Machine Learning Report

03

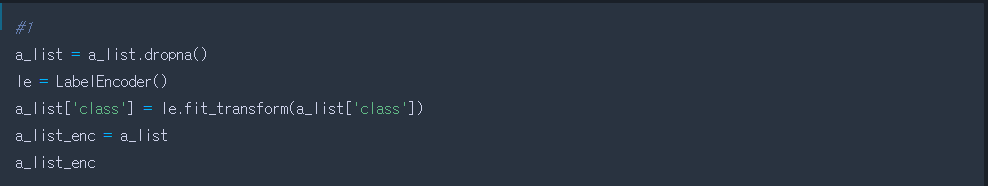
학번: 2017112200

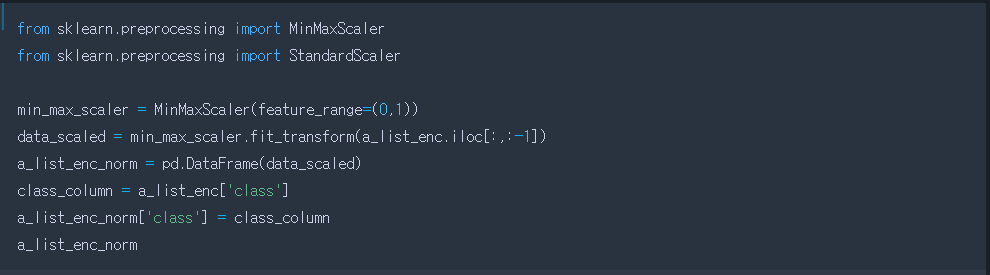
이름: 신현호

1. Preprocessing

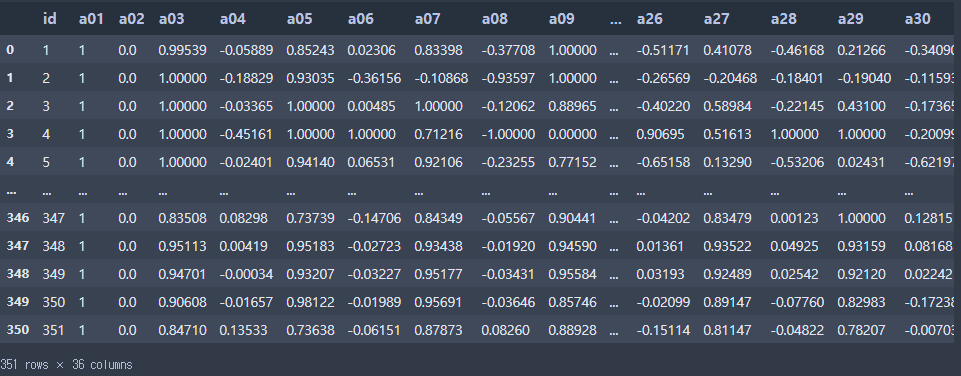
If your data needs preprocssing (Label Encoding & Normalization), please do that.

