**Introduction**

1. The problem is dimension reduction and the comparison of different classification models for identifying letters in various fonts. It can be done in 3 steps: data preprocessing, model fitting, and dimension reduction.

In order to complete the task efficiently, the optimal classifier needs to be selected. And there are many factors that can be considered to choosing the best classifier. The factors are times costs, accuracy, memory usage, ease of implementation, degree of fit, and some other factors.

1. Binary classification is the task of classifying the elements of a set into two groups on the basis of a classification rule. For the third problem, I choose ‘A’ and ‘B’. I think pairs that are more similar are harder to classify like they have the same width and the same height. So ‘H’ and ‘K’ are the hardest to classify, and ‘M’ and ‘Y’ are the easiest to classify.
2. The dimension reduction is good for reducing time costs, but it will cause the accuracy to decrease. The more dimension reduced, the lower accuracy I got.

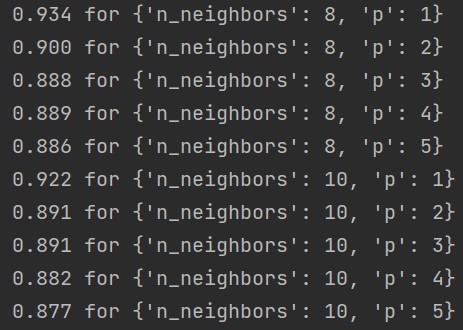
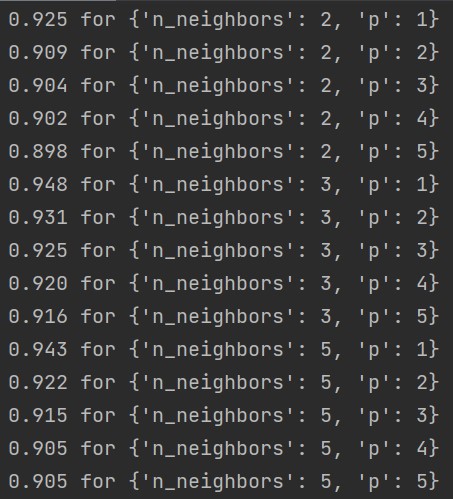
**Results**

1. KNN
2. The kNN algorithm is when predicting a new value x, judging the category of x according to the category of the K points closest to it.

Advantage: Easy to implement.

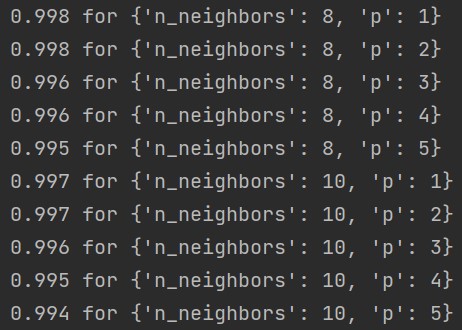
Disadvantage: When the data dimension is high, the amount of computation is large; samples with close distances may not belong to the same category.

1. Pair 1: H and K



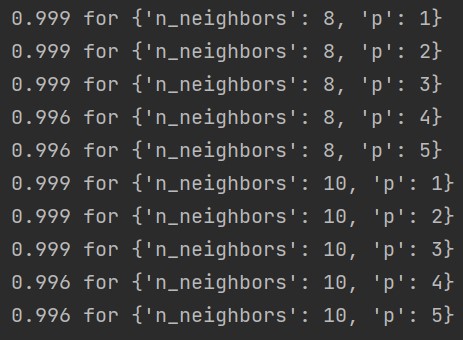
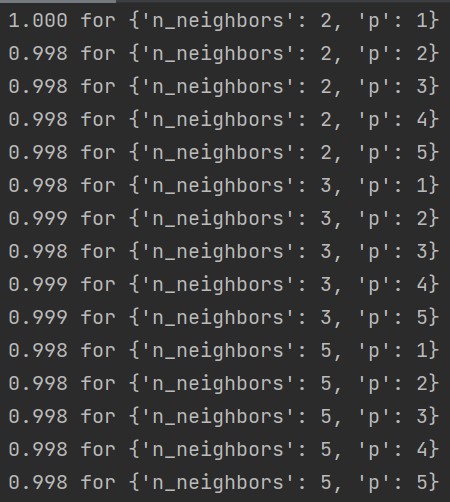
Best result in cross-validation: 0.9479245283018868.

Pair 2: M and Y



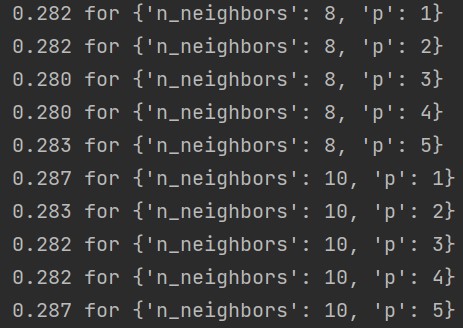
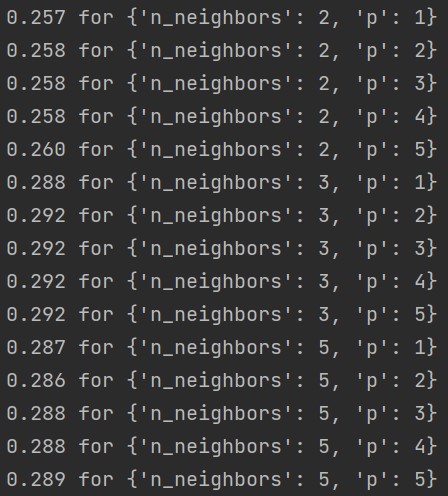
Best result in cross-validation: 0.9985915492957746.

