Learning-Based Indoor Localization by Using Wi-Fi CSI

Introduction

There are training stage and testing stage in the CSI localization system. The training stage is used to build the database model, and make the system predict the localization during the testing stage.

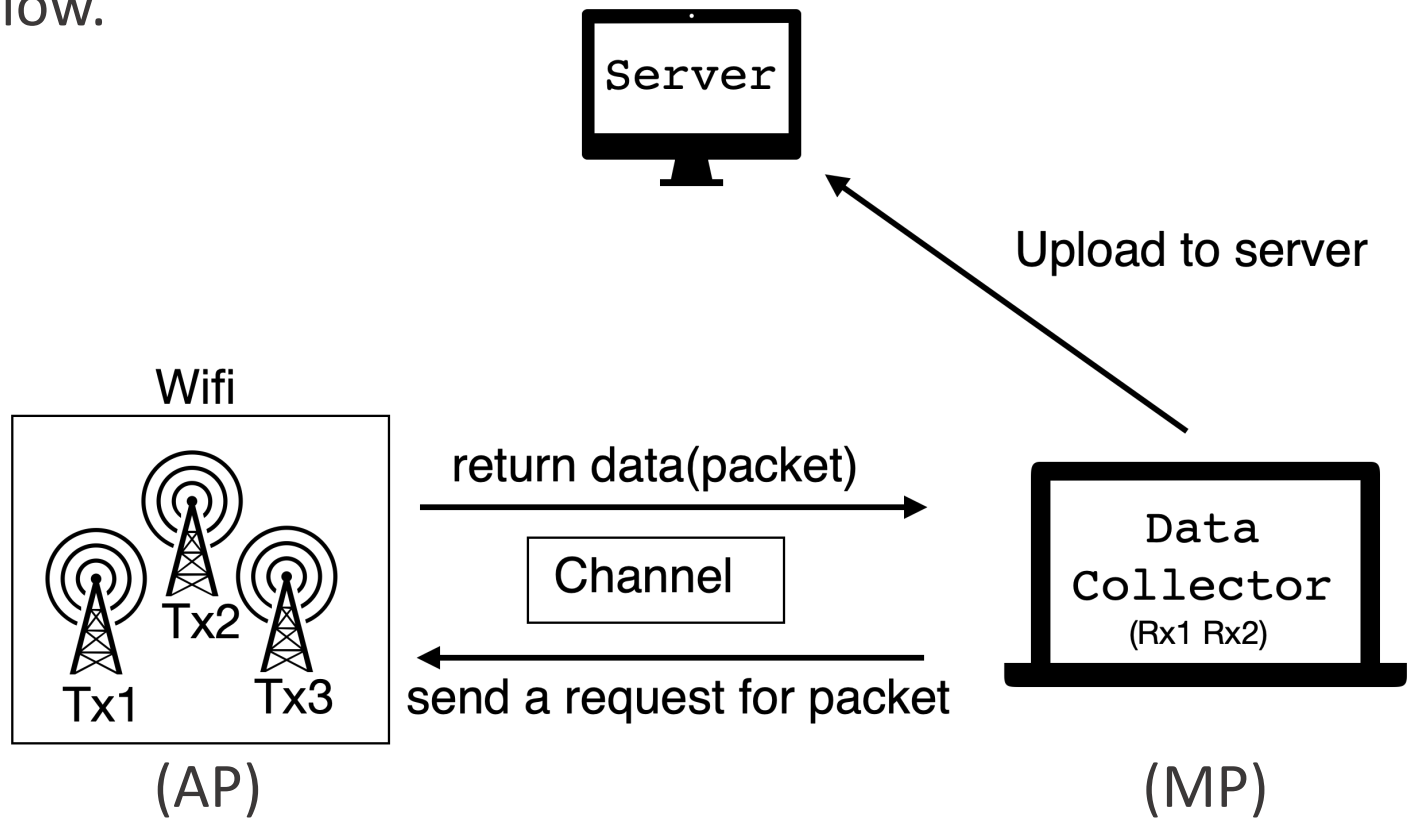
With the popularity of wireless networks and mobile devices today, CSI is positioned Technology has grown at this stage, medical long-term care (elderly tracking), factory warehousing (asset monitoring), airport station (routing Citation), department stores (indoor shopping guide)... etc. are indoor fixed application of bit technology.

Project Goal

Through the combination of WIFI and CSI, the information collected by the AP is used to accurately locate the location of the person.

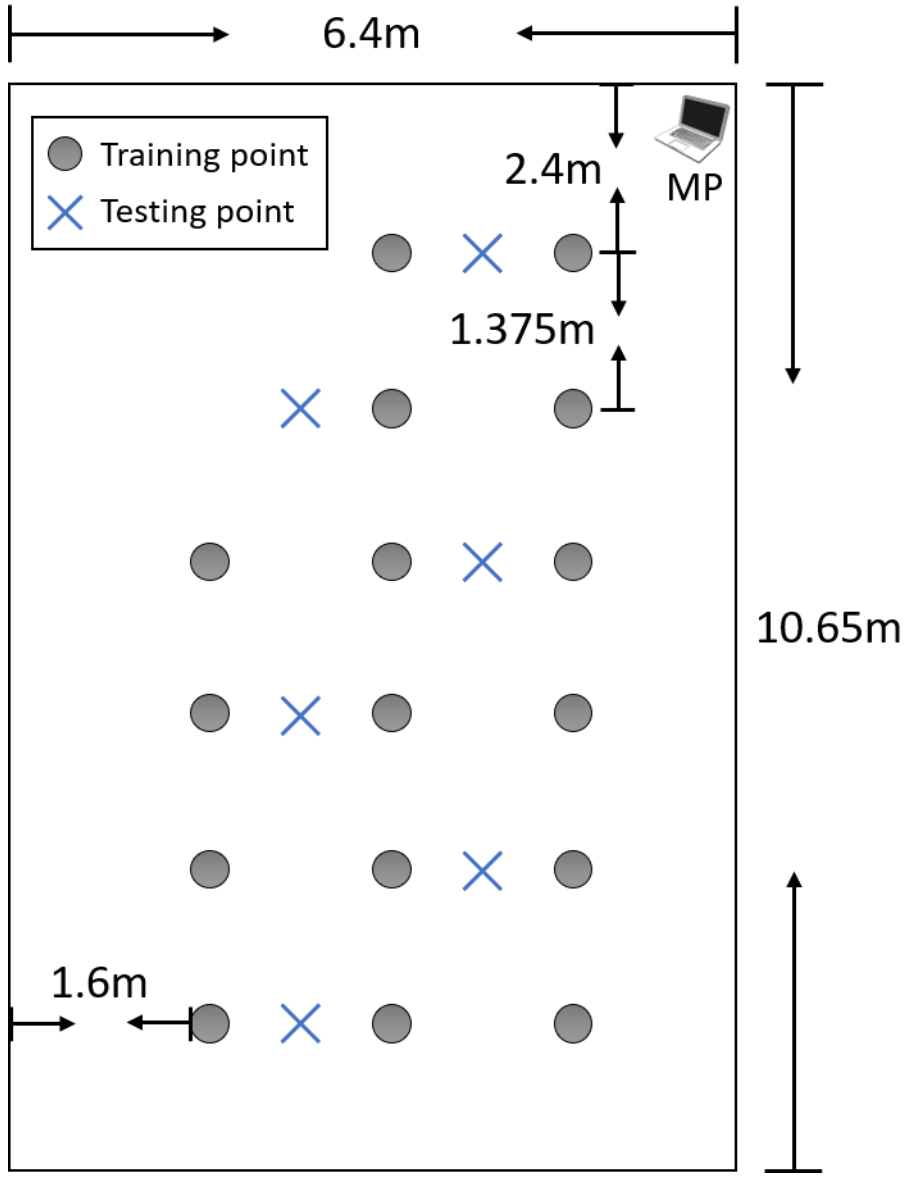
Data Collection

Use a laptop with an Intel WiFi Link 5300 (iwl5300) 802.11n network card as the receiver, and a MAP-AC1750 router as the transmitter, as shown in the figure below.

