 | 0.286229 | 0.000467152 |
| DU | 0.290174 | 0.0328268 | 0.285981 | 0.000279845 |
| DV | 0.299971 | 0.0253546 | 0.285799 | 0.000351183 |
| DW | 0.283826 | 0.0611075 | 0.28626 | 4.86E-05 |
| DX | 0.294249 | 0.0266527 | 0.286316 | 2.53E-05 |
| DY | 0.304679 | nan | 0.286396 | 0.000849354 |
| DZ | 0.304679 | nan | 0.286396 | 0.000849354 |
| EA | 0.304679 | nan | 0.286396 | 0.000849354 |
| EB | 0.304679 | nan | 0.286396 | 0.000849354 |
| EC | 0.304679 | nan | 0.286396 | 0.000849354 |
| ED | 0.304679 | nan | 0.286396 | 0.000849354 |
| EE | 0.304679 | nan | 0.286396 | 0.000849354 |
| EF | 0.304679 | nan | 0.286396 | 0.000849354 |
| EG | 0.304679 | nan | 0.286396 | 0.000849354 |
| EH | 0.304679 | nan | 0.286396 | 0.000849354 |
| EI | 0.304679 | nan | 0.286396 | 0.000849354 |
| EJ | 0.304679 | nan | 0.286396 | 0.000849354 |
| EK | 0.304679 | nan | 0.286396 | 0.000849354 |
| EL | 0.304679 | nan | 0.286396 | 0.000849354 |
| EM | 0.304679 | nan | 0.286396 | 0.000849354 |
| EN | 0.304679 | nan | 0.286396 | 0.000849354 |
| EO | 0.304679 | nan | 0.286396 | 0.000849354 |
| EP | 281.074 | 0.262086 | 0.286303 | 0.000799774 |
| EQ | 0.304679 | nan | 0.286396 | 0.000849354 |
| ER | 0.304679 | nan | 0.286865 | 3.76E-06 |
| ES | 0.304679 | nan | 0.28685 | 3.71E-06 |
| ET | 0.304679 | 4.90E-16 | 0.286396 | 0.000849354 |
| EU | 0.243782 | 0.481253 | 0.286976 | 1.20E-05 |
| EV | 0.304679 | 3.77E-17 | 0.286152 | 0.000927382 |
| EW | 0.243767 | 0.481253 | 0.286294 | 0.000304145 |
| EX | 0.304679 | 0 | 0.286396 | 0.000849789 |
| EY | 0.304679 | 0 | 0.286396 | 0.000849354 |
| EZ | 3706.07 | 0.00163996 | 0.286984 | 1.74E-06 |
| FA | 0.304679 | nan | 0.286396 | 0.000849354 |
| FB | 0.304679 | -4.90E-16 | 0.286396 | 0.000849354 |
| FC | 0.304679 | -4.90E-16 | 0.286396 | 0.000849354 |
| FD | 0.244653 | 0.487594 | 0.286813 | 2.94E-06 |
| FE | 0.244653 | 0.487594 | 0.286813 | 2.94E-06 |
| FF | 0.304679 | nan | 0.286117 | 0.000937072 |
| FG | 0.304679 | 1.55E-16 | 0.286411 | 0.000825812 |
| FH | 0.180953 | 0.365801 | 0.286813 | 2.94E-06 |
| FI | 0.304679 | nan | 0.286827 | 2.74E-06 |
| FJ | 0.304493 | 0.00532489 | 0.286396 | 0.000849354 |
| FK | 0.304636 | 0.0318438 | 0.286872 | 2.18E-06 |
| FL | 0.244178 | 0.488175 | 0.286075 | 0.00102854 |
| FM | 0.243716 | 0.488572 | 0.286378 | 0.000886778 |
| FN | 0.244032 | 0.488343 | 0.286439 | 0.000839698 |
| FO | 0.304622 | 0.0667928 | 0.286872 | 2.18E-06 |
| FP | 0.24411 | 0.488209 | 0.286075 | 0.00102854 |
| FQ | 0.243716 | 0.488572 | 0.286378 | 0.000886778 |
| FR | 0.244032 | 0.488343 | 0.286439 | 0.000839698 |
| FS | 0.304683 | 0.00041653 | 0.286396 | 0.000849354 |
| FT | 0.304679 | nan | 0.286396 | 0.000849354 |
| FU | 0.304679 | nan | 0.286396 | 0.000849354 |
| FV | 0.304679 | nan | 0.286319 | 0.00029881 |
| FW | 0.304679 | nan | 0.286396 | 0.000849354 |
| FX | 0.304679 | nan | 0.286396 | 0.000849354 |
| FY | 0.304679 | nan | 0.286396 | 0.000849354 |
| FZ | 0.304679 | nan | 0.286396 | 0.000849354 |
| GA | 0.304679 | nan | 0.286396 | 0.000849354 |
| GB | 0.304679 | nan | 0.286396 | 0.000849354 |
| GC | 0.304679 | nan | 0.286396 | 0.000849354 |
| GD | 0.304679 | nan | 0.286396 | 0.000849354 |
| GE | 0.304679 | nan | 0.286396 | 0.000849354 |
| GF | 8.14037 | 0.0420259 | 0.286637 | 0.000588662 |
| GG | 0.304679 | nan | 0.286396 | 0.000849354 |
| GH | 0.304679 | nan | 0.286036 | 0.000337613 |
| GI | 0.304679 | nan | 0.286021 | 0.000335466 |
| GJ | 0.304679 | 0 | 0.286396 | 0.000849354 |
| GK | 0.304679 | nan | 0.28616 | 0.00123143 |
| GL | 0.304679 | nan | 0.286421 | 0.000813665 |
| GM | 0.304679 | nan | 0.286304 | 0.000303111 |
| GN | 0.304679 | nan | 0.286396 | 0.000849354 |
| GO | 0.304679 | nan | 0.286431 | 0.000846946 |
| GP | 0.250381 | 0.479359 | 0.286396 | 0.000854539 |
| GQ | 0.267675 | 0.0676537 | 0.287539 | 0.00138376 |
| GR | 0.230579 | 0.202423 | 0.286689 | 0.000336835 |
| GS | 0.260024 | 0.136443 | 0.286145 | 0.000932241 |
| GT | 0.307018 | 0.000304026 | 0.286767 | 9.19E-06 |
| GU | 0.158432 | 0.442802 | 0.28639 | 0.000849904 |
| GV | 0.304679 | nan | 0.286396 | 0.000849643 |
| GW | 0.304679 | 0 | 0.286396 | 0.000850326 |
| GX | 0.304679 | 0.000494462 | 0.286395 | 0.000889896 |
| GY | 0.304681 | 0.002775 | 0.286396 | 0.000850268 |
| GZ | 0.304679 | 0 | 0.286381 | 0.000896644 |
| HA | 0.304684 | 0.000304866 | 0.28637 | 0.000898589 |
| HB | 0.304679 | 5.44E-17 | 0.286396 | 0.000850213 |
| HC | 0.304676 | 5.78E-05 | 0.286396 | 0.000849746 |
| HD | 0.304679 | 0 | 0.286396 | 0.000849376 |
| HE | 0.304679 | 0 | 0.286396 | 0.000849354 |
| HF | 0.304679 | nan | 0.286396 | 0.000849354 |
| HG | 0.304679 | nan | 0.286396 | 0.000849354 |

table 4. regression on FACTORY data using nu-SVR

As shown above, we conducted the regression experiment and collected all results on each feature. From the table, we could found that there existed some features performed well in regressing, therefore, which would also perform well in classifying. What’s more, on empirical methods, we also knew that the whole SVM classifier was capable of classifying the dataset correctly. So we continued to conduct the classification experiment.

# 5.CLASSIFICATION EXPERIMENT

Based on the regression experiment above, we continued to conduct the classification experiment, which was the main part of the whole job.

We will start the explanation by talking about the kernels. One of the basic theory of SVM is to cast the data in the low dimension into the high dimension to find a proper hyper plane to achieve the goal of separating the data. There were several kernels could be used, and we chose two types of kernels: linear and rbf (radial basis function). The linear kernel is the simplest and most straight-forward kernel, it just inputs points in low dimension into a linear function and get the output of high dimension, while the rbf uses the radial basis function to get the output. Linear kernel is simple to understand but is slow, while the rbf kernel is stable and fast. So we regarded the result done by rbf kernel as the main result while the result done by linear function as the contrast.