Pair 3: A and B



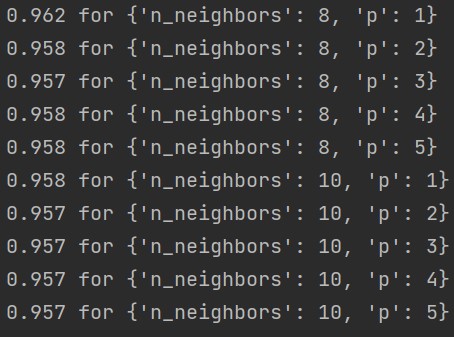
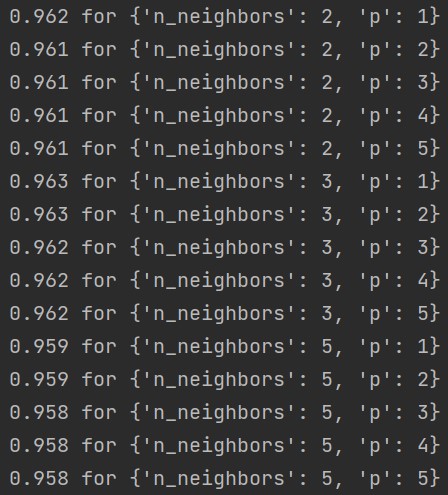
Best result in cross-validation: 1.0.

1. The method is Low Variance of Simple Quality Filtering. And the threshold is 4.85 to get the final 4 dimensions for the pair1, threshold is 8 for pair2, threshold is 4.815 for pair3. If a feature’s value changes very little across the samples, it’s probably not informative, and the method will put these features away.
2. Pair 1: H and K



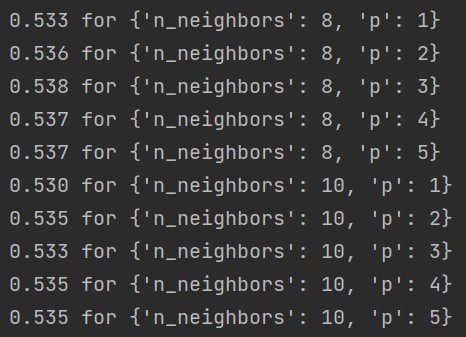
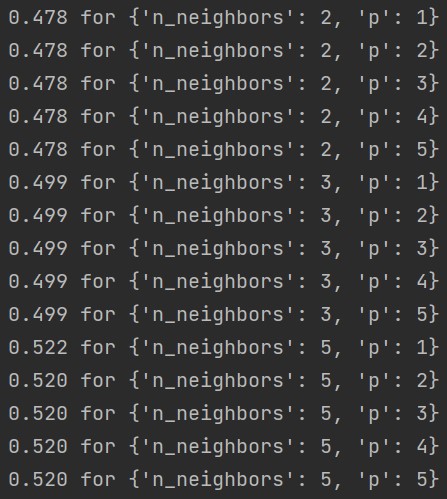
Best result in cross-validation: 0.2920754716981132.

Pair 2: M and Y



Best result in cross-validation: 0.9626760563380282.

Pair 3: A and B



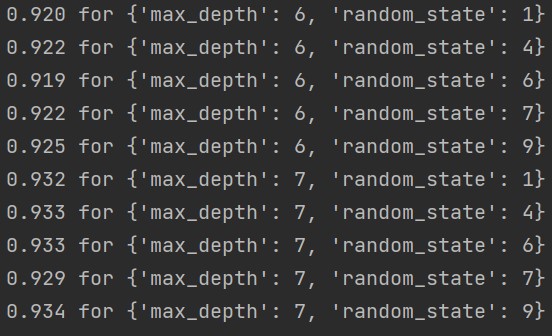
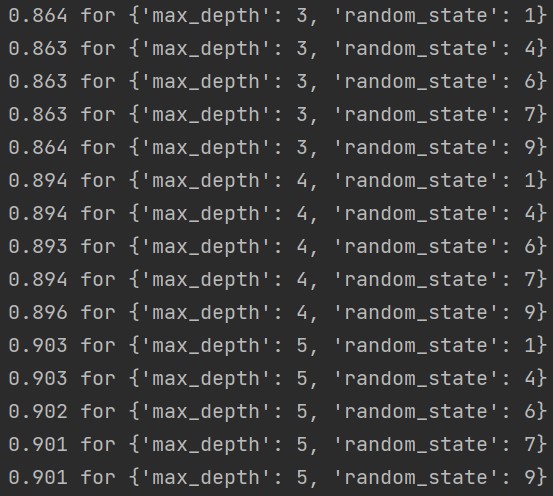
Best result in cross-validation: 0.5375166410650282.

1. Hyperparameters are ‘n\_neighbors’ and ‘p’.
2. Decision Tree
3. A Decision Tree is a tree-like structure in which each internal node represents a test output on each attribute in a test, and each leaf node represents a category.

Advantage: less time complexity, easy to understand and explain.