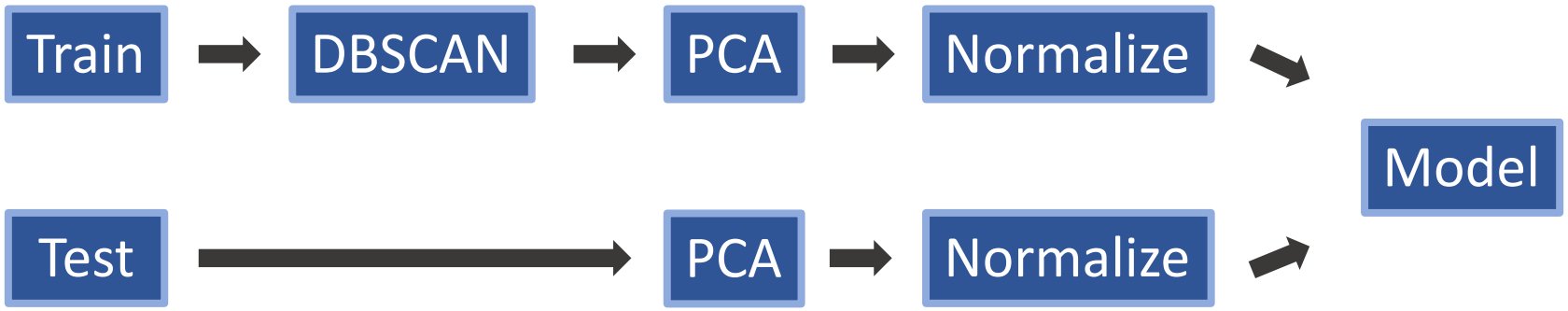


Data Collection

- Venue: Building 4, Room 710
- The sampling data of 16 reference points in the classroom are used as training points, and 6 nearby points are used as testing points, as shown in the figure on the right.



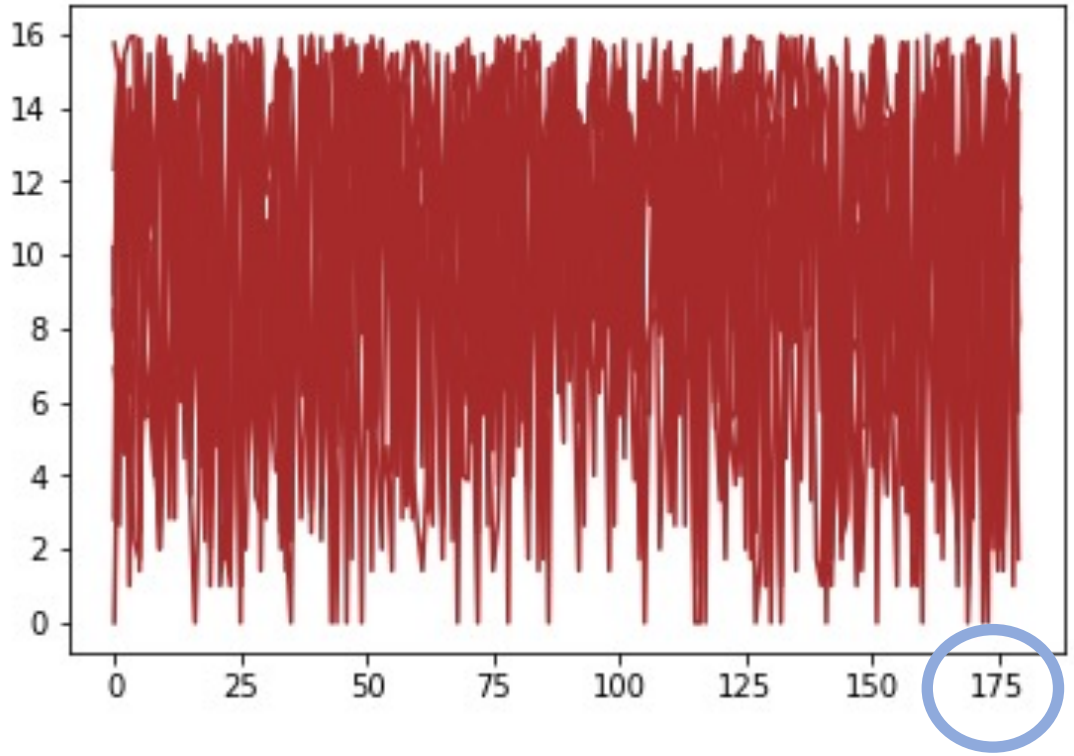
Experiment framework



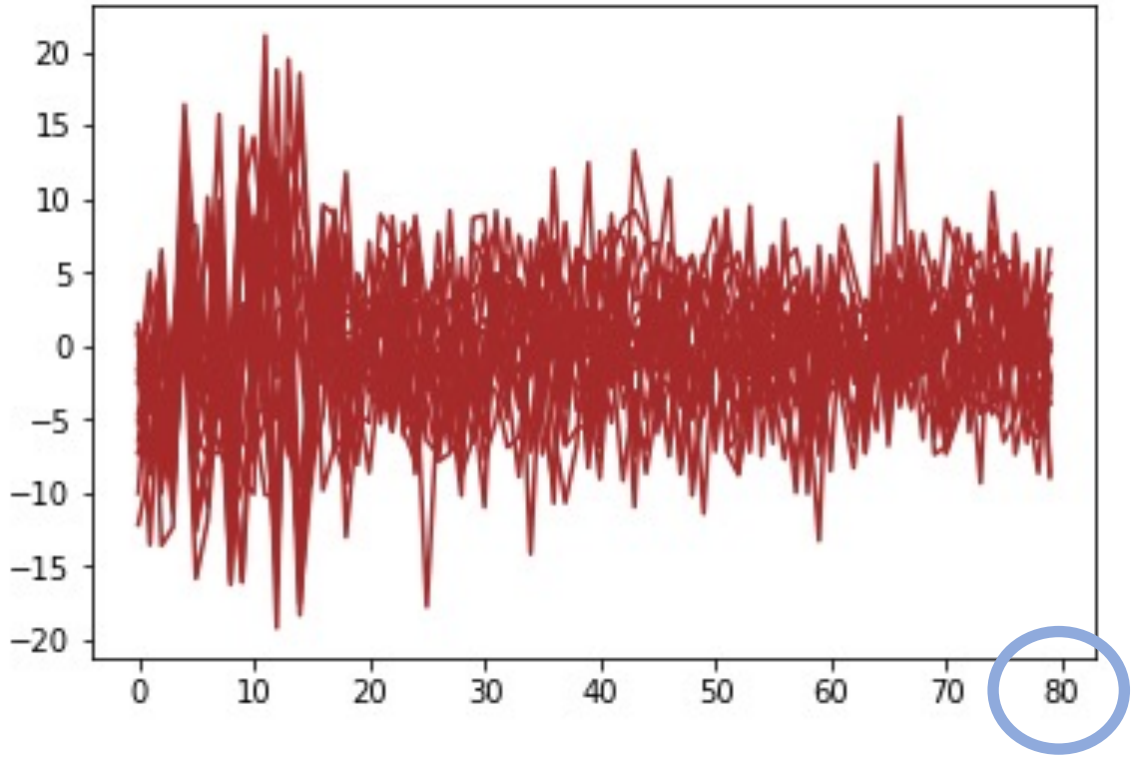
PCA

The original data is disassembled into more representative principal components through PCA, and the dimensionality is reduced at the same time.