Crucial parameters of SVM are c and gamma (radial kernel only). The value c reflects the tolerance of SVM to the error. The bigger c is, the system is more incapable of tolerancing errors. The parameter gamma decides the distribution of the data when they are casted into the high-dimension space.

As mentioned in the regression experiment, we used the titles as the same, except the parameter c. We used 2 kernels, 3 values of c, “single” and “reverse”, so there were 2\*3\*2=12 sets of results in total.

We judged the importance of a feature by observing the correct rate, which was the main entity in the result table. The meaning of the term “Correct Rate” is as follows. We used some data to train the model of predicting the result of each set of data with its label which was called “training data”, then inputted some “test data” without its labels and get a label through our predicting model trained above. If the predicted label was identical to the origin one, then it was classified correctly, or wrong otherwise. At last, we counted the total quantity of the correct ones, and divided by the whole number of dataset.

The correct rate was done by the k-fold cross validation (reference [3]), which was a popular strategy in the machine learning field. At the beginning, we split the whole data set into k (say, 5) subsets evenly, then we trained the model by subset 1, 2, 3 and 4, and used subset 5 as the test set to get the result. Then we used subset 1, 3, 4, 5 to train and subset 2 to test, and so on. Finally, we take the average of 5 results, as the final result – correct rate.

## 5.1 classification experiment on ICI data

The result tables were as follows. The columns titled by “single” are the higher the better, while the columns titled by “reverse” are the lower the better. Two tables are different with each other by the selection of the kernel type.

Thus, we could find the features which played important role in separating the data correctly: feature K, N, V, whose single experiment had the correct rate over 50%. Among these three features, the feature N is the best, which had the correct rate over 55%.

In addition, we have done another experiment to find the most important features subset. By the select strategy introduced by the reference [1], we started the experiment using a small subset of random features combination, then add or drop feature dynamically. Finally, we got the combination of feature K, N, V, which was identical to the most important subset mentioned in the paragraph above. We could achieve about 75% correct rate using the combination by k-fold cross validation, and get about 82% using the whole data set to train.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | ICI |  |  |  |  |  |
| kernel | linear |  |  |  |  |  |
| c | c=0.5 |  | c=1 |  | c=5 |  |
| features | single | reverse | single | reverse | single | reverse |
| A | 42.4008 | 40.7007 | 42.3493 | 36.3215 | 37.3004 | 37.3004 |
| B | 38.4853 | 33.1788 | 38.4853 | 36.6306 | 31.2725 | 31.2725 |
| C | 22.7718 | 34.5698 | 16.5379 | 37.558 | 14.9408 | 14.9408 |
| D | 26.5327 | 37.6095 | 20.1443 | 36.0639 | 23.7506 | 23.7506 |
| E | 30.3967 | 35.6002 | 36.9397 | 35.8578 | 25.5538 | 25.5538 |
| F | 50.9016 | 33.694 | 50.8501 | 37.4034 | 43.1221 | 43.1221 |
| G | 48.5317 | 34.0031 | 48.5317 | 33.694 | 48.5317 | 48.5317 |
| H | 31.8908 | 33.9001 | 23.3385 | 30.5513 | 24.3174 | 24.3174 |
| I | 46.3163 | 38.3823 | 46.3679 | 36.6306 | 46.4194 | 46.4194 |
| J | 34.6728 | 35.4456 | 34.6728 | 37.6095 | 34.6728 | 34.6728 |
| K | 34.6728 | 26.7388 | 34.6728 | 30.7058 | 34.6728 | 34.6728 |
| L | 35.3426 | 34.7244 | 35.3426 | 31.9423 | 35.3426 | 35.3426 |
| M | 35.3426 | 33.7455 | 35.3426 | 35.8578 | 35.3426 | 35.3426 |
| N | 44.7707 | 31.3756 | 44.7707 | 34.7244 | 44.7707 | 44.7707 |
| O | 36.476 | 35.4456 | 39.0005 | 34.1061 | 39.1036 | 39.1036 |
| P | 34.6728 | 34.1577 | 34.6728 | 38.1762 | 34.6728 | 34.6728 |
| Q | 34.6728 | 36.0639 | 34.6728 | 34.5698 | 34.6728 | 34.6728 |
| R | 46.677 | 33.7455 | 49.356 | 36.373 | 49.4075 | 49.4075 |
| S | 34.6728 | 33.7455 | 34.6728 | 36.1669 | 34.6728 | 34.6728 |
| T | 44.7707 | 29.6754 | 45.0799 | 33.9001 | 45.1314 | 45.1314 |
| U | 34.6213 | 31.5301 | 34.6213 | 33.1273 | 34.6213 | 34.6213 |
| V | 52.035 | 38.846 | 52.035 | 35.7548 | 52.0866 | 52.0866 |
| W | 49.253 | 33.2303 | 49.253 | 36.9912 | 49.5621 | 49.5621 |
| X | 44.9253 | 29.4693 | 44.8223 | 37.0428 | 44.7707 | 44.7707 |
| Y | 36.2184 | 32.1999 | 36.27 | 36.8367 | 36.2184 | 36.2184 |
| Z | 34.6728 | 32.3545 | 34.6728 | 32.7151 | 34.6728 | 34.6728 |
| AB | 41.6795 | 35.9093 | 41.7826 | 32.8697 | 41.9371 | 41.9371 |

table 5. classification on ICI data using linear

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  |  |
| kernel | rbf |  |  |  |  |  |
| c | c=0.5 |  | c=1 |  | c=5 |  |
| features | single | reverse | single | reverse | single | reverse |
| A | 49.253 | 34.6728 | 47.2952 | 34.6728 | 45.0283 | 34.6728 |
| B | 49.3045 | 34.6728 | 48.1195 | 34.6728 | 47.3467 | 34.6728 |
| C | 34.6728 | 34.6728 | 34.6728 | 34.6728 | 34.7759 | 34.6728 |
| D | 34.6728 | 34.6728 | 34.6728 | 34.6728 | 34.6213 | 34.6728 |
| E | 36.8882 | 34.6728 | 44.3071 | 34.6728 | 44.101 | 34.6728 |
| F | 50.0258 | 34.6728 | 49.2014 | 34.6728 | 47.4498 | 34.6728 |
| G | 47.2437 | 34.6728 | 46.4709 | 34.6728 | 45.389 | 34.6728 |
| H | 34.6728 | 34.6728 | 34.3637 | 34.6728 | 34.2092 | 34.6728 |
| I | 46.1618 | 34.6728 | 45.6466 | 34.6728 | 45.3375 | 34.6728 |
| J | 36.2184 | 34.6728 | 35.4456 | 34.6728 | 35.1365 | 34.6728 |
| K | 53.6837 | 34.6728 | 53.9413 | 34.6728 | 53.9928 | 34.6728 |
| L | 35.3426 | 34.6728 | 35.3426 | 34.6728 | 35.3426 | 34.6728 |
| M | 35.3426 | 34.6728 | 35.3426 | 34.6728 | 35.3426 | 34.6728 |
| N | 55.3838 | 34.6728 | 55.3838 | 34.6728 | 55.3838 | 34.6728 |
| O | 39.3612 | 34.6728 | 39.5672 | 34.6728 | 40.237 | 34.6728 |
| P | 34.6728 | 34.6728 | 34.6728 | 34.6728 | 34.6728 | 34.6728 |
| Q | 34.6728 | 34.6728 | 34.4668 | 34.6728 | 34.7759 | 34.6728 |
| R | 49.3045 | 34.6728 | 49.4075 | 34.6728 | 49.4075 | 34.6728 |
| S | 34.6728 | 34.6728 | 34.6728 | 34.6728 | 34.6728 | 34.6728 |
| T | 44.9253 | 34.6728 | 44.9768 | 34.6728 | 45.0283 | 34.6728 |
| U | 34.8789 | 34.6728 | 35.2911 | 34.6728 | 35.4456 | 34.6728 |
| V | 52.6533 | 34.6728 | 52.7048 | 34.6728 | 52.5502 | 34.6728 |
| W | 49.5106 | 34.6728 | 49.356 | 34.6728 | 49.5106 | 34.6728 |
| X | 45.6981 | 34.6728 | 45.9042 | 34.6728 | 46.5739 | 34.6728 |
| Y | 36.373 | 34.6728 | 37.7125 | 34.6728 | 39.7218 | 34.6728 |
| Z | 34.6728 | 34.6728 | 34.6728 | 34.6728 | 34.5183 | 34.6728 |
| AB | 42.916 | 34.6728 | 43.1736 | 34.6728 | 44.8738 | 34.6728 |