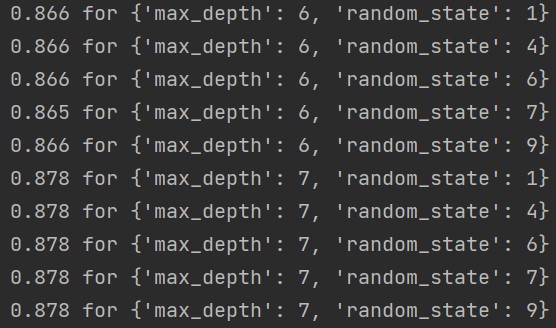
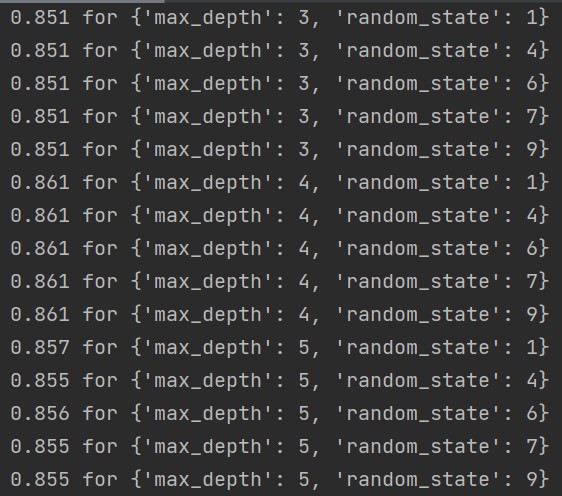
Disadvantage: easy to overfit, ignore dependencies between attributes.

1. Pair 1: H and K



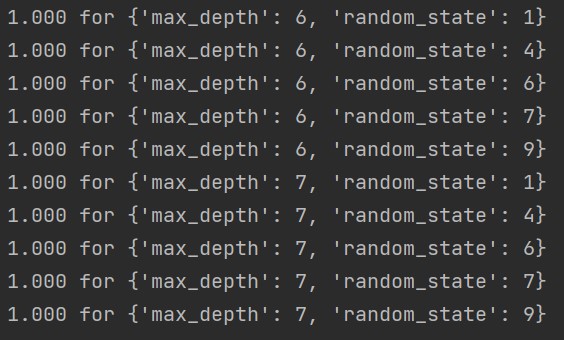
Best result in cross-validation: 0.9343396226415095.

Pair 2: M and Y



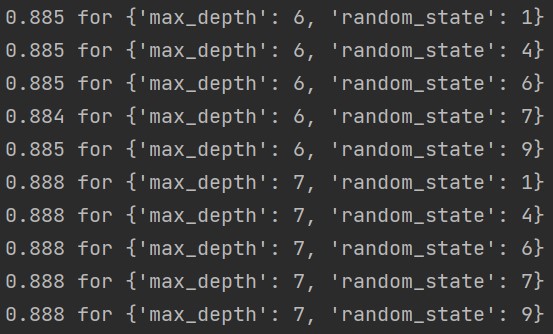
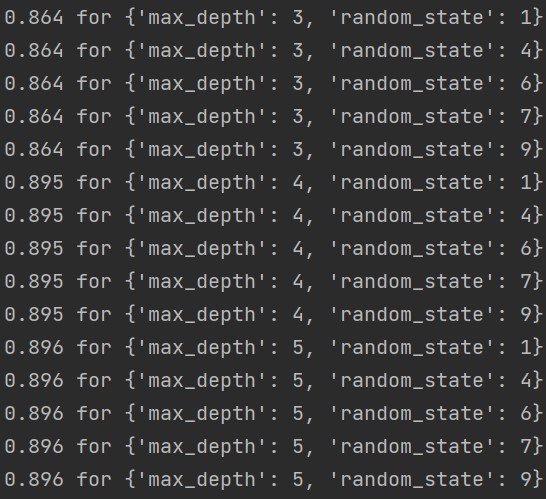
Best result in cross-validation: 0.8784905660377358.

Pair: A and B



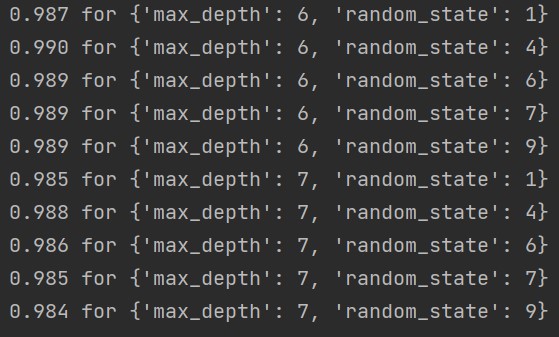
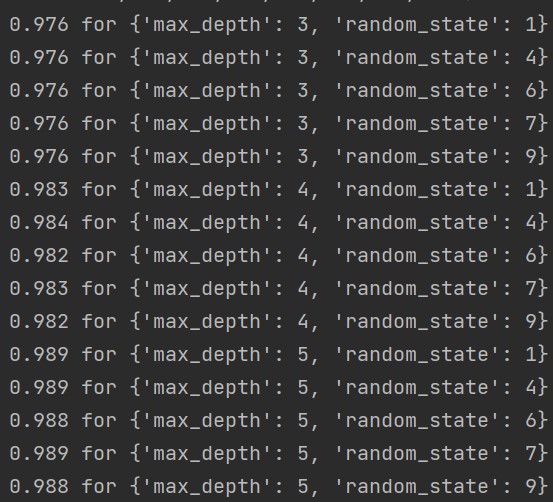
Best result in cross-validation: 1.0.