before

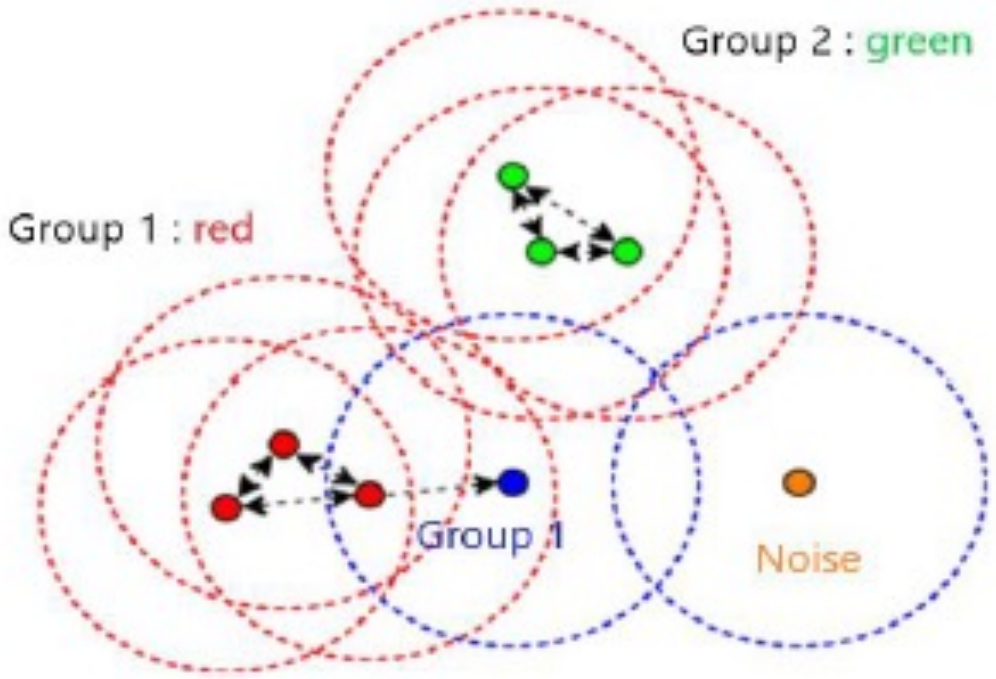


after

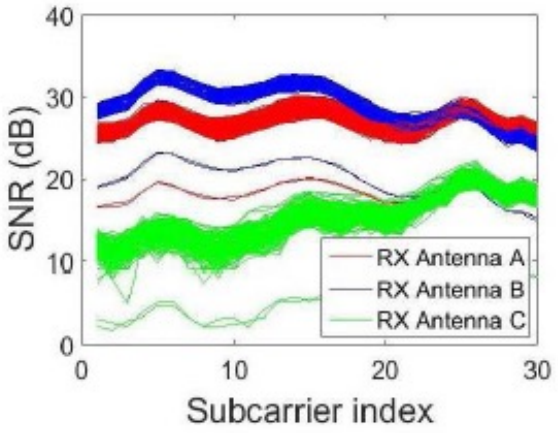


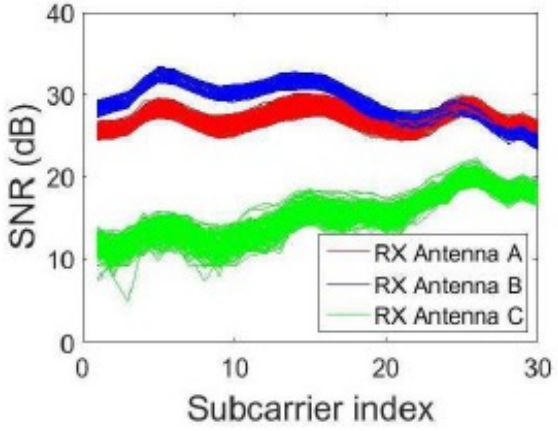
DBSCAN

Divide data into multiple groups through DBSCAN, and mark points with large deviations.