table 6. classification on ICI data using rbf

## 5.2 classification experiment on FACTORY data

The illustration was the same as that of ICI data.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | factory |  |  |  |  |  |
| kernel | linear |  |  |  |  |  |
| c | c=0.5 |  | c=1 |  | c=5 |  |
| features | single | reverse | single | reverse | single | reverse |
| A | 92.383 | 92.2742 | 92.383 | 92.2742 | 92.383 | 92.2742 |
| B | 92.383 | 92.2742 | 88.901 | 92.2742 | 88.901 | 92.2742 |
| C | 92.383 | 92.2742 | 92.383 | 92.2742 | 92.383 | 92.2742 |
| D | 92.383 | 92.2742 | 92.383 | 92.2742 | 92.383 | 92.2742 |
| E | 92.383 | 91.0773 | 92.383 | 91.0773 | 92.383 | 91.0773 |
| F | 93.6888 | 91.0773 | 93.6888 | 91.0773 | 93.6888 | 91.0773 |
| G | 27.4211 | 88.5745 | 27.4211 | 88.5745 | 36.3439 | 88.5745 |
| H | 84.5484 | 90.8596 | 84.5484 | 90.8596 | 83.2427 | 84.8749 |
| I | 83.2427 | 91.0773 | 75.4081 | 91.0773 | 75.4081 | 91.0773 |
| J | 92.383 | 92.2742 | 92.383 | 92.2742 | 92.383 | 92.2742 |
| K | 93.6888 | 91.0773 | 93.6888 | 91.0773 | 93.6888 | 91.0773 |
| L | 21.5452 | 84.3308 | 21.5452 | 84.3308 | 54.2982 | 84.3308 |
| M | 92.383 | 92.2742 | 92.383 | 92.2742 | 92.383 | 92.2742 |
| N | 92.383 | 92.2742 | 92.383 | 92.2742 | 92.383 | 92.2742 |
| O | 92.383 | 92.2742 | 92.383 | 92.2742 | 92.383 | 92.2742 |
| P | 92.383 | 92.2742 | 92.383 | 92.2742 | 92.383 | 92.2742 |
| Q | 92.383 | 92.2742 | 92.383 | 92.2742 | 92.383 | 92.2742 |
| R | 79.2165 | 92.2742 | 82.9162 | 92.2742 | 89.5539 | 92.2742 |
| S | 92.383 | 92.2742 | 92.383 | 92.2742 | 92.383 | 92.2742 |
| T | 92.383 | 86.5071 | 92.383 | 86.5071 | 92.383 | 86.5071 |
| U | 92.383 | 92.2742 | 92.383 | 92.2742 | 92.383 | 92.2742 |
| V | 92.383 | 92.2742 | 92.383 | 92.2742 | 92.383 | 92.2742 |
| W | 92.383 | 92.2742 | 92.383 | 92.2742 | 92.383 | 92.2742 |
| X | 92.383 | 92.2742 | 92.383 | 92.2742 | 92.383 | 92.2742 |
| Y | 92.383 | 92.2742 | 92.383 | 92.2742 | 92.383 | 92.2742 |
| Z | 92.383 | 92.2742 | 92.383 | 92.2742 | 92.383 | 92.2742 |
| AA | 92.383 | 92.2742 | 92.383 | 92.2742 | 92.383 | 92.2742 |
| AB | 92.383 | 92.2742 | 92.383 | 92.2742 | 92.383 | 92.2742 |
| AC | 92.383 | 92.2742 | 92.383 | 92.2742 | 92.383 | 92.2742 |
| AD | 92.383 | 92.2742 | 92.383 | 92.2742 | 92.383 | 92.2742 |
| AE | 92.383 | 92.2742 | 92.383 | 92.2742 | 92.383 | 92.2742 |
| AF | 93.6888 | 92.2742 | 93.6888 | 92.2742 | 93.6888 | 92.2742 |
| AG | 100 | 92.2742 | 100 | 92.2742 | 100 | 92.2742 |
| AH | 92.383 | 92.2742 | 92.383 | 92.2742 | 92.383 | 92.2742 |
| AI | 92.383 | 92.2742 | 92.383 | 92.2742 | 92.383 | 92.2742 |
| AJ | 86.9423 | 92.2742 | 92.383 | 92.2742 | 92.383 | 92.2742 |
| AK | 92.383 | 92.2742 | 92.383 | 92.2742 | 92.383 | 92.2742 |
| AL | 92.383 | 92.2742 | 92.383 | 92.2742 | 92.383 | 92.2742 |
| AM | 92.383 | 92.2742 | 92.383 | 92.2742 | 92.383 | 92.2742 |
| AN | 92.383 | 92.2742 | 92.383 | 92.2742 | 92.383 | 92.2742 |
| AO | 12.1872 | 86.5071 | 14.6899 | 86.5071 | 13.3841 | 86.5071 |
| AP | 92.383 | 92.2742 | 92.383 | 92.2742 | 92.383 | 92.2742 |
| AQ | 93.6888 | 91.0773 | 93.6888 | 91.0773 | 93.6888 | 91.0773 |
| AR | 92.383 | 92.2742 | 92.383 | 92.2742 | 92.383 | 92.2742 |
| AS | 92.383 | 92.2742 | 92.383 | 92.2742 | 92.383 | 92.2742 |
| AT | 92.383 | 92.2742 | 92.383 | 92.2742 | 92.383 | 92.2742 |
| AU | 57.2361 | 92.2742 | 57.2361 | 92.2742 | 43.9608 | 92.383 |
| AV | 92.383 | 92.2742 | 92.383 | 92.2742 | 92.383 | 92.2742 |
| AW | 92.383 | 92.2742 | 92.383 | 92.2742 | 92.383 | 92.2742 |
| AX | 92.383 | 92.2742 | 92.383 | 92.2742 | 92.383 | 92.2742 |
| AY | 27.4211 | 88.5745 | 27.4211 | 88.5745 | 36.3439 | 88.5745 |
| AZ | 92.383 | 92.2742 | 92.383 | 92.2742 | 92.383 | 92.2742 |
| BA | 93.0359 | 92.2742 | 93.0359 | 92.2742 | 93.0359 | 92.2742 |
| BB | 92.383 | 92.2742 | 92.383 | 92.2742 | 92.383 | 92.2742 |
| BC | 92.383 | 92.2742 | 92.383 | 92.2742 | 92.383 | 92.2742 |
| BD | 92.383 | 92.2742 | 92.383 | 92.2742 | 92.383 | 92.2742 |
| BE | 92.383 | 92.2742 | 92.383 | 92.2742 | 92.383 | 92.2742 |
| BF | 92.383 | 92.2742 | 92.383 | 92.2742 | 92.383 | 92.2742 |
| BG | 92.383 | 92.2742 | 92.383 | 92.2742 | 92.383 | 92.2742 |
| BH | 92.383 | 92.2742 | 92.383 | 92.2742 | 92.383 | 92.2742 |
| BI | 92.383 | 92.2742 | 92.383 | 92.2742 | 92.383 | 92.2742 |
| BJ | 92.383 | 92.2742 | 92.383 | 92.2742 | 92.383 | 92.2742 |
| BK | 92.383 | 92.2742 | 92.383 | 92.2742 | 92.383 | 92.2742 |
| BL | 92.383 | 91.0773 | 92.383 | 91.0773 | 92.383 | 91.0773 |
| BM | 92.383 | 92.2742 | 92.383 | 92.2742 | 92.383 | 92.2742 |
| BN | 92.383 | 92.2742 | 92.383 | 92.2742 | 92.383 | 92.2742 |
| BO | 92.383 | 92.2742 | 92.383 | 92.2742 | 92.383 | 92.2742 |
| BP | 92.383 | 92.2742 | 92.383 | 92.2742 | 92.383 | 92.2742 |