1. The dimension reduction method is Univariate Feature Selection of Feature Selection. It can get the top k features with the highest score.
2. Pair 1: H and K



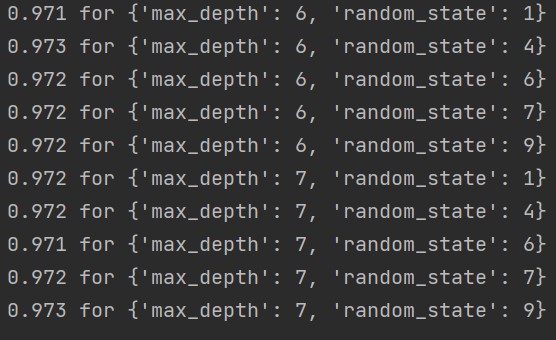
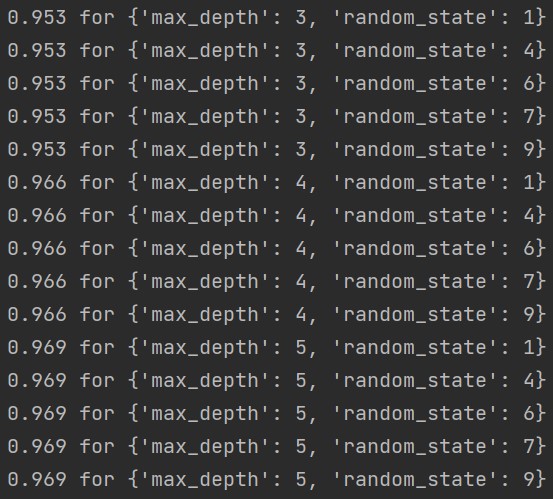
Best result in cross-validation: 0.8958490566037736.

Pair 2: M and Y



Best result in cross-validation: 0.9901408450704224.

Pair 3: A and B



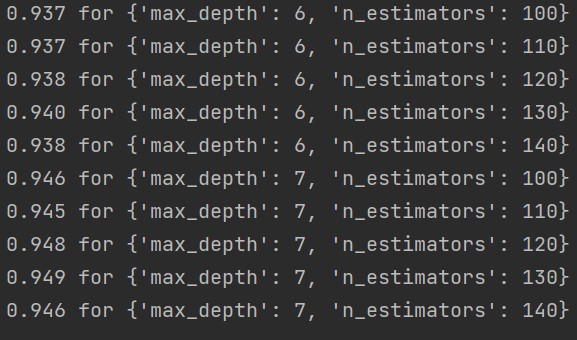
Best result in cross-validation: 0.9728315412186381.

1. Hyperparameters are ‘max\_depth’ and ‘random\_state’.
2. Random Forest
3. Random Forest is an algorithm that integrates multiple trees through the idea of ​​ensemble learning. Its basic unit is a decision tree.

Advantage: good performance, strong anti-interference ability.

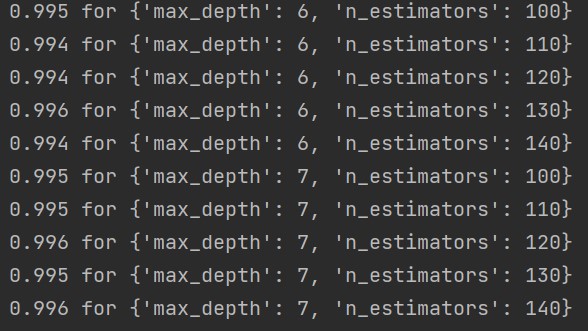
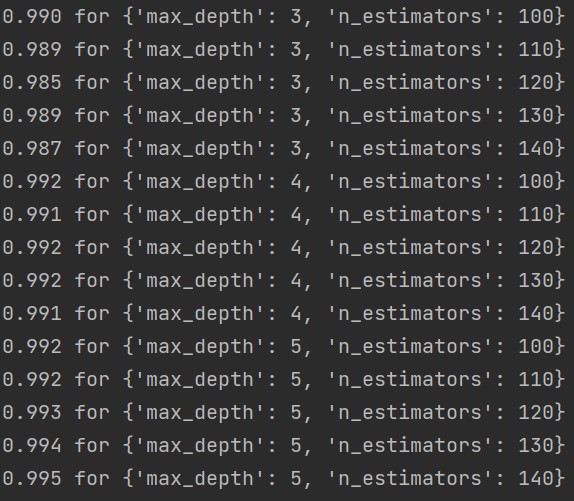
Disadvantage: many similar decision trees masking the real results, easy to overfit.

1. Pair 1: H and K



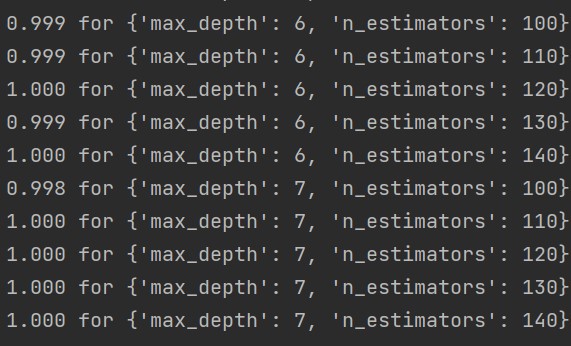
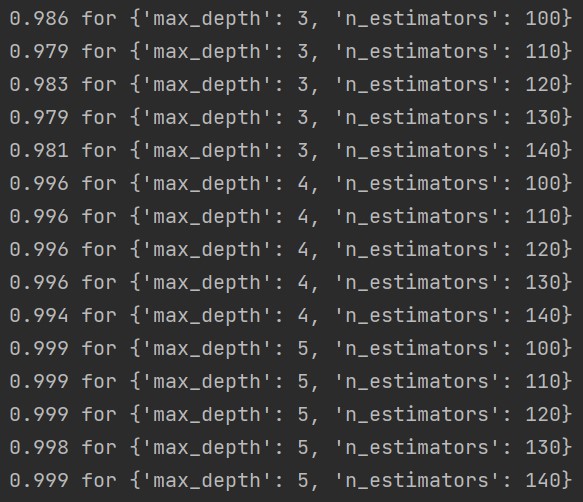
Best result in cross-validation: 0.9486792452830188.