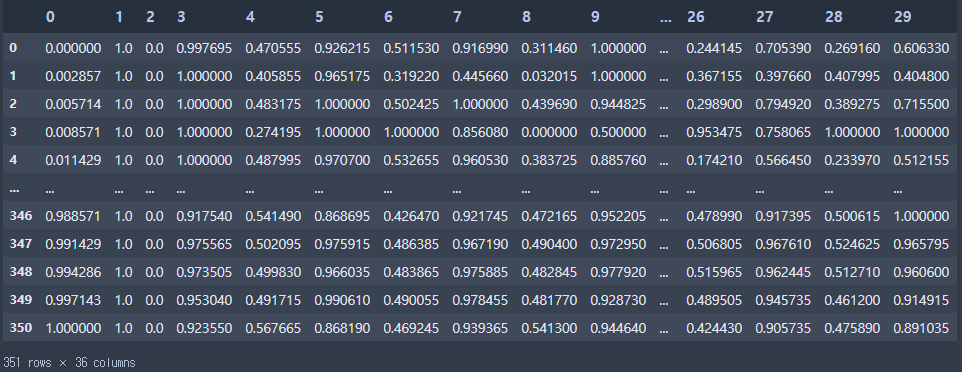
<Program Code>





<Result>

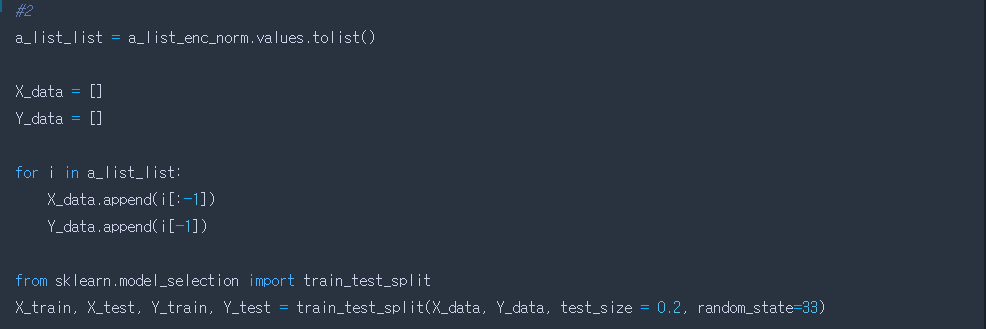




1. Divide into train & test

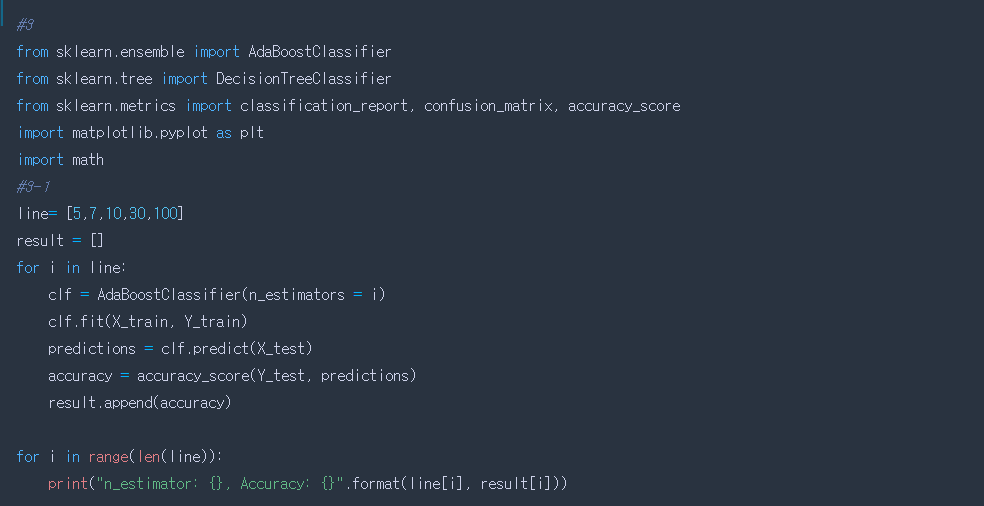
From a\_list, construct X\_train, X\_test, Y\_train, Y\_test