| BQ | 92.383 | 92.2742 | 92.383 | 92.2742 | 92.383 | 92.2742 |
| BR | 92.383 | 92.2742 | 92.383 | 92.2742 | 92.383 | 92.2742 |
| BS | 92.383 | 92.2742 | 92.383 | 92.2742 | 92.383 | 92.2742 |
| BT | 92.383 | 92.2742 | 92.383 | 92.2742 | 92.383 | 92.2742 |
| BU | 92.383 | 92.2742 | 92.383 | 92.2742 | 92.383 | 92.2742 |
| BV | 92.383 | 92.2742 | 92.383 | 92.2742 | 92.383 | 92.2742 |
| BW | 92.383 | 92.2742 | 92.383 | 92.2742 | 92.383 | 92.2742 |
| BX | 92.383 | 92.2742 | 92.383 | 92.2742 | 92.383 | 92.2742 |
| BY | 92.383 | 92.2742 | 92.383 | 92.2742 | 92.383 | 92.2742 |
| BZ | 92.383 | 92.2742 | 92.383 | 92.2742 | 92.383 | 92.2742 |
| CA | 92.383 | 92.2742 | 92.383 | 92.2742 | 92.383 | 92.2742 |
| CB | 92.383 | 92.2742 | 92.383 | 92.2742 | 92.383 | 92.2742 |
| CC | 92.383 | 92.2742 | 92.383 | 92.2742 | 92.383 | 92.2742 |
| CD | 92.383 | 92.2742 | 92.383 | 92.2742 | 92.383 | 92.2742 |
| CE | 92.383 | 92.2742 | 92.383 | 92.2742 | 92.383 | 92.2742 |
| CF | 92.383 | 92.2742 | 92.383 | 92.2742 | 92.383 | 92.2742 |
| CG | 92.383 | 92.2742 | 92.383 | 92.2742 | 92.383 | 92.2742 |
| CH | 92.383 | 92.2742 | 92.383 | 92.2742 | 92.383 | 92.2742 |
| CI | 92.383 | 91.0773 | 92.383 | 91.0773 | 92.383 | 91.0773 |
| CJ | 92.383 | 91.0773 | 92.383 | 91.0773 | 92.383 | 91.0773 |
| CK | 92.383 | 92.2742 | 92.383 | 91.0773 | 92.383 | 92.2742 |
| CL | 92.383 | 92.2742 | 92.383 | 91.0773 | 92.383 | 92.2742 |
| CM | 92.383 | 92.2742 | 92.383 | 92.2742 | 92.383 | 92.2742 |
| CN | 92.383 | 92.2742 | 92.383 | 92.2742 | 92.383 | 92.2742 |
| CO | 92.383 | 92.2742 | 92.383 | 92.2742 | 92.383 | 92.2742 |
| CP | 92.383 | 92.2742 | 92.383 | 92.2742 | 92.383 | 92.2742 |
| CQ | 92.383 | 92.2742 | 92.383 | 92.2742 | 92.383 | 92.2742 |
| CR | 92.383 | 92.2742 | 92.383 | 92.2742 | 92.383 | 92.2742 |
| CS | 92.383 | 92.2742 | 92.383 | 92.2742 | 92.383 | 92.2742 |
| CT | 92.383 | 92.2742 | 92.383 | 92.2742 | 92.383 | 92.2742 |
| CU | 92.383 | 92.2742 | 92.383 | 92.2742 | 92.383 | 92.2742 |
| CV | 92.383 | 92.2742 | 92.383 | 92.2742 | 92.383 | 92.2742 |
| CW | 92.383 | 92.2742 | 92.383 | 92.2742 | 92.383 | 92.2742 |
| CX | 92.383 | 92.2742 | 92.383 | 92.2742 | 92.383 | 92.2742 |
| CY | 92.383 | 92.2742 | 92.383 | 92.2742 | 92.383 | 92.2742 |
| CZ | 92.383 | 92.2742 | 92.383 | 92.2742 | 92.383 | 92.2742 |
| DA | 92.383 | 91.0773 | 92.383 | 91.0773 | 92.383 | 91.0773 |
| DB | 92.383 | 92.2742 | 92.383 | 92.2742 | 92.383 | 92.2742 |
| DC | 92.383 | 91.0773 | 92.383 | 91.0773 | 92.383 | 91.0773 |
| DD | 92.383 | 92.2742 | 92.383 | 92.2742 | 92.383 | 92.2742 |
| DE | 92.383 | 92.2742 | 92.383 | 92.2742 | 92.383 | 92.2742 |
| DF | 92.383 | 91.0773 | 92.383 | 91.0773 | 92.383 | 91.0773 |
| DG | 92.383 | 92.2742 | 92.383 | 92.2742 | 92.383 | 92.2742 |
| DH | 92.383 | 92.2742 | 92.383 | 92.2742 | 92.383 | 92.2742 |
| DI | 92.383 | 91.0773 | 92.383 | 91.0773 | 92.383 | 91.0773 |
| DJ | 92.383 | 91.0773 | 92.383 | 91.0773 | 92.383 | 91.0773 |
| DK | 92.383 | 91.0773 | 92.383 | 91.0773 | 92.383 | 91.0773 |
| DL | 92.383 | 91.0773 | 92.383 | 91.0773 | 92.383 | 91.0773 |
| DM | 92.383 | 92.2742 | 92.383 | 92.2742 | 92.383 | 92.2742 |
| DN | 92.383 | 92.2742 | 92.383 | 92.2742 | 92.383 | 92.2742 |
| DO | 92.383 | 92.2742 | 92.383 | 92.2742 | 92.383 | 92.2742 |
| DP | 92.383 | 91.0773 | 92.383 | 91.0773 | 92.383 | 91.0773 |
| DQ | 92.383 | 91.0773 | 92.383 | 91.0773 | 92.383 | 91.0773 |
| DR | 92.383 | 91.0773 | 92.383 | 91.0773 | 92.383 | 91.0773 |
| DS | 92.383 | 92.2742 | 92.383 | 92.2742 | 92.383 | 92.2742 |
| DT | 92.383 | 92.2742 | 92.383 | 92.2742 | 92.383 | 92.2742 |
| DU | 92.383 | 92.2742 | 92.383 | 92.2742 | 92.383 | 92.2742 |
| DV | 92.383 | 92.2742 | 92.383 | 92.2742 | 92.383 | 92.2742 |
| DW | 92.383 | 92.2742 | 92.383 | 92.2742 | 92.383 | 92.2742 |
| DX | 92.383 | 92.2742 | 92.383 | 92.2742 | 92.383 | 92.2742 |
| DY | 92.383 | 92.2742 | 92.383 | 92.2742 | 92.383 | 92.2742 |
| DZ | 92.383 | 92.2742 | 92.383 | 92.2742 | 92.383 | 92.2742 |
| EA | 92.383 | 92.2742 | 92.383 | 92.2742 | 92.383 | 92.2742 |
| EB | 92.383 | 92.2742 | 92.383 | 92.2742 | 92.383 | 92.2742 |
| EC | 92.383 | 92.2742 | 92.383 | 92.2742 | 92.383 | 92.2742 |
| ED | 92.383 | 92.2742 | 92.383 | 92.2742 | 92.383 | 92.2742 |
| EE | 92.383 | 92.2742 | 92.383 | 92.2742 | 92.383 | 92.2742 |
| EF | 92.383 | 92.2742 | 92.383 | 92.2742 | 92.383 | 92.2742 |
| EG | 92.383 | 92.2742 | 92.383 | 92.2742 | 92.383 | 92.2742 |
| EH | 92.383 | 92.2742 | 92.383 | 92.2742 | 92.383 | 92.2742 |
| EI | 92.383 | 92.2742 | 92.383 | 92.2742 | 92.383 | 92.2742 |
| EJ | 92.383 | 92.2742 | 92.383 | 92.2742 | 92.383 | 92.2742 |
| EK | 92.383 | 92.2742 | 92.383 | 92.2742 | 92.383 | 92.2742 |
| EL | 92.383 | 92.2742 | 92.383 | 92.2742 | 92.383 | 92.2742 |
