The chosen algorithm is the K-nearest neighbour classification. This algorithm is easy to understand, but I think it has many in-depth problems that deserve me to study and explore.

(1) Feature Normalization

(2) If the difference in the data volume of the different classifications is pronounced, I need to add the distance weights.

(3) About the choice of the optimal hyperparameters, I would apply the grid search method. Hyperparameters should include K, the way to measure distance.

(4) About the performance of the K-nearest neighbour, I would explore the k-dimension tree to optimize computing speed.

**Questions**: In K-nearest neighbour, there is no model parameter. So, is the following statements reasonable? Crossing validation is unmeaningful. We can replace crossing validation with grid search.

About the dataset choice, what range do you think the dimension of the dataset should be selected for being suitable for the k-nearest neighbour? What about data volume?

Do you have any suggestions about the above problems I want to study?