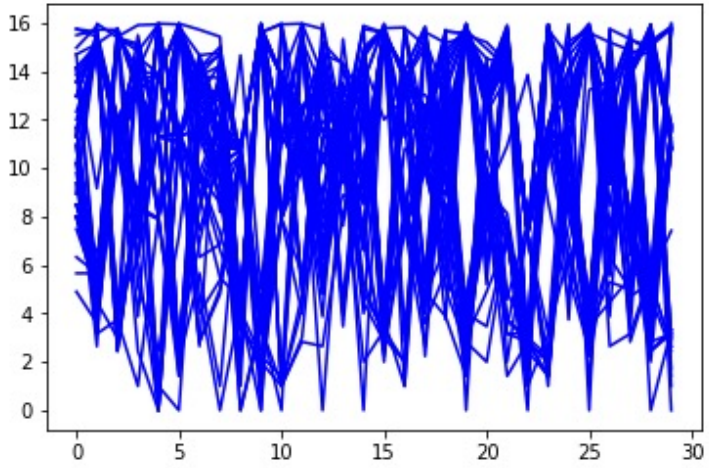


DBSCAN

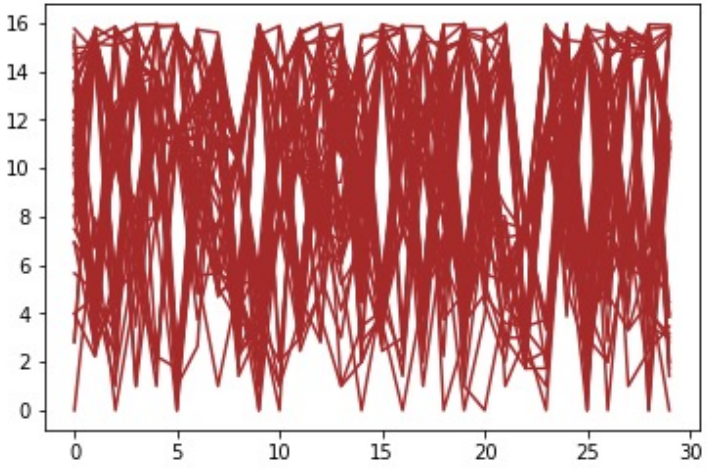




Without denoise



With denoise



Model

Linear regression

SVM - SVC

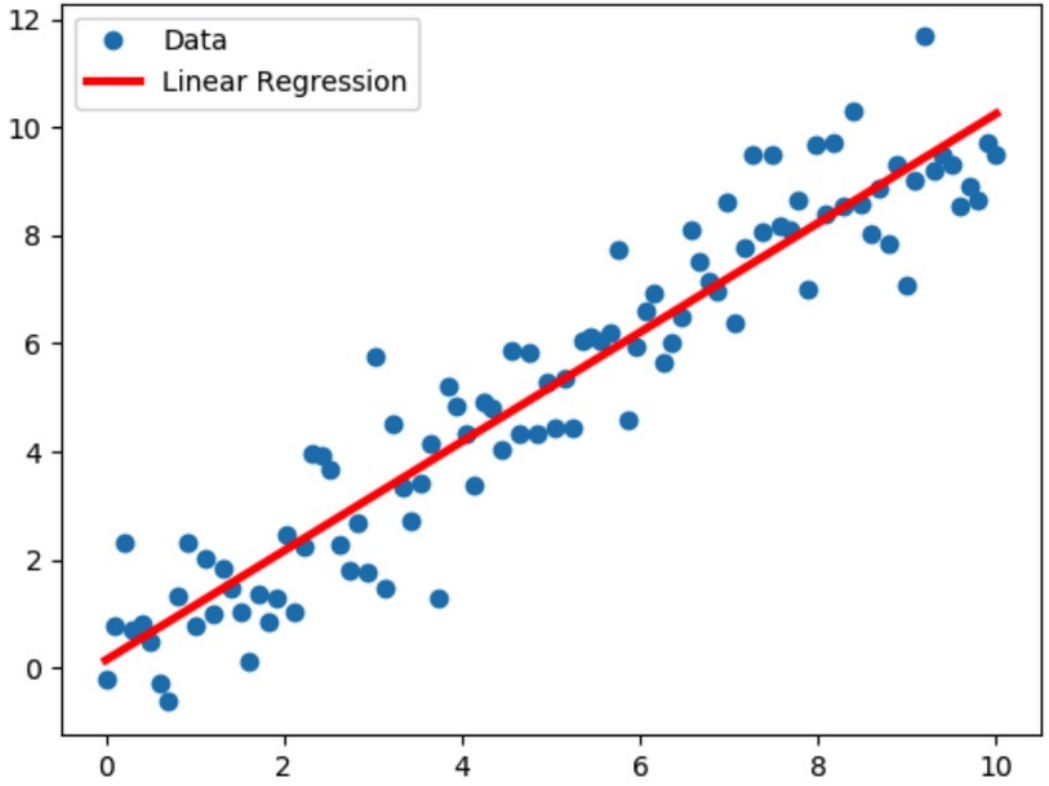
SVM - SVR

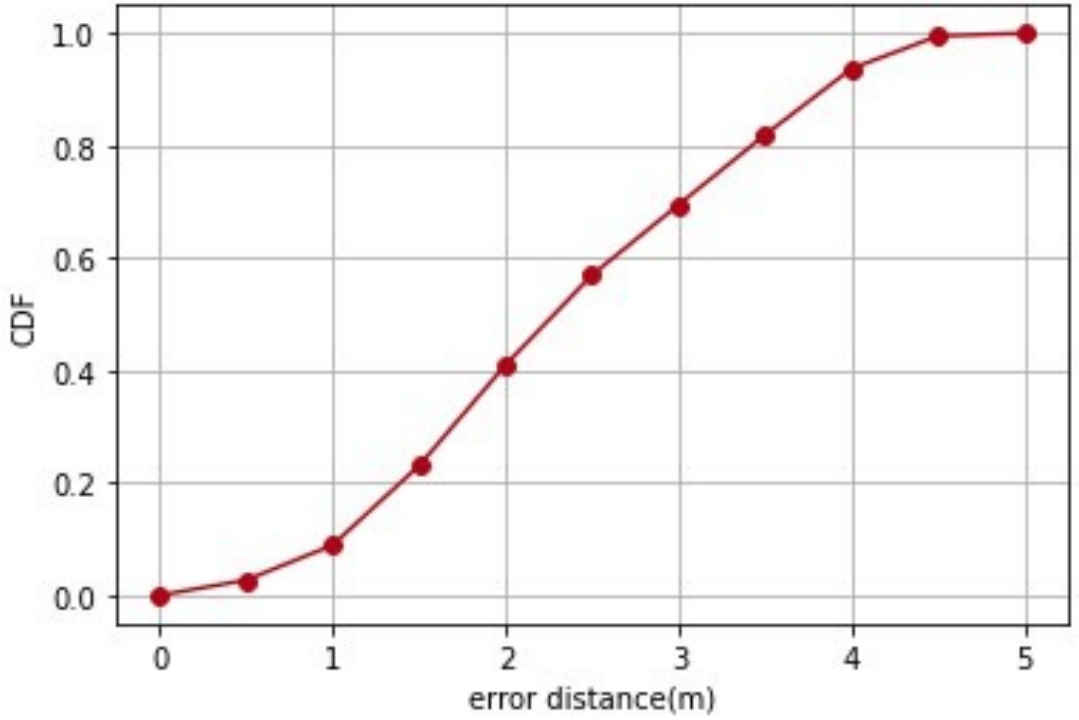
KNN

NN

Linear Regression

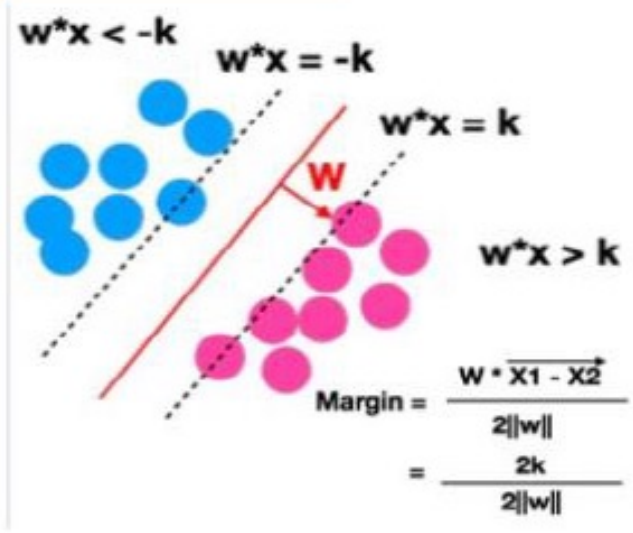
$Y = aX + b$





SVM-SVC

Find a hyperplane (Hyperplane), which can effectively classify or regression analysis the data to mark, while ensuring that the samples on both sides of the hyperplane are as far away from the hyperplane as possible.




```
clf=SVC(kernel='rbf',probability=True)
```

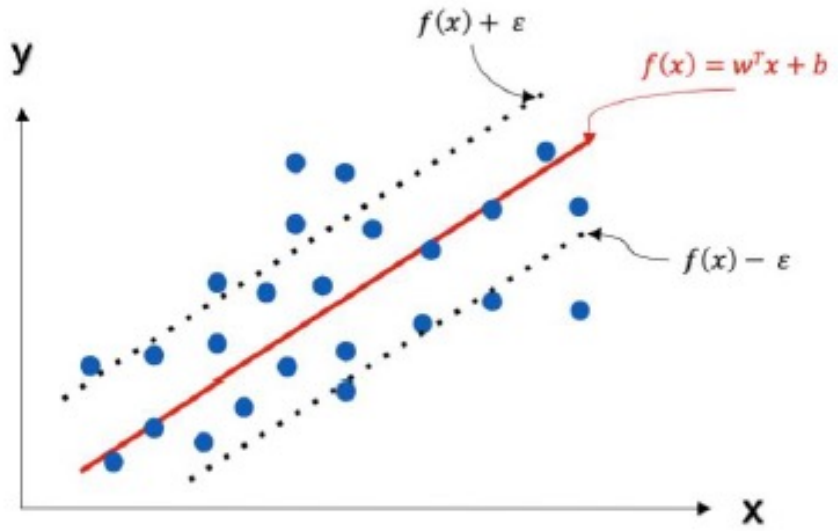
$$\text{minimize } \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{n}\xi_i$$

$$\text{subject to } \begin{cases} c_i(w^T r_i + b) \geq 1 - \xi_i \\ C > 0, \quad \xi_i \geq 0 \end{cases}$$

$$f(r) = sign(\sum_{i=1}^{n} c_i \alpha_i K(r_i, r) + b)$$

$$K(x_i, x_j) = exp(-\gamma \|x_i - x_j\|^2)$$

SVM-SVR




```
clf_svr=SVR(kernel='rbf')
```

$$\text{minimize } \frac{1}{2}\|w\|^2 + C(v\epsilon + \frac{1}{n}\sum_{i=1}^{n}(\xi_i^* + \xi_i))$$