| EM | 92.383 | 92.2742 | 92.383 | 92.2742 | 92.383 | 92.2742 |
| EN | 92.383 | 92.2742 | 92.383 | 92.2742 | 92.383 | 92.2742 |
| EO | 92.383 | 92.2742 | 92.383 | 92.2742 | 92.383 | 92.2742 |
| EP | 21.5452 | 62.1328 | 21.5452 | 62.1328 | 21.5452 | 62.1328 |
| EQ | 92.383 | 92.2742 | 92.383 | 92.2742 | 92.383 | 92.2742 |
| ER | 92.383 | 92.2742 | 92.383 | 92.2742 | 92.383 | 92.2742 |
| ES | 92.383 | 91.0773 | 92.383 | 91.0773 | 92.383 | 91.0773 |
| ET | 92.383 | 92.2742 | 92.383 | 92.2742 | 92.383 | 92.2742 |
| EU | 93.9064 | 92.2742 | 93.9064 | 92.2742 | 93.9064 | 92.2742 |
| EV | 92.383 | 92.2742 | 92.383 | 92.2742 | 92.383 | 92.2742 |
| EW | 92.383 | 92.2742 | 92.383 | 92.2742 | 92.383 | 92.2742 |
| EX | 92.383 | 91.0773 | 92.383 | 91.0773 | 92.383 | 91.0773 |
| EY | 92.383 | 92.2742 | 92.383 | 92.2742 | 92.383 | 92.2742 |
| EZ | 67.0294 | 92.383 | 67.0294 | 92.383 | 67.0294 | 92.383 |
| FA | 92.383 | 92.2742 | 92.383 | 92.2742 | 92.383 | 92.2742 |
| FB | 92.383 | 92.2742 | 92.383 | 92.2742 | 92.383 | 92.2742 |
| FC | 92.383 | 92.2742 | 92.383 | 92.2742 | 92.383 | 92.2742 |
| FD | 92.2742 | 92.2742 | 92.2742 | 92.2742 | 92.2742 | 92.2742 |
| FE | 92.2742 | 92.2742 | 92.2742 | 92.2742 | 92.2742 | 92.2742 |
| FF | 92.383 | 92.2742 | 92.383 | 92.2742 | 92.383 | 92.2742 |
| FG | 92.383 | 92.2742 | 92.383 | 92.2742 | 92.383 | 92.2742 |
| FH | 92.383 | 91.0773 | 92.383 | 91.0773 | 92.383 | 91.0773 |
| FI | 92.383 | 92.2742 | 92.383 | 92.2742 | 92.383 | 92.2742 |
| FJ | 92.383 | 92.2742 | 92.383 | 92.2742 | 92.383 | 92.2742 |
| FK | 92.383 | 91.0773 | 92.383 | 91.0773 | 92.383 | 91.0773 |
| FL | 93.9064 | 92.2742 | 93.9064 | 92.2742 | 93.9064 | 92.2742 |
| FM | 92.383 | 92.2742 | 92.383 | 92.2742 | 92.383 | 92.2742 |
| FN | 92.383 | 91.0773 | 92.383 | 91.0773 | 92.383 | 91.0773 |
| FO | 92.383 | 91.0773 | 92.383 | 91.0773 | 92.383 | 91.0773 |
| FP | 93.9064 | 92.2742 | 93.9064 | 92.2742 | 93.9064 | 92.2742 |
| FQ | 92.383 | 92.2742 | 92.383 | 92.2742 | 92.383 | 92.2742 |
| FR | 92.383 | 91.0773 | 92.383 | 91.0773 | 92.383 | 91.0773 |
| FS | 92.383 | 92.2742 | 92.383 | 92.2742 | 92.383 | 92.2742 |
| FT | 92.383 | 92.2742 | 92.383 | 92.2742 | 92.383 | 92.2742 |
| FU | 92.383 | 92.2742 | 92.383 | 92.2742 | 92.383 | 92.2742 |
| FV | 92.383 | 92.2742 | 92.383 | 92.2742 | 92.383 | 92.2742 |
| FW | 92.383 | 92.2742 | 92.383 | 92.2742 | 92.383 | 92.2742 |
| FX | 92.383 | 92.2742 | 92.383 | 92.2742 | 92.383 | 92.2742 |
| FY | 92.383 | 92.2742 | 92.383 | 92.2742 | 92.383 | 92.2742 |
| FZ | 92.383 | 92.2742 | 92.383 | 92.2742 | 92.383 | 92.2742 |
| GA | 92.383 | 92.2742 | 92.383 | 92.2742 | 92.383 | 92.2742 |
| GB | 92.383 | 92.2742 | 92.383 | 92.2742 | 92.383 | 92.2742 |
| GC | 92.383 | 92.2742 | 92.383 | 92.2742 | 92.383 | 92.2742 |
| GD | 92.383 | 92.2742 | 92.383 | 92.2742 | 92.383 | 92.2742 |
| GE | 92.383 | 92.2742 | 92.383 | 92.2742 | 92.383 | 92.2742 |
| GF | 57.4538 | 85.3101 | 57.4538 | 92.2742 | 44.2873 | 92.2742 |
| GG | 92.383 | 92.2742 | 92.383 | 92.2742 | 92.383 | 92.2742 |
| GH | 92.383 | 92.2742 | 92.383 | 92.2742 | 92.383 | 92.2742 |
| GI | 92.383 | 91.0773 | 92.383 | 91.0773 | 92.383 | 91.0773 |
| GJ | 92.383 | 92.2742 | 92.383 | 92.2742 | 92.383 | 92.2742 |
| GK | 92.383 | 92.2742 | 92.383 | 92.2742 | 92.383 | 92.2742 |
| GL | 92.383 | 92.2742 | 92.383 | 92.2742 | 92.383 | 92.2742 |
| GM | 92.383 | 92.2742 | 92.383 | 92.2742 | 92.383 | 92.2742 |
| GN | 92.383 | 91.0773 | 92.383 | 91.0773 | 92.383 | 91.0773 |
| GO | 92.383 | 92.2742 | 92.383 | 92.2742 | 92.383 | 92.2742 |
| GP | 93.6888 | 92.2742 | 93.6888 | 92.2742 | 93.6888 | 92.2742 |
| GQ | 92.7095 | 92.2742 | 92.7095 | 92.2742 | 82.5898 | 92.2742 |
| GR | 92.8183 | 92.2742 | 92.8183 | 92.2742 | 92.8183 | 92.2742 |
| GS | 92.383 | 92.2742 | 92.383 | 92.2742 | 92.383 | 92.2742 |
| GT | 92.383 | 92.2742 | 92.383 | 92.2742 | 92.383 | 92.2742 |
| GU | 94.3417 | 92.2742 | 93.9064 | 92.2742 | 76.605 | 92.2742 |
| GV | 92.383 | 92.2742 | 92.383 | 92.2742 | 92.383 | 92.2742 |
| GW | 92.383 | 92.2742 | 92.383 | 92.2742 | 92.383 | 92.2742 |
| GX | 92.383 | 92.2742 | 92.383 | 92.2742 | 92.383 | 92.2742 |
| GY | 92.383 | 92.2742 | 92.383 | 92.2742 | 92.383 | 92.2742 |
| GZ | 92.383 | 92.2742 | 92.383 | 92.2742 | 92.383 | 92.2742 |
| HA | 92.383 | 92.2742 | 92.383 | 92.2742 | 92.383 | 92.2742 |
| HB | 92.383 | 92.2742 | 92.383 | 92.2742 | 92.383 | 92.2742 |
| HC | 92.383 | 92.2742 | 92.383 | 92.2742 | 92.383 | 92.2742 |
| HD | 92.383 | 92.2742 | 92.383 | 92.2742 | 92.383 | 92.2742 |
| HE | 92.383 | 92.2742 | 92.383 | 92.2742 | 92.383 | 92.2742 |
| HF | 92.383 | 92.2742 | 92.383 | 92.2742 | 92.383 | 92.2742 |
| HG | 92.383 | 92.2742 | 92.383 | 92.2742 | 92.383 | 92.2742 |