Pair 2: M and Y



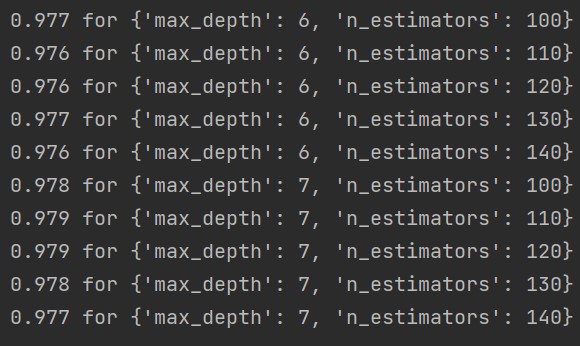
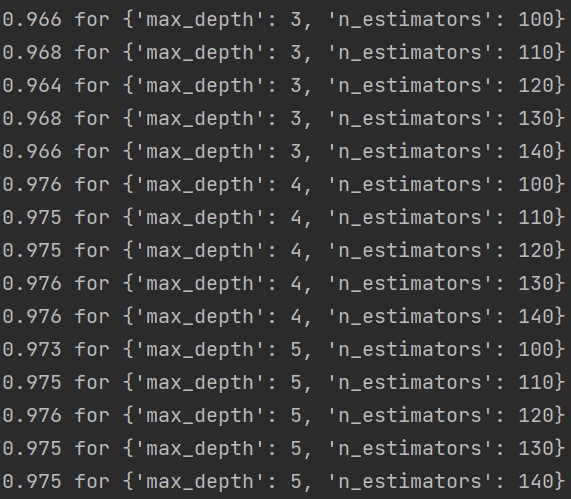
Best result in cross-validation: 0.995774647887324.

Pair 3: A and B



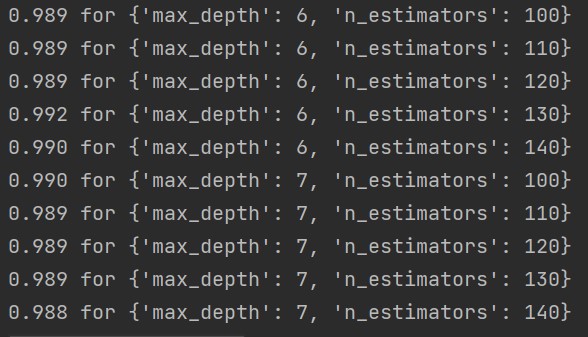
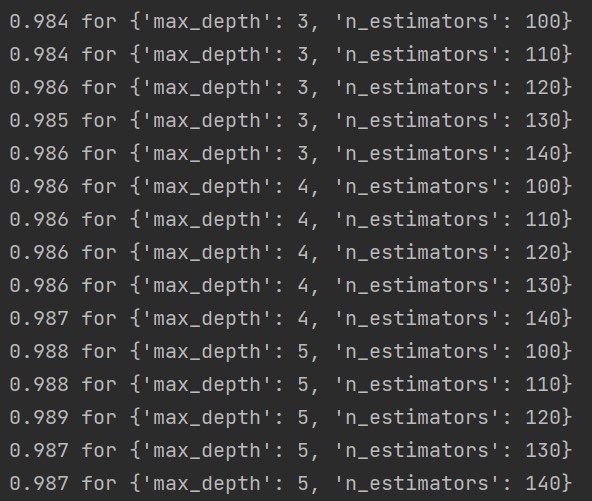
Best result in cross-validation: 1.0.

1. The dimension reduction method is Univariate Feature Selection of Feature Selection. It can get the top k features with the highest score.
2. Pair 1: H and K



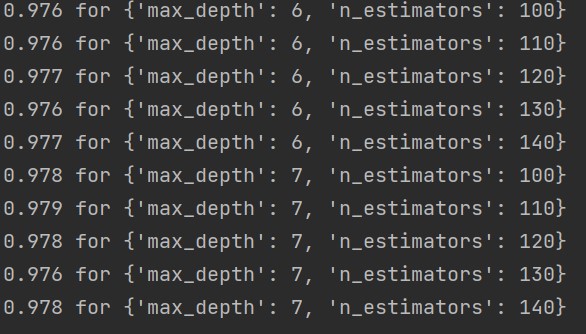
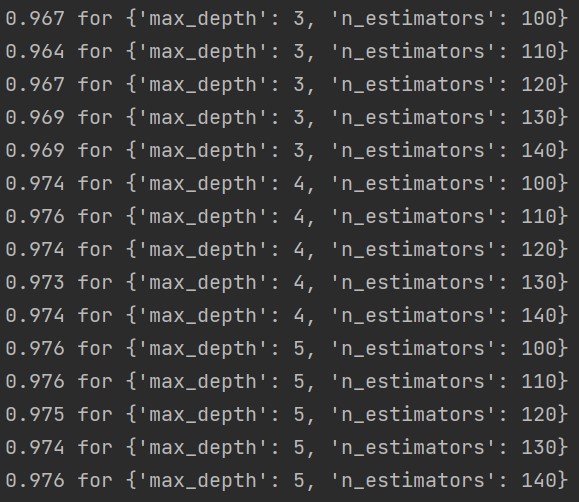
Best result in cross-validation: 0.979267793138761.