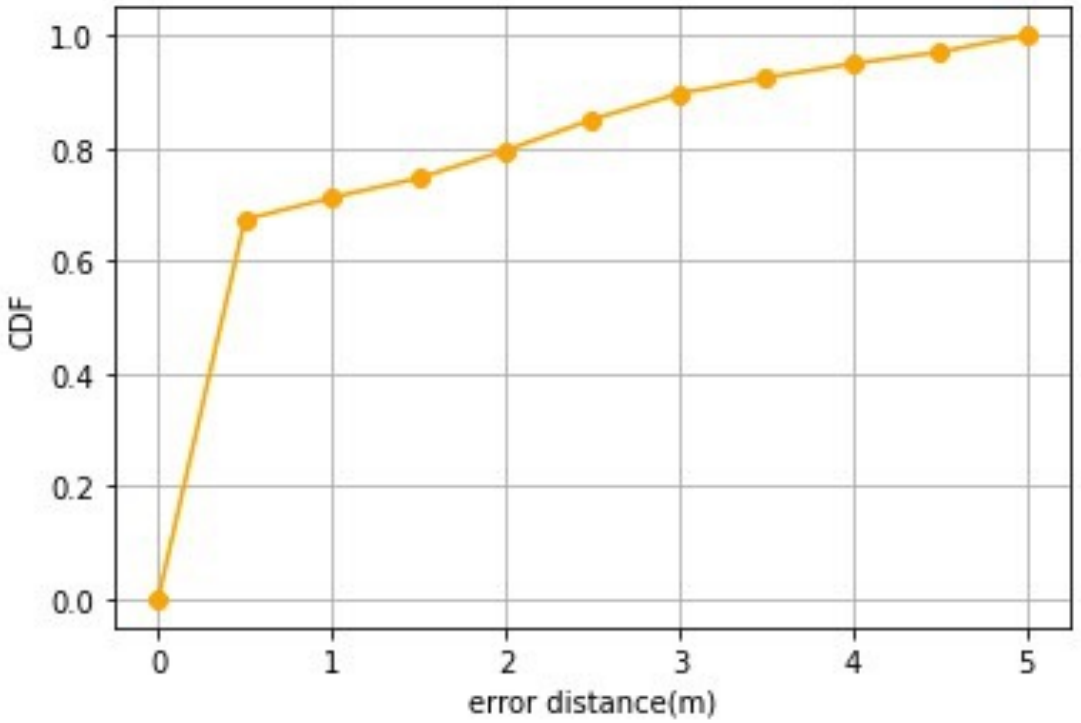
$$\text{subject to } \begin{cases} (w^T r_i + b) - c_i \le \epsilon + \xi_i \\ c_i - (w^T r_i + b) \le \epsilon + \xi_i^* \\ \xi_i, \xi_i^*, \quad \epsilon \ge 0 \end{cases}$$

$$f(r) = \sum_{i=1}^{n}(\alpha_i - \alpha_i^*)K(r_i, r) + b$$

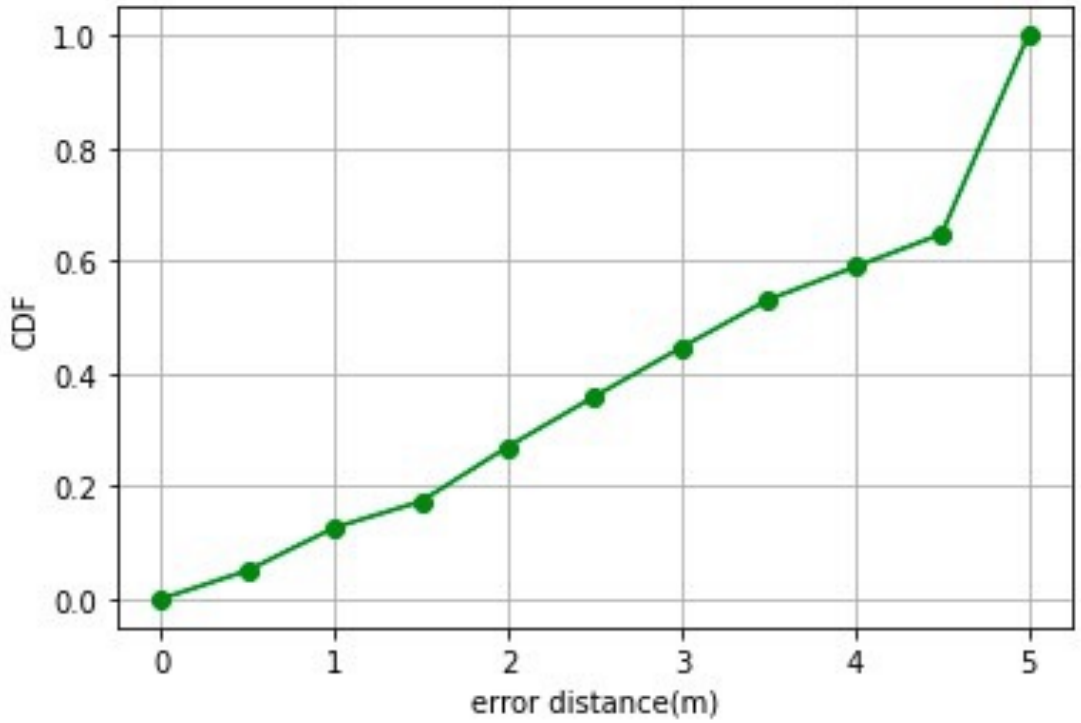
$$K(x_i, x_j) = exp(-\gamma \|x_i - x_j\|^2)$$