table 7. classification on FACTORY data using linear

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  |  |
| kernel | rbf |  |  |  |  |  |
| c | c=0.5 |  | c=1 |  | c=5 |  |
| features | single | reverse | single | reverse | single | reverse |
| A | 95.8651 | 92.383 | 95.8651 | 92.383 | 95.8651 | 92.383 |
| B | 92.383 | 92.383 | 92.383 | 92.383 | 94.8857 | 92.383 |
| C | 92.383 | 92.383 | 92.383 | 92.383 | 92.383 | 92.383 |
| D | 92.383 | 92.383 | 92.383 | 92.383 | 92.383 | 92.383 |
| E | 93.6888 | 92.383 | 93.6888 | 92.383 | 93.6888 | 92.383 |
| F | 95.8651 | 92.383 | 95.8651 | 92.383 | 95.8651 | 92.383 |
| G | 93.9064 | 92.383 | 93.9064 | 92.383 | 93.9064 | 92.383 |
| H | 92.383 | 92.383 | 92.383 | 92.383 | 92.383 | 92.383 |
| I | 92.383 | 92.383 | 92.383 | 92.383 | 92.383 | 92.383 |
| J | 92.383 | 92.383 | 92.383 | 92.383 | 92.383 | 92.383 |
| K | 95.8651 | 92.383 | 95.8651 | 92.383 | 95.8651 | 92.383 |
| L | 92.383 | 92.383 | 92.383 | 92.383 | 92.383 | 92.383 |
| M | 92.383 | 92.383 | 92.383 | 92.383 | 92.383 | 92.383 |
| N | 92.383 | 92.383 | 92.383 | 92.383 | 92.383 | 92.383 |
| O | 92.383 | 92.383 | 92.383 | 92.383 | 92.383 | 92.383 |
| P | 92.383 | 92.383 | 92.383 | 92.383 | 92.383 | 92.383 |
| Q | 92.383 | 92.383 | 92.383 | 92.383 | 92.383 | 92.383 |
| R | 92.383 | 92.383 | 92.383 | 92.383 | 94.7769 | 92.383 |
| S | 92.383 | 92.383 | 92.383 | 92.383 | 92.383 | 92.383 |
| T | 91.9478 | 92.383 | 92.0566 | 92.383 | 91.1861 | 92.383 |
| U | 92.383 | 92.383 | 92.383 | 92.383 | 92.383 | 92.383 |
| V | 92.383 | 92.383 | 92.383 | 92.383 | 92.383 | 92.383 |
| W | 92.383 | 92.383 | 92.1654 | 92.383 | 92.0566 | 92.383 |
| X | 92.383 | 92.383 | 92.383 | 92.383 | 92.383 | 92.383 |
| Y | 92.383 | 92.383 | 92.383 | 92.383 | 92.383 | 92.383 |
| Z | 92.383 | 92.383 | 92.383 | 92.383 | 92.383 | 92.383 |
| AA | 92.383 | 92.383 | 92.383 | 92.383 | 92.383 | 92.383 |
| AB | 92.383 | 92.383 | 92.383 | 92.383 | 92.383 | 92.383 |
| AC | 92.383 | 92.383 | 92.383 | 92.383 | 92.383 | 92.383 |
| AD | 92.383 | 92.383 | 92.383 | 92.383 | 92.383 | 92.383 |
| AE | 92.383 | 92.383 | 92.383 | 92.383 | 92.0566 | 92.383 |
| AF | 93.6888 | 92.383 | 93.6888 | 92.383 | 93.6888 | 92.383 |
| AG | 100 | 92.383 | 100 | 92.383 | 100 | 92.383 |
| AH | 92.383 | 92.383 | 92.383 | 92.383 | 92.383 | 92.383 |
| AI | 92.383 | 92.383 | 92.383 | 92.383 | 92.383 | 92.383 |
| AJ | 92.383 | 92.383 | 91.7301 | 92.383 | 89.6627 | 92.383 |
| AK | 94.4505 | 92.383 | 95.4298 | 92.383 | 95.4298 | 92.383 |
| AL | 92.383 | 92.383 | 92.383 | 92.383 | 92.383 | 92.383 |
| AM | 92.383 | 92.383 | 92.383 | 92.383 | 92.383 | 92.383 |
| AN | 92.383 | 92.383 | 92.383 | 92.383 | 92.383 | 92.383 |
| AO | 93.6888 | 92.383 | 93.6888 | 92.383 | 93.6888 | 92.383 |
| AP | 92.383 | 92.383 | 91.2949 | 92.383 | 90.8596 | 92.383 |
| AQ | 93.6888 | 92.383 | 93.6888 | 92.383 | 93.6888 | 92.383 |
| AR | 92.383 | 92.383 | 92.383 | 92.383 | 92.383 | 92.383 |
| AS | 92.383 | 92.383 | 92.383 | 92.383 | 92.383 | 92.383 |
| AT | 92.383 | 92.383 | 92.383 | 92.383 | 92.383 | 92.383 |
| AU | 90.2067 | 92.383 | 80.8487 | 92.383 | 80.8487 | 92.383 |
| AV | 92.383 | 92.383 | 92.383 | 92.383 | 92.383 | 92.383 |
| AW | 92.383 | 92.383 | 92.383 | 92.383 | 92.383 | 92.383 |
| AX | 92.383 | 92.383 | 92.383 | 92.383 | 92.383 | 92.383 |
| AY | 93.9064 | 92.383 | 93.9064 | 92.383 | 93.9064 | 92.383 |
| AZ | 92.383 | 92.383 | 91.0773 | 92.383 | 90.642 | 92.383 |
| BA | 93.6888 | 92.383 | 93.6888 | 92.383 | 93.6888 | 92.383 |
| BB | 92.383 | 92.383 | 92.383 | 92.383 | 92.383 | 92.383 |
| BC | 92.383 | 92.383 | 92.383 | 92.383 | 92.383 | 92.383 |
| BD | 92.383 | 92.383 | 92.383 | 92.383 | 92.383 | 92.383 |
| BE | 92.383 | 92.383 | 92.383 | 92.383 | 92.383 | 92.383 |
| BF | 92.383 | 92.383 | 92.383 | 92.383 | 92.383 | 92.383 |
| BG | 92.383 | 92.383 | 92.383 | 92.383 | 92.383 | 92.383 |
| BH | 93.58 | 92.383 | 93.2535 | 92.383 | 92.8183 | 92.383 |
| BI | 92.383 | 92.383 | 92.383 | 92.383 | 92.383 | 92.383 |
| BJ | 92.383 | 92.383 | 92.383 | 92.383 | 92.383 | 92.383 |
| BK | 92.383 | 92.383 | 92.383 | 92.383 | 92.383 | 92.383 |
| BL | 93.1447 | 92.383 | 93.2535 | 92.383 | 93.2535 | 92.383 |
| BM | 92.383 | 92.383 | 92.383 | 92.383 | 92.383 | 92.383 |
| BN | 92.383 | 92.383 | 92.383 | 92.383 | 92.383 | 92.383 |
| BO | 92.383 | 92.383 | 92.383 | 92.383 | 92.383 | 92.383 |
| BP | 92.383 | 92.383 | 92.383 | 92.383 | 92.383 | 92.383 |
| BQ | 92.383 | 92.383 | 92.383 | 92.383 | 92.383 | 92.383 |
| BR | 92.383 | 92.383 | 92.383 | 92.383 | 92.383 | 92.383 |
| BS | 92.383 | 92.383 | 92.383 | 92.383 | 92.383 | 92.383 |
| BT | 92.383 | 92.383 | 92.383 | 92.383 | 92.383 | 92.383 |
| BU | 92.383 | 92.383 | 92.383 | 92.383 | 92.383 | 92.383 |
| BV | 92.383 | 92.383 | 92.383 | 92.383 | 92.383 | 92.383 |
| BW | 92.383 | 92.383 | 92.383 | 92.383 | 92.383 | 92.383 |
| BX | 92.383 | 92.383 | 92.383 | 92.383 | 92.383 | 92.383 |
| BY | 92.383 | 92.383 | 92.383 | 92.383 | 92.383 | 92.383 |
| BZ | 92.383 | 92.383 | 92.383 | 92.383 | 92.383 | 92.383 |
| CA | 92.383 | 92.383 | 92.383 | 92.383 | 92.383 | 92.383 |
| CB | 92.383 | 92.383 | 92.383 | 92.383 | 92.383 | 92.383 |
| CC | 92.383 | 92.383 | 92.383 | 92.383 | 92.383 | 92.383 |
| CD | 92.383 | 92.383 | 92.383 | 92.383 | 92.383 | 92.383 |
| CE | 92.383 | 92.383 | 92.383 | 92.383 | 92.383 | 92.383 |
| CF | 92.383 | 92.383 | 92.383 | 92.383 | 92.383 | 92.383 |
| CG | 92.383 | 92.383 | 92.383 | 92.383 | 92.383 | 92.383 |
| CH | 92.383 | 92.383 | 92.383 | 92.383 | 92.383 | 92.383 |
| CI | 92.383 | 92.383 | 92.383 | 92.383 | 92.383 | 92.383 |
| CJ | 92.383 | 92.383 | 92.383 | 92.383 | 92.383 | 92.383 |
| CK | 92.383 | 92.383 | 92.383 | 92.383 | 92.383 | 92.383 |
| CL | 92.383 | 92.383 | 92.383 | 92.383 | 92.383 | 92.383 |
| CM | 92.383 | 92.383 | 92.383 | 92.383 | 92.383 | 92.383 |
| CN | 92.383 | 92.383 | 92.383 | 92.383 | 92.383 | 92.383 |
| CO | 92.383 | 92.383 | 92.383 | 92.383 | 92.383 | 92.383 |
| CP | 92.383 | 92.383 | 92.383 | 92.383 | 92.383 | 92.383 |
| CQ | 92.383 | 92.383 | 92.383 | 92.383 | 92.383 | 92.383 |
| CR | 92.383 | 92.383 | 92.383 | 92.383 | 92.383 | 92.383 |
| CS | 92.383 | 92.383 | 92.383 | 92.383 | 92.383 | 92.383 |
| CT | 92.383 | 92.383 | 92.383 | 92.383 | 92.383 | 92.383 |
| CU | 92.383 | 92.383 | 92.383 | 92.383 | 92.383 | 92.383 |
| CV | 92.383 | 92.383 | 92.383 | 92.383 | 92.383 | 92.383 |
| CW | 92.383 | 92.383 | 92.383 | 92.383 | 92.383 | 92.383 |
| CX | 92.383 | 92.383 | 92.383 | 92.383 | 92.383 | 92.383 |
| CY | 92.383 | 92.383 | 92.383 | 92.383 | 92.383 | 92.383 |
| CZ | 92.383 | 92.383 | 92.383 | 92.383 | 92.383 | 92.383 |
| DA | 96.518 | 92.383 | 96.3003 | 92.383 | 96.518 | 92.383 |
| DB | 92.383 | 92.383 | 92.383 | 92.383 | 92.383 | 92.383 |