Pair 2: M and Y



Best result in cross-validation: 0.991549295774648.

Pair 3: A and B



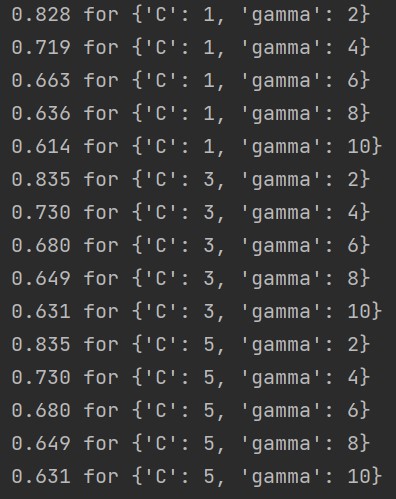
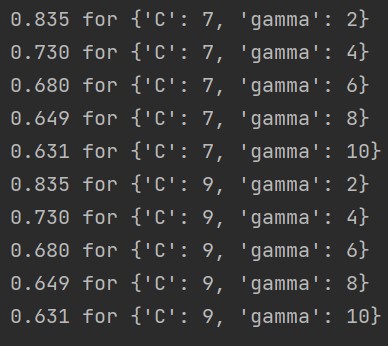
Best result in cross-validation: 0.9785535074244752.

1. Hyperparameters are ‘max\_depth’ and ‘n\_estimators’.
2. SVM
3. SVM is a two-class classifier, and its goal is to find a hyperplane, using two classes of data as far away from the hyperplane as possible, so that it can classify new data more accurately, even if the classifier is more robust.

Advantage: performs well on various datasets.

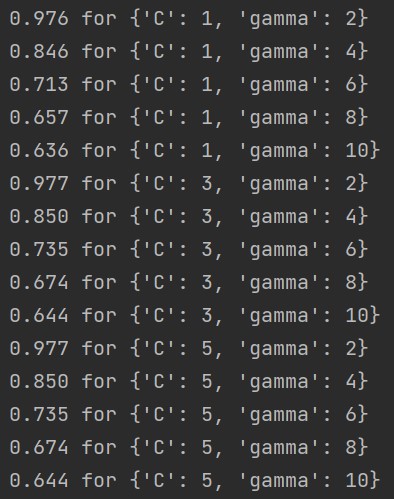
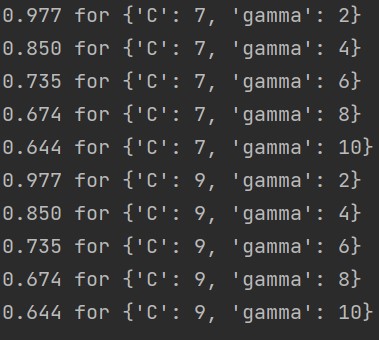
Disadvantage: not suitable for large sample data.

1. Pair 1: H and K



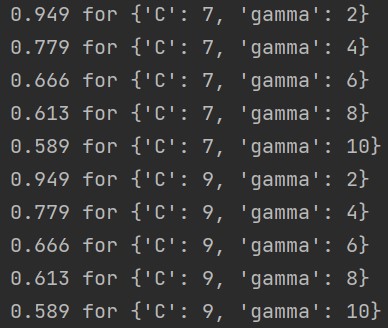
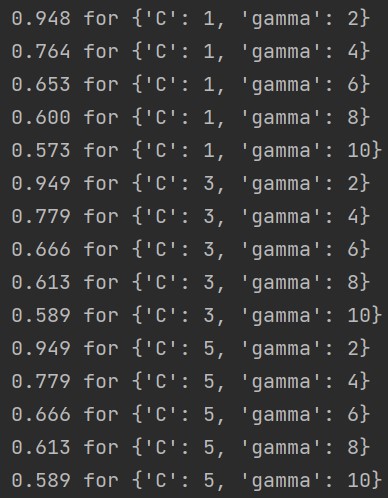
Best result in cross-validation: 0.8347169811320754.

Pair 2: M and Y



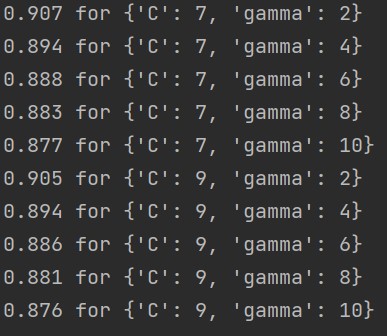
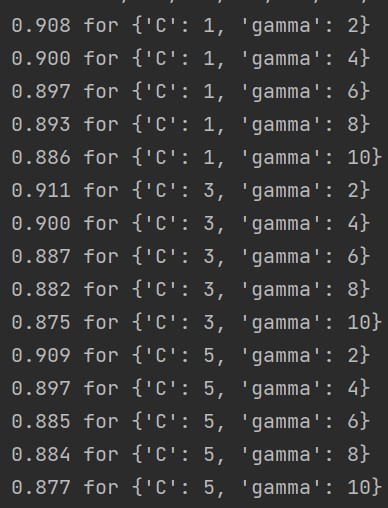
Best result in cross-validation: 0.9774647887323944.