SVM

SVC

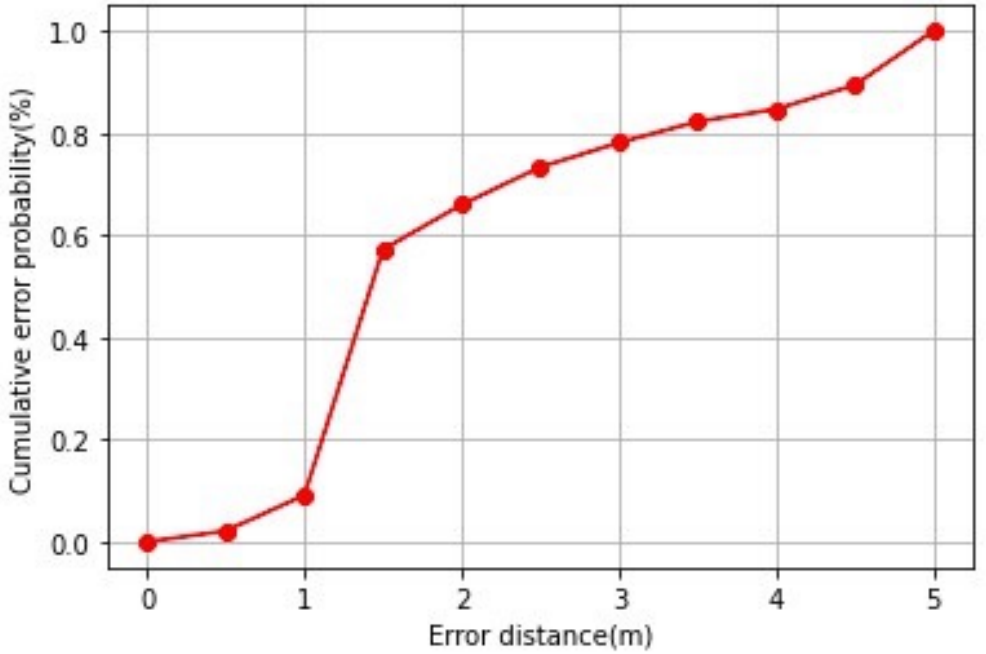

1.0
0.8
0.6
0.4
0.2
0.0
CDF
0
1
2
3
4
5
error distance(m)

SVR

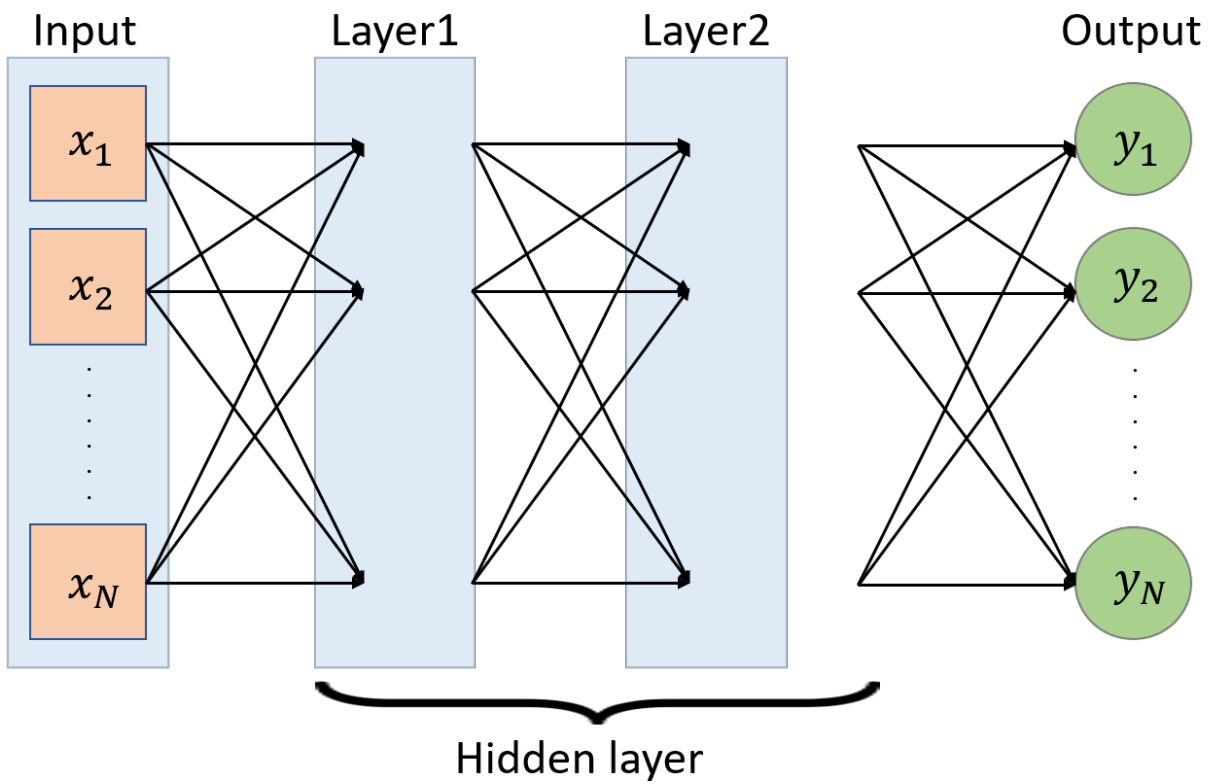

1.0
0.8
0.6
0.4
0.2
0.0
CDF
0
1
2
3
4
5
error distance(m)

KNN

Find the nearest k points from the newly added data, and add the group with the largest number of them.