| DC | 92.383 | 92.383 | 93.4712 | 92.383 | 93.3624 | 92.383 |
| DD | 92.383 | 92.383 | 92.383 | 92.383 | 92.383 | 92.383 |
| DE | 92.2742 | 92.383 | 92.383 | 92.383 | 92.1654 | 92.383 |
| DF | 92.383 | 92.383 | 92.383 | 92.383 | 92.383 | 92.383 |
| DG | 92.383 | 92.383 | 92.383 | 92.383 | 92.383 | 92.383 |
| DH | 92.2742 | 92.383 | 92.1654 | 92.383 | 92.4918 | 92.383 |
| DI | 92.1654 | 92.383 | 92.1654 | 92.383 | 91.9478 | 92.383 |
| DJ | 92.2742 | 92.383 | 92.0566 | 92.383 | 91.5125 | 92.383 |
| DK | 92.2742 | 92.383 | 92.0566 | 92.383 | 92.383 | 92.383 |
| DL | 92.383 | 92.383 | 92.1654 | 92.383 | 91.6213 | 92.383 |
| DM | 92.0566 | 92.383 | 91.9478 | 92.383 | 91.6213 | 92.383 |
| DN | 92.383 | 92.383 | 92.383 | 92.383 | 91.9478 | 92.383 |
| DO | 92.383 | 92.383 | 92.383 | 92.383 | 91.7301 | 92.383 |
| DP | 92.0566 | 92.383 | 92.1654 | 92.383 | 91.7301 | 92.383 |
| DQ | 92.7095 | 92.383 | 93.1447 | 92.383 | 93.9064 | 92.383 |
| DR | 92.383 | 92.383 | 92.2742 | 92.383 | 91.839 | 92.383 |
| DS | 92.383 | 92.383 | 92.6007 | 92.383 | 92.4918 | 92.383 |
| DT | 92.2742 | 92.383 | 92.0566 | 92.383 | 91.5125 | 92.383 |
| DU | 91.9478 | 92.383 | 92.0566 | 92.383 | 91.5125 | 92.383 |
| DV | 92.0566 | 92.383 | 92.2742 | 92.383 | 91.4037 | 92.383 |
| DW | 92.383 | 92.383 | 92.7095 | 92.383 | 92.2742 | 92.383 |
| DX | 92.2742 | 92.383 | 91.839 | 92.383 | 91.4037 | 92.383 |
| DY | 92.383 | 92.383 | 92.383 | 92.383 | 92.383 | 92.383 |
| DZ | 92.383 | 92.383 | 92.383 | 92.383 | 92.383 | 92.383 |
| EA | 92.383 | 92.383 | 92.383 | 92.383 | 92.383 | 92.383 |
| EB | 92.383 | 92.383 | 92.383 | 92.383 | 92.383 | 92.383 |
| EC | 92.383 | 92.383 | 92.383 | 92.383 | 92.383 | 92.383 |
| ED | 92.383 | 92.383 | 92.383 | 92.383 | 92.383 | 92.383 |
| EE | 92.383 | 92.383 | 92.383 | 92.383 | 92.383 | 92.383 |
| EF | 92.383 | 92.383 | 92.383 | 92.383 | 92.383 | 92.383 |
| EG | 92.383 | 92.383 | 92.383 | 92.383 | 92.383 | 92.383 |
| EH | 92.383 | 92.383 | 92.383 | 92.383 | 92.383 | 92.383 |
| EI | 92.383 | 92.383 | 92.383 | 92.383 | 92.383 | 92.383 |
| EJ | 92.383 | 92.383 | 92.383 | 92.383 | 92.383 | 92.383 |
| EK | 92.383 | 92.383 | 92.383 | 92.383 | 92.383 | 92.383 |
| EL | 92.383 | 92.383 | 92.383 | 92.383 | 92.383 | 92.383 |
| EM | 92.383 | 92.383 | 92.383 | 92.383 | 92.383 | 92.383 |
| EN | 92.383 | 92.383 | 92.383 | 92.383 | 92.383 | 92.383 |
| EO | 92.383 | 92.383 | 92.383 | 92.383 | 92.383 | 92.383 |
| EP | 93.9064 | 92.383 | 92.383 | 92.383 | 92.383 | 92.383 |
| EQ | 92.383 | 92.383 | 92.383 | 92.383 | 92.383 | 92.383 |
| ER | 92.383 | 92.383 | 92.383 | 92.383 | 92.383 | 92.383 |
| ES | 92.383 | 92.383 | 92.383 | 92.383 | 92.383 | 92.383 |
| ET | 92.383 | 92.383 | 92.383 | 92.383 | 92.383 | 92.383 |
| EU | 93.9064 | 92.383 | 93.9064 | 92.383 | 93.9064 | 92.383 |
| EV | 92.383 | 92.383 | 92.383 | 92.383 | 92.383 | 92.383 |
| EW | 93.9064 | 92.383 | 93.9064 | 92.383 | 93.9064 | 92.383 |
| EX | 92.383 | 92.383 | 92.383 | 92.383 | 92.383 | 92.383 |
| EY | 92.383 | 92.383 | 92.383 | 92.383 | 92.383 | 92.383 |
| EZ | 94.124 | 92.383 | 94.124 | 92.383 | 94.124 | 92.383 |
| FA | 92.383 | 92.383 | 92.383 | 92.383 | 92.383 | 92.383 |
| FB | 92.383 | 92.383 | 92.383 | 92.383 | 92.383 | 92.383 |
| FC | 92.383 | 92.383 | 92.383 | 92.383 | 92.383 | 92.383 |
| FD | 93.9064 | 92.383 | 93.9064 | 92.383 | 93.9064 | 92.383 |
| FE | 93.9064 | 92.383 | 93.9064 | 92.383 | 93.9064 | 92.383 |
| FF | 92.383 | 92.383 | 92.383 | 92.383 | 92.383 | 92.383 |
| FG | 92.383 | 92.383 | 92.383 | 92.383 | 92.383 | 92.383 |
| FH | 92.383 | 92.383 | 92.0566 | 92.383 | 93.9064 | 92.383 |
| FI | 92.383 | 92.383 | 92.383 | 92.383 | 92.383 | 92.383 |
| FJ | 92.383 | 92.383 | 92.383 | 92.383 | 92.383 | 92.383 |
| FK | 92.383 | 92.383 | 92.383 | 92.383 | 92.383 | 92.383 |
| FL | 93.9064 | 92.383 | 93.9064 | 92.383 | 93.9064 | 92.383 |
| FM | 93.9064 | 92.383 | 93.9064 | 92.383 | 93.9064 | 92.383 |
| FN | 93.9064 | 92.383 | 93.9064 | 92.383 | 93.9064 | 92.383 |
| FO | 92.383 | 92.383 | 92.383 | 92.383 | 92.383 | 92.383 |
| FP | 93.9064 | 92.383 | 93.9064 | 92.383 | 93.9064 | 92.383 |
| FQ | 93.9064 | 92.383 | 93.9064 | 92.383 | 93.9064 | 92.383 |
| FR | 93.9064 | 92.383 | 93.9064 | 92.383 | 93.9064 | 92.383 |
| FS | 92.383 | 92.383 | 92.383 | 92.383 | 92.383 | 92.383 |
| FT | 92.383 | 92.383 | 92.383 | 92.383 | 92.383 | 92.383 |
| FU | 92.383 | 92.383 | 92.383 | 92.383 | 92.383 | 92.383 |
| FV | 92.383 | 92.383 | 92.383 | 92.383 | 92.383 | 92.383 |
| FW | 92.383 | 92.383 | 92.383 | 92.383 | 92.383 | 92.383 |
| FX | 92.383 | 92.383 | 92.383 | 92.383 | 92.383 | 92.383 |
| FY | 92.383 | 92.383 | 92.383 | 92.383 | 92.383 | 92.383 |
| FZ | 92.383 | 92.383 | 92.383 | 92.383 | 92.383 | 92.383 |
| GA | 92.383 | 92.383 | 92.383 | 92.383 | 92.383 | 92.383 |
| GB | 92.383 | 92.383 | 92.383 | 92.383 | 92.383 | 92.383 |
| GC | 92.383 | 92.383 | 92.383 | 92.383 | 92.383 | 92.383 |
| GD | 92.383 | 92.383 | 92.383 | 92.383 | 92.383 | 92.383 |
| GE | 92.383 | 92.383 | 92.383 | 92.383 | 92.383 | 92.383 |
| GF | 90.2067 | 92.383 | 81.0664 | 92.383 | 81.0664 | 92.383 |
| GG | 92.383 | 92.383 | 92.383 | 92.383 | 92.383 | 92.383 |
| GH | 92.383 | 92.383 | 92.383 | 92.383 | 92.383 | 92.383 |
| GI | 92.383 | 92.383 | 92.383 | 92.383 | 92.383 | 92.383 |
| GJ | 92.383 | 92.383 | 92.383 | 92.383 | 92.383 | 92.383 |
| GK | 92.383 | 92.383 | 92.383 | 92.383 | 92.383 | 92.383 |
| GL | 92.383 | 92.383 | 92.383 | 92.383 | 92.383 | 92.383 |
| GM | 92.383 | 92.383 | 92.383 | 92.383 | 92.383 | 92.383 |
| GN | 92.383 | 92.383 | 92.383 | 92.383 | 92.383 | 92.383 |
| GO | 92.383 | 92.383 | 92.383 | 92.383 | 92.383 | 92.383 |
| GP | 92.9271 | 92.383 | 92.9271 | 92.383 | 92.9271 | 92.383 |
| GQ | 92.383 | 92.383 | 92.0566 | 92.383 | 92.1654 | 92.383 |
| GR | 92.7095 | 92.383 | 93.1447 | 92.383 | 93.0359 | 92.383 |
| GS | 93.58 | 92.383 | 93.4712 | 92.383 | 93.4712 | 92.383 |
| GT | 92.383 | 92.383 | 92.2742 | 92.383 | 92.2742 | 92.383 |
| GU | 93.6888 | 92.383 | 94.0152 | 92.383 | 94.0152 | 92.383 |
| GV | 92.383 | 92.383 | 92.383 | 92.383 | 92.383 | 92.383 |
| GW | 92.383 | 92.383 | 92.383 | 92.383 | 92.383 | 92.383 |
| GX | 92.383 | 92.383 | 92.383 | 92.383 | 92.383 | 92.383 |
| GY | 92.383 | 92.383 | 92.383 | 92.383 | 92.383 | 92.383 |
| GZ | 92.383 | 92.383 | 92.383 | 92.383 | 92.383 | 92.383 |
| HA | 92.383 | 92.383 | 92.383 | 92.383 | 92.383 | 92.383 |
| HB | 92.383 | 92.383 | 92.383 | 92.383 | 92.383 | 92.383 |
| HC | 92.383 | 92.383 | 92.383 | 92.383 | 92.383 | 92.383 |
| HD | 92.383 | 92.383 | 92.383 | 92.383 | 92.383 | 92.383 |
| HE | 92.383 | 92.383 | 92.383 | 92.383 | 92.383 | 92.383 |
| HF | 92.383 | 92.383 | 92.383 | 92.383 | 92.383 | 92.383 |
| HG | 92.383 | 92.383 | 92.383 | 92.383 | 92.383 | 92.383 |