<Program Code>

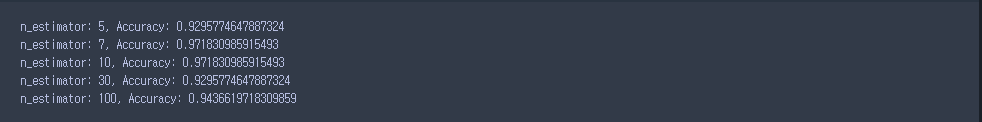


1. Running AdaBoost
2. run by changing n\_estimators = 5, 7, 10, 30, 100, and show the accuracies of each run.

<Program Code>

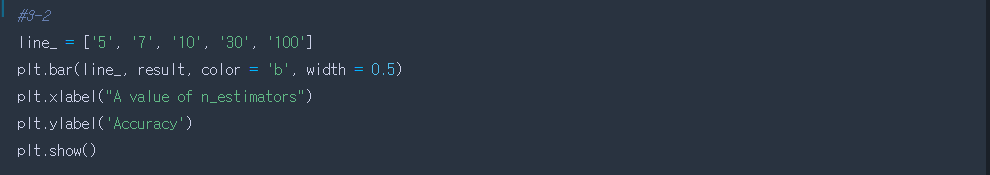


<Result>

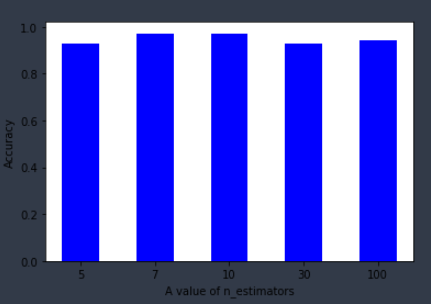


1. Plot your results and explain the effect of the n\_estimators.

<Program Code>



<Result>

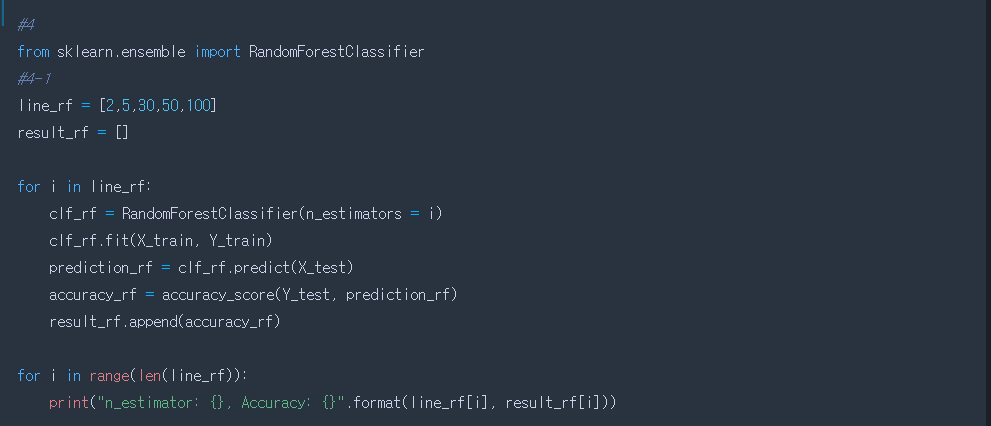


1. Compare the best performance of AdaBoostClassifier with that of IBL. Which is better ? Explain the results.

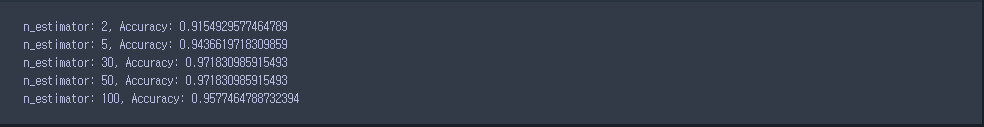
* 저번 과제에서 IBL을 사용했을 때에는 최대 정확치가 0.87 정도였지만, 이번 과제에서의 AdaBoostClassifier을 사용했을 때는 모든 정확치가 0.9를 넘었다. 따라서 AdaBoostClassifier가 IBL에 비해 더 좋은 성능을 보였다.

1. Running Random Forest
2. Run by changing n\_estimators = 2, 5, 30, 50, 100, respectively, and show the accuracies of each run.