Pair 3: A and B



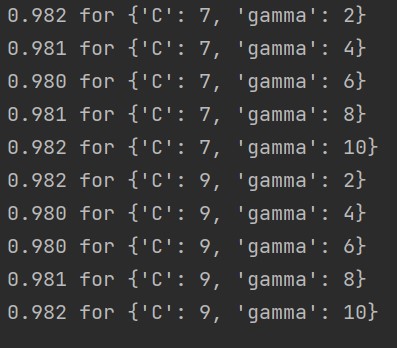
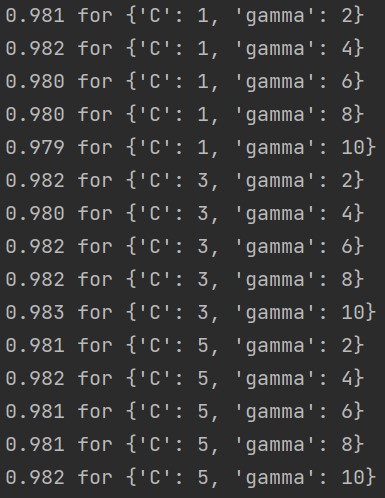
Best result in cross-validation: 0.9485330261136713.

1. The method is PCA. In PCA, data is transformed from the original coordinate system to the new coordinate system. The choice of a new coordinate system is determined by the data itself. The first new axis selects the direction with the largest variance in the original data, and the second new axis selects the direction that is orthogonal to the first axis and has the greatest variance. This process is repeated for the number of features in the original data. You will find that most of the variance is contained in the first few new axes. Therefore, we can only select the first few coordinate axes, that is, the data is dimensionally reduced. Advantage: reduce the computational cost of the algorithm, avoid noisy. Disadvantage: feature decomposition has limitations.
2. Pair 1: H and K



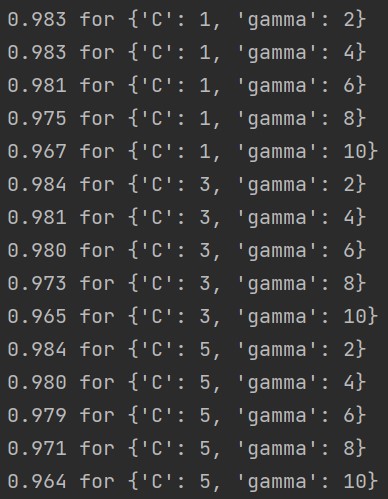
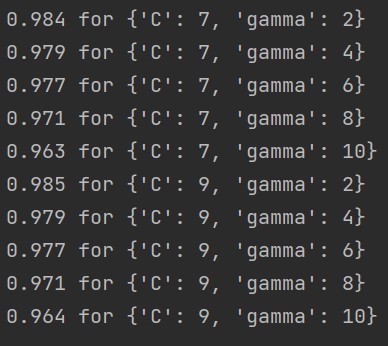
Best result in cross-validation: 0.9109433962264152.

Pair 2: M and Y



Best result in cross-validation: 0.9830985915492958.

Pair 3: A and B



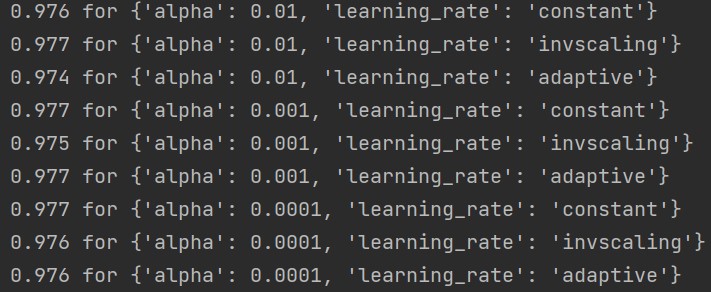
Best result in cross-validation: 0.9849897593445981.

1. Hyperparameters are ‘C’ and ‘gamma’.
2. Artificial Neural Network
3. A single neuron collects the information transmitted by other neurons on the dendrite. After the amount of information reaches a threshold, a pulse signal is generated, which is transmitted through the axon to the synapse, and then transmitted to the next neuron.

Advantage: ability to find optimal solutions at high speed, high classification accuracy.

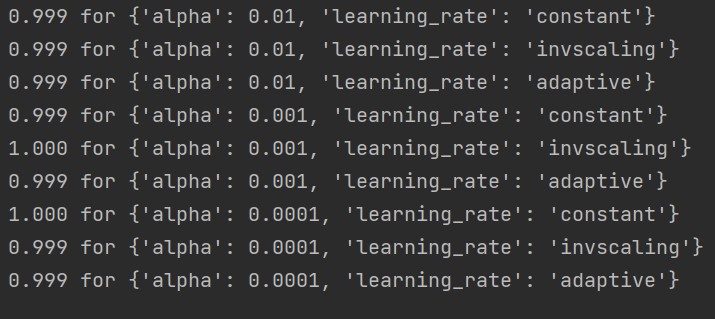
Disadvantage: require a large number of parameters, it may fall into a local minimum.

1. Pair 1: H and K



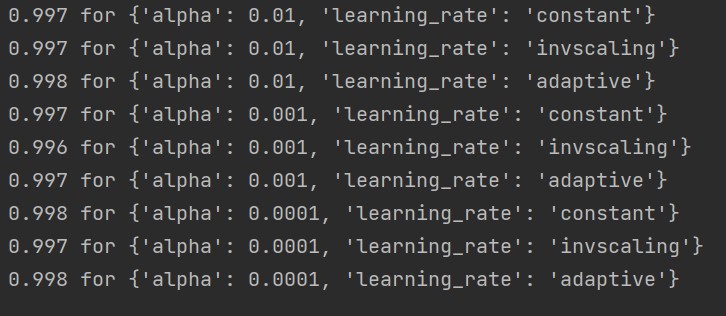
Best result in cross-validation: 0.9773584905660379.