table 8. classification on FACTORY data using rbf

Thus, we find the most important feature AG, it can predict the result exactly right. In the origin data set, feature AG took two values: 10 or 11. When feature AG took 10, the product was qualified, while it took 11, the product was with defect (error code was omitted).

We also used the same strategy above to find the feature combination. As seen in the table below, the feature F, K, DA performs better in the “single” training experiment respectively, so we set these 3 features as the initial subset to iterate. The result tested our hypothesis. After iterating using the strategy introduced by the reference, we got the combination of features F, K, DA, as the subset with greatest impact.

# 6.PARAMETER OPTIMIZATION

After getting the result and feature combination, we used R to optimize the parameter, which were c and gamma. R studio and package e1071 offered a simple method to do this job.

We had tried about 50 sets of parameters. Here’s the best subset of the whole combinations and the best one.

Gamma cost error dispersion

1e-03 10 0.5032217 0.04674353

1e-02 10 0.4651985 0.04579680

1e-01 10 0.4304053 0.02701354

1e+00 10 0.3286104 0.03011281

1e+01 10 0.3137883 0.02487371

1e-03 100 0.4639082 0.04446581

1e-02 100 0.4677626 0.04556134

1e-01 100 0.3930314 0.03302645

1e+00 100 0.3144458 0.02186001

1e+01 100 0.3169810 0.02456874

. . . .

1e+00 1000 0.3035874 0.02145914

From the data above, we can observe several sets of parameter. In our optimization process, we tried different combinations of c and gamma to get the minimum error. Finally, we get gamma = 10 and c =10 or gamma = 1 and cost = 1000, as the best combination of parameters.

# 7.CONCLUSION

In this project, we used several classifiers, such as SVM, MaxEnt and KNN, to test and get the best feature subsets that effect the quality of the product. Alas, we got the result via a series of smart strategy including greedy algorithm and iterating algorithm. Finally, the subset of features that have the greatest impact on the quality of the product was found.

We select several classifiers as candidates to do the main job of classifying the data, after designing the steps of necessary processing and cleaning of data. Then we clean the data for the classifiers and use the one with best performance (SVM) to do the main job.

After trying different combinations of parameters with the same corpus, we have got the best parameter combination - gamma = 10 and c =10 or gamma = 1 and cost = 1000.

We judged the performance of each feature or subset of features by “correct rate”. It is the accuracy of the classifier to tell the good products from the bad ones.

We conduct the classification experiment to get the best combination of features. The feature of the greatest impact of ICI data is N (correct rate over 55%), while the combination of features of the greatest impact is K, N and V (correct rate over 82%).

The feature of the greatest impact of Factory data is AG (correct rate equals to 100%), while the combination of features of the greatest impact is F, K and DA (on the condition feature AG was omitted and correct rate over 95%).

# 8.REFERENCE

[1] Hai Zhao, Xiaotian Zhang, Chunyu Kit. 2013. Integrative Semantic Dependency Parsing via Efficient Large-scale Feature Selection. In Journal of Artificial Intelligence Research 46 (2013) 203-233.

[2] J. A. K. Suykens, Support vector machines: A nonlinear modelling and control perspective, Eur. J. Control 2001, 7, 311-327.

[3] https://en.wikipedia.org/wiki/Cross-validation\_(statistics)