<Program Code>

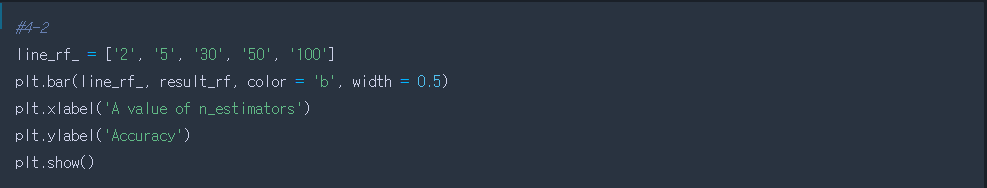


<Result>

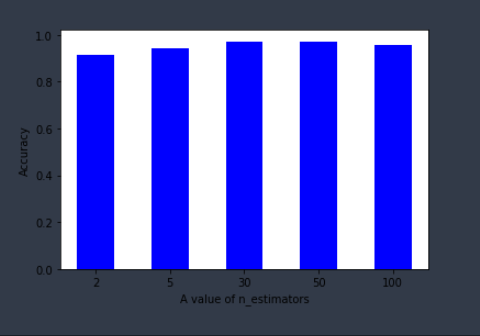


1. Plot your results and explain the effect of the n\_estimators

<Program Code>

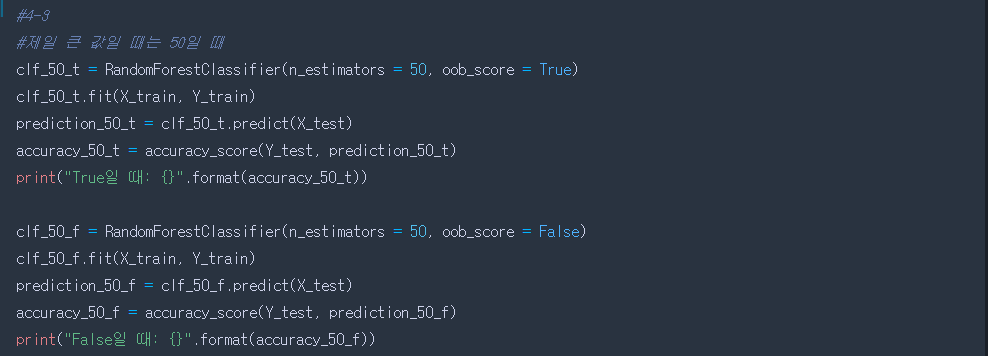


<Result>



1. Choose the optimal n\_estimators from q. 1), and run the model by changing oob\_score = True/False. respectively. Show the accuracies of each run, and explain the effect of the oob\_score.

<Program Code>



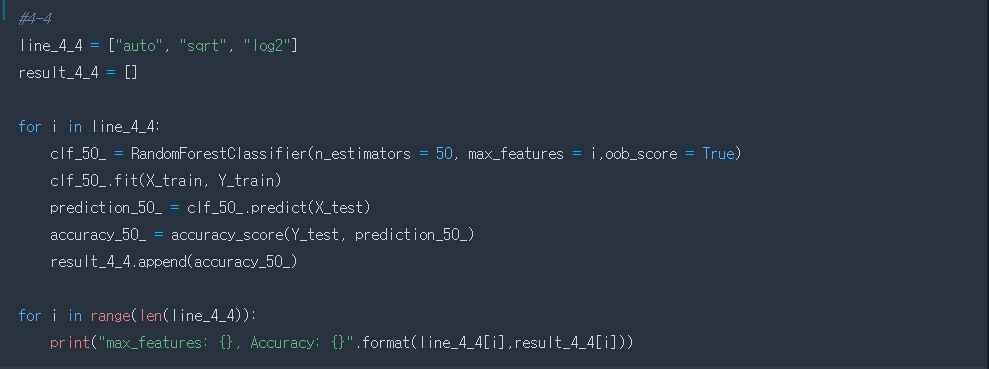
<Result>



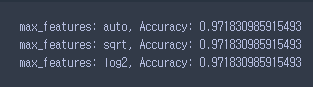
* oob\_score가 True일 때의 정확도가 False일 때보다 높다.

1. Choose the optimal n\_estimators from q. 1), and run the model by changing max\_features = “auto”, “sqrt”, “log2”, respectively. Show the accuracies of each run, and explain the effect of the max\_features.

<Program Code>



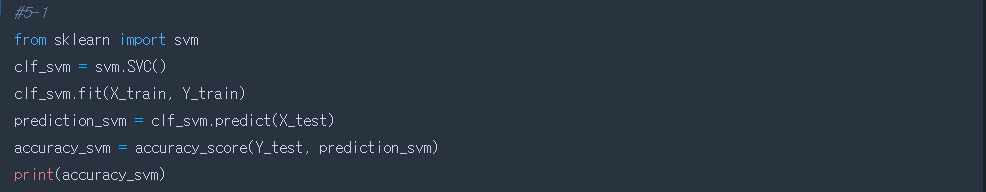
<Result>



* 결과값을 보면, 정확도가 모두 같기 때문에, 영향을 미치지 않았다.

1. Running SVM
2. Calculate the accuracy of SVC

<Program Code>

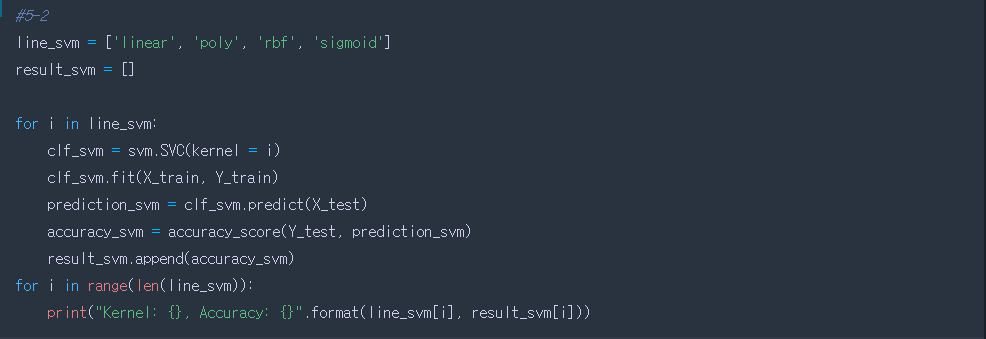


<Result>

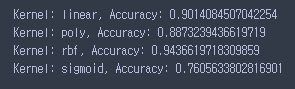


1. Run SVC by changing kernel to ‘linear’, ‘poly’, ‘rbf’, and ‘sigmoid’, and show the accuracies of each. Which kernel function shows the best accuracy ? and explain why ?

<Program Code>



<Result>



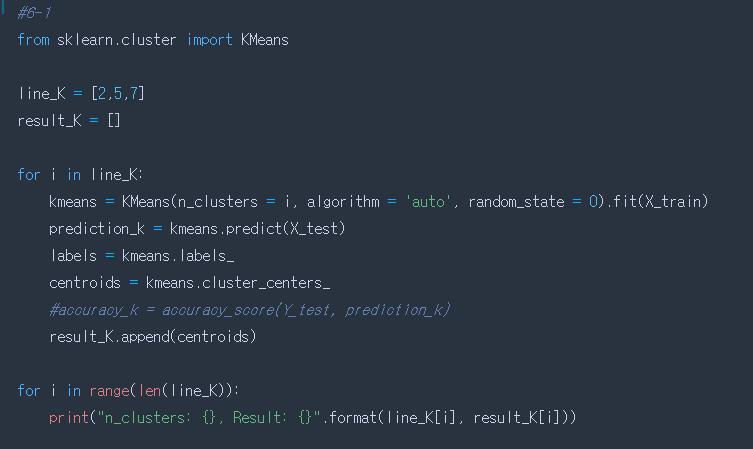
* rbf 커널을 사용할 때, 가장 결과값이 좋았다.