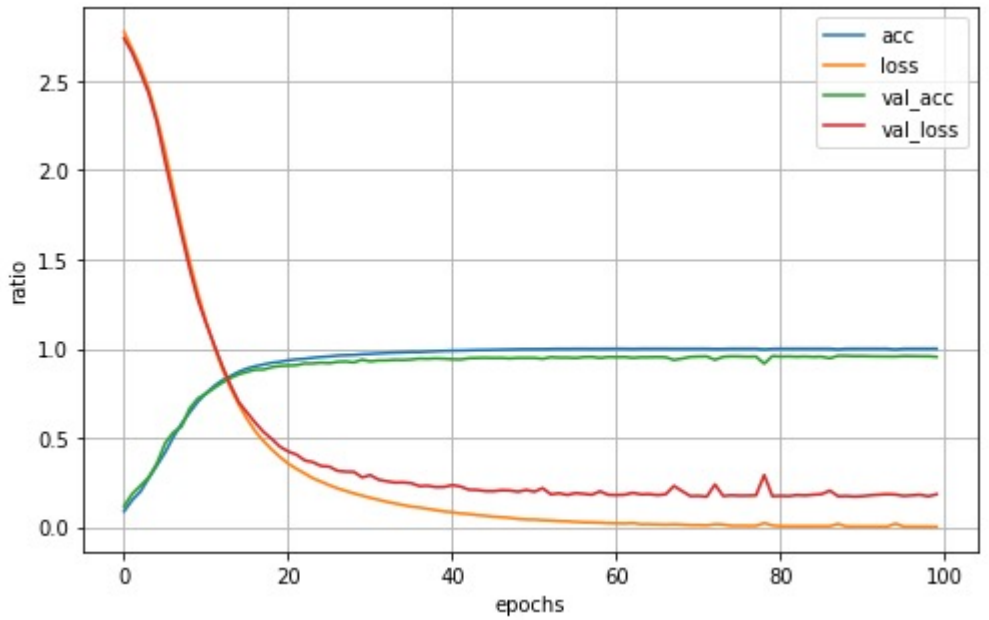


NN



Result - model

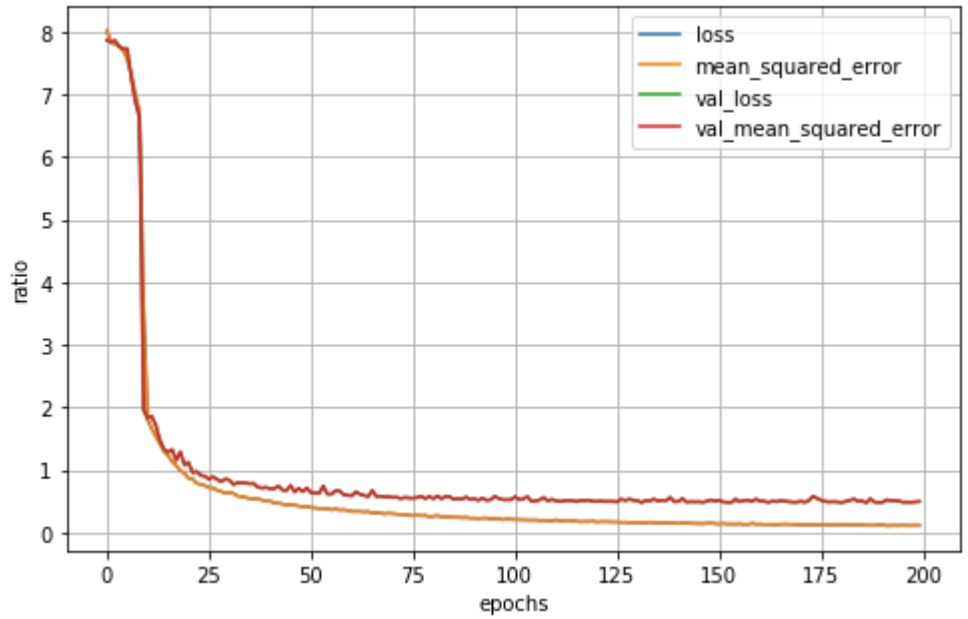
- The input of Categorical crossentropy is the CSI of the training point and the corresponding label passes through the one-hot encoder.
- The output of Categorical crossentropy is the probability of testing point on each training point



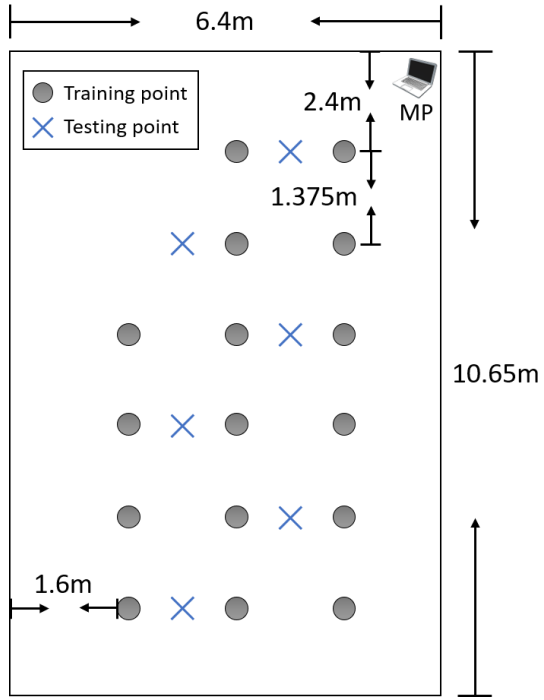
▲categorical crossentropy

Result - model

- The input of Mean square error is the CSI of the training point and the corresponding coordinate points.
- The output of the Mean square error is the predicted coordinates of the testing point

