3. Conclusion

3.1.1　　(3 points) Which conclusions can we draw about the five algorithms examined during this assignment? For each algorithm briefly discuss the key thing you noticed about it during this assignment. Use no more than 250 words in total (+- 50 words per algorithm).

- Gaussian:

Due to the limitation of the assumption of independent predictors, Gaussian performs worst in both tasks. Therefore, when dealing with the tasks relevant to social investigation and image recognition, it may not be the first choice.

- DecisionTreeClassifier:

DecisionTreeClassifier ranks at third place in the first task but performs worse in the second task. Since predicting continuous values is weak for it, when dealing with the dataset containing lots of continuous features, it cannot be the first choice. In American census dataset, it at least has 8 discrete features, so it performs not the worst one.

- KNeighborsClassifier:

Rather than "drawing boundaries for different classes", it uses "similarity between the neighbor" to judge the class of a sample. When people distinguish two images, they compare the similarity between pixels. Similarly, so it can perform well in the second task. Also, after re-scale preprocessing, it can perform even better.

- SVC:

SVC is the best classifier in our tasks because it can change its kernel to fit various datasets. It works well with even unstructured and semi-structured data like images and trees. It is a good tool to solve local optima problem with high dimensions.

- LogisticRegression:

In both tasks, although it is not the best one, it performs better than several algorithms. We thought this kind of algorithms can be extended to be a more complex and powerful classifier. Indeed, we found that many other powerful and widely used algorithms are based on it, like neural networks.