1. Compare the accuracy of SVC with that of IBL, RandomForest, and AdaBoost, respectively. Explain the results.

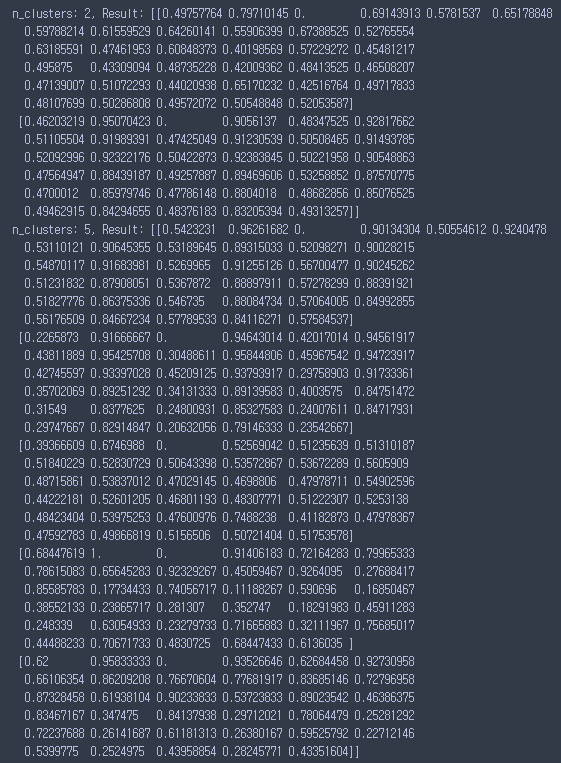
* 저번 과제에서의 IBL의 경우는 정확도가 0.8대였고, AdaBoostClassifier와 RandomForest를 썼을 경우는 0.9대였다. 둘의 정확도는 거의 엇비슷했으며, 제일 성능이 낮은 Classifier는 IBL이었다.

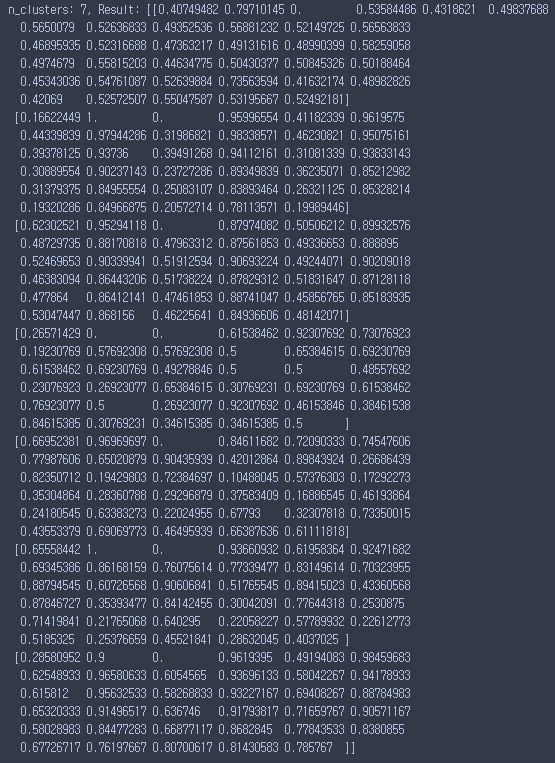
1. Clustering(K Means)
2. Run KMeans 3 times by changing n\_clusters = 2, 5, 7 respectively and show the mean of each cluster.

<Program Code>



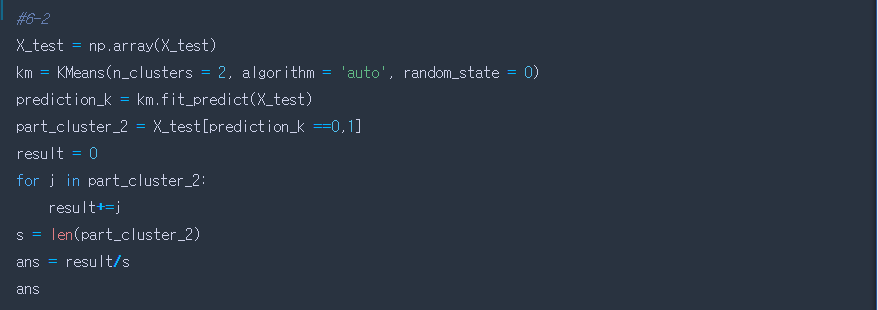
<Result>





1. For the clustering of k=2, pick one cluster. Calculate the average value of each attribute of the data in that cluster.

<Program Code>