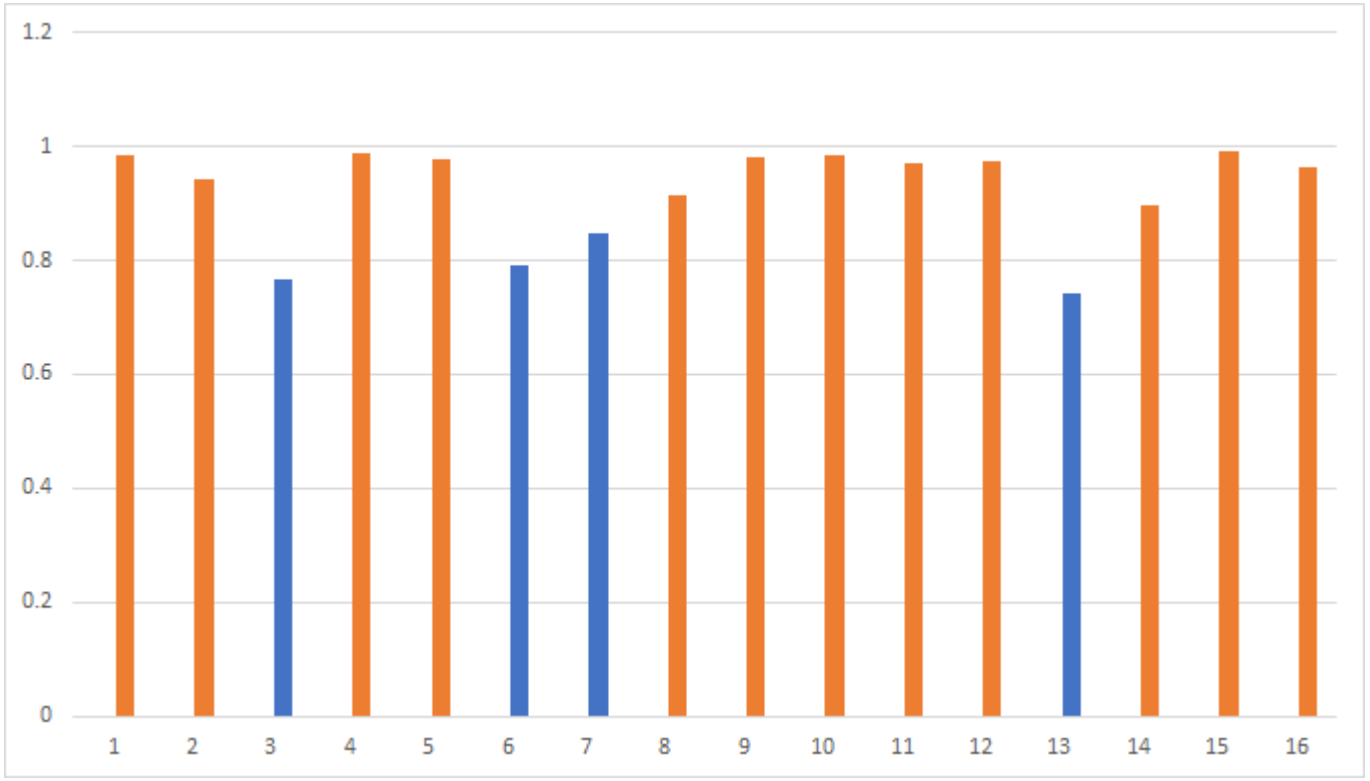


▲mean square error

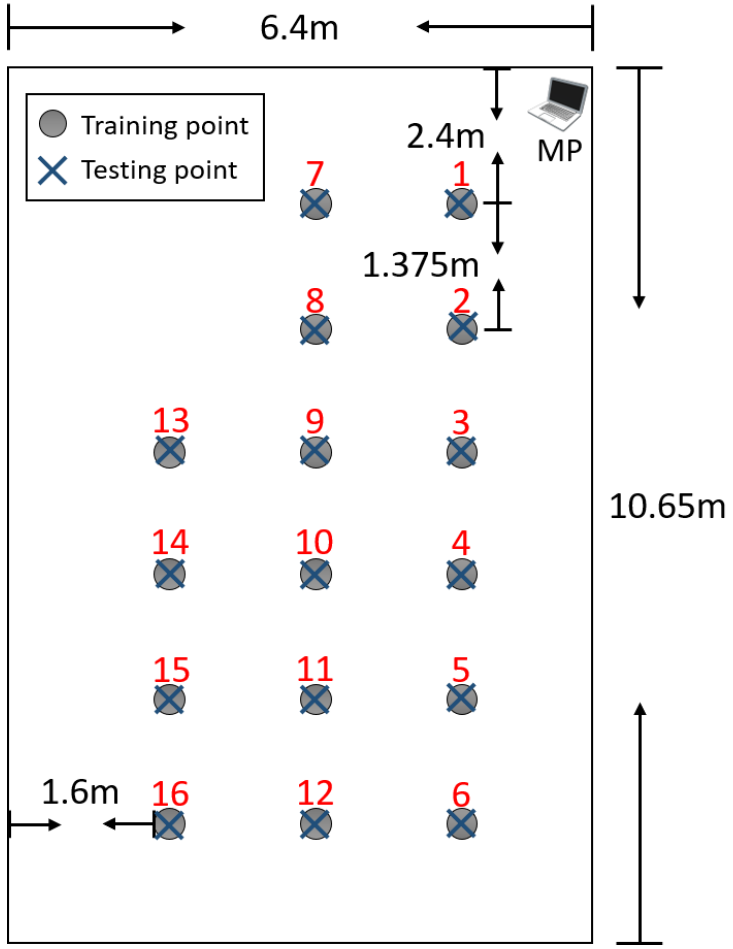


▲coordinate diagram

Conclusion - categorization



▲accuracy of testing point

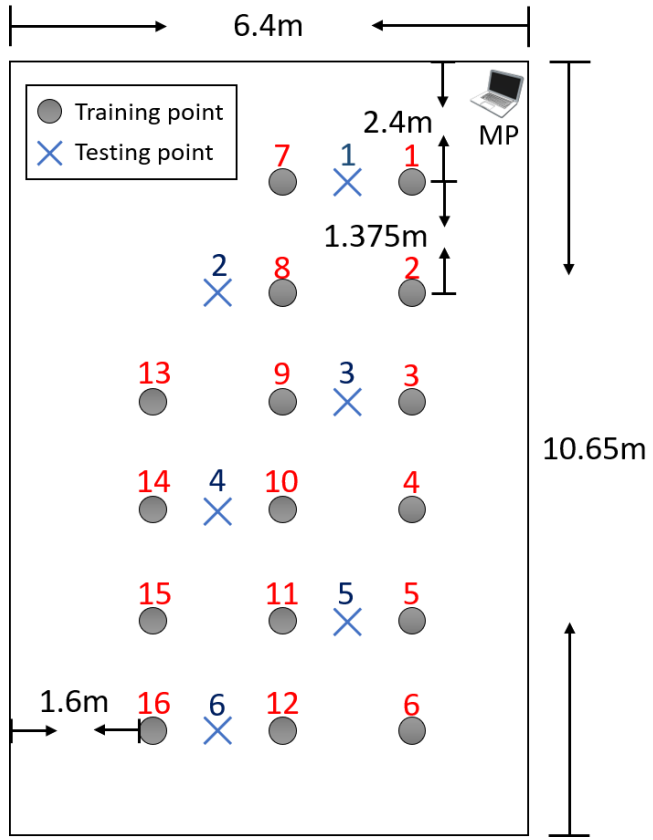


▲testing point

Conclusion - prediction

When we look for factors affecting the accuracy rate, we find that the error of the first point was too large, so we calculated the average error of each point.

And the error of the first and second point is the largest.

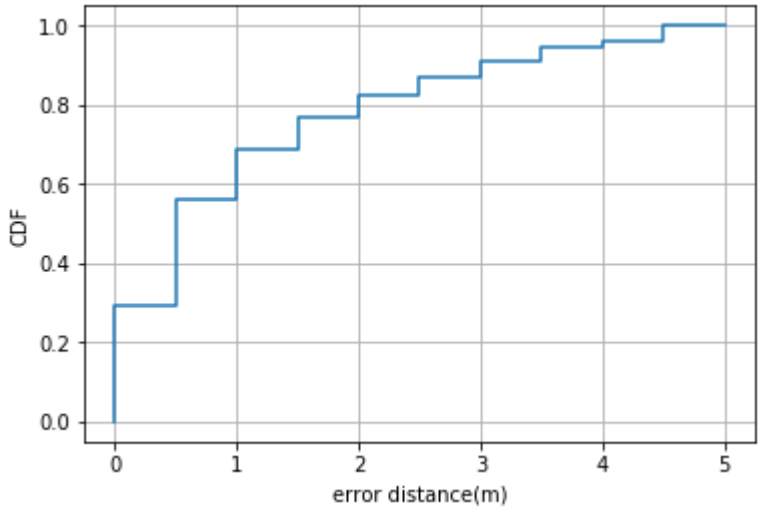


▲testing point

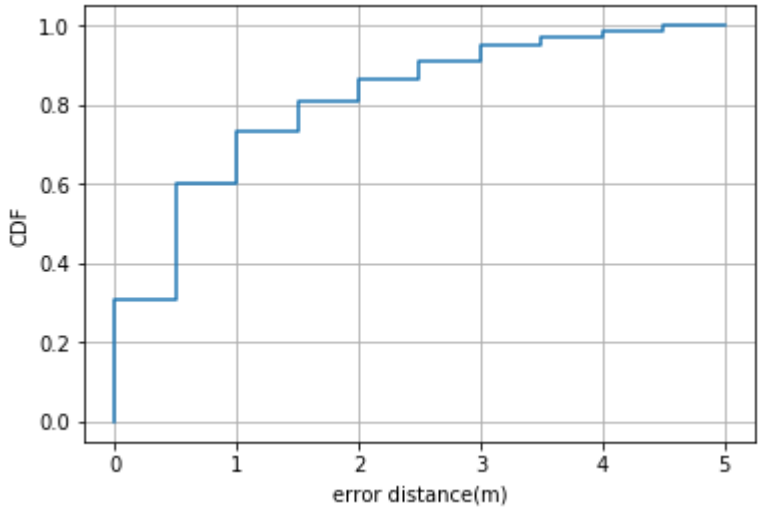
	1	2	3	4	5	6
average error	4.030719	3.802222	1.384805	2.787186	2.389254	2.248498

▲average error of testing point

Conclusion - CDF

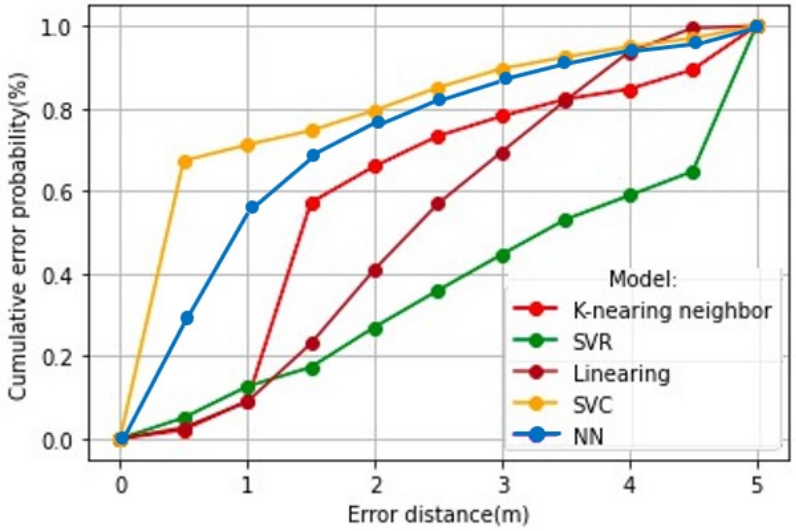


▲with testing point 1



▲without testing point 1

Conclusion - final



1. The accuracy of training points are higher than testing points.
2. When the error distance is within 1m, the effect of SVC will be the best, and NN will be second.