Pair 2: M and Y



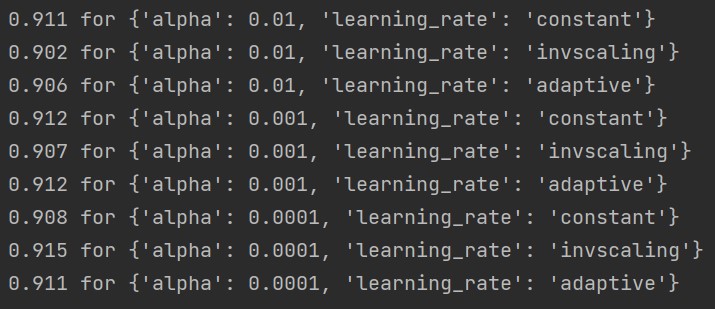
Best result in cross-validation: 1.0.

Pair 3: A and B



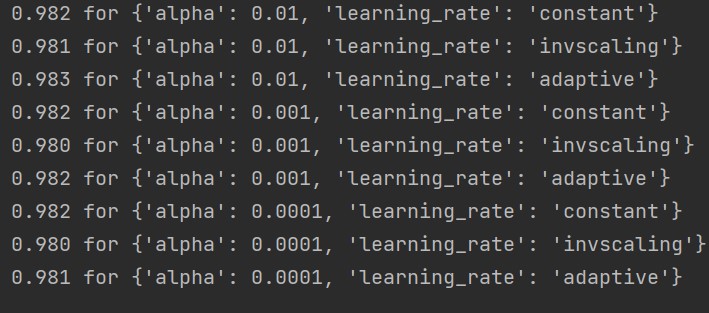
Best result in cross-validation: 0.9978571428571428.

1. The method is PCA. The method is PCA. In PCA, data is transformed from the original coordinate system to the new coordinate system. The choice of a new coordinate system is determined by the data itself. The first new axis selects the direction with the largest variance in the original data, and the second new axis selects the direction that is orthogonal to the first axis and has the greatest variance. This process is repeated for the number of features in the original data. You will find that most of the variance is contained in the first few new axes. Therefore, we can only select the first few coordinate axes, that is, the data is dimensionally reduced. Advantage: reduce the computational cost of the algorithm, avoid noisy.
2. Pair 1: H and K



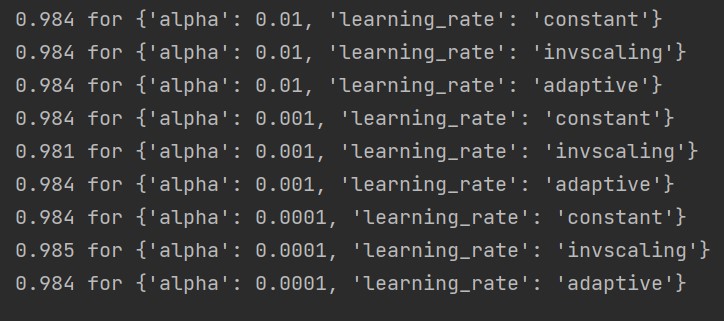
Best result in cross-validation: 0.9154716981132076.

Pair 2: M and Y



Best result in cross-validation: 0.9830985915492958.

Pair 3: A and B



Best result in cross-validation: 0.9849948796722992.

1. Hyperparameters are ‘alpha’ and ‘learning-rate’.

**Discussion**

1. Before dimension reduction

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  |  | KNN | Decision Tree | Random Forest | SVM | Artificial Neural Network |
| H and K | Performance(Best result in cross-validation) | 0.9479245283018868 | 0.9343396226415095 | 0.9486792452830188 | 0.8347169811320754 | 0.9773584905660379 |
| Run time(fit time) | 0.00030794 | 0.00283959 | 0.10282288 | 0.03439016 | 1.76284561 |
| M and Y | performance | 0.9985915492957746 | 0.8784905660377358 | 0.995774647887324 | 0.9774647887323944 | 1.0 |
| Run time | 0.00040011 | 0.00159845 | 0.10129547 | 0.03834763 | 0.83972168 |
| A and B | performance | 1.0 | 1.0 | 1.0 | 0.9485330261136713 | 0.9978571428571428 |
| Run time | 0.00059272 | 0.00139999 | 0.1081883 | 0.04058228 | 0.882418 |

1. After dimension reduction

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  |  | KNN | Decision Tree | Random Forest | SVM | Artificial Neural Network |
| H and K | performance | 0.2920754716981132 | 0.8958490566037736 | 0.979267793138761 | 0.9109433962264152 | 0.9154716981132076 |
| Run time | 0.00090333 | 0.00082231 | 0.09867988 | 0.0255929 | 3.00140815 |
| M and Y | performance | 0.9626760563380282 | 0.9901408450704224 | 0.991549295774648 | 0.9830985915492958 | 0.9830985915492958 |
| Run time | 0.00078974 | 0.00080872 | 0.09299951 | 0.02519207 | 6.15156178 |
| A and B | performance | 0.5375166410650282 | 0.9728315412186381 | 0.9785535074244752 | 0.9849897593445981 | 0.9849948796722992 |
| Run time | 0.00060096 | 0.0007987 | 0.10241613 | 0.02301206 | 3.6501545 |

1. I would choose Decision Tree. Because it accuracy is high and time costs is low. Dimension reduction will reduce the run time but decrease the accuracy for most of the models. For the new dataset, I will see if there some missing values, and then use the suit model according to the dataset’s size. And I find in this problem, some models are all work well, like Decision Tree, Random Forest and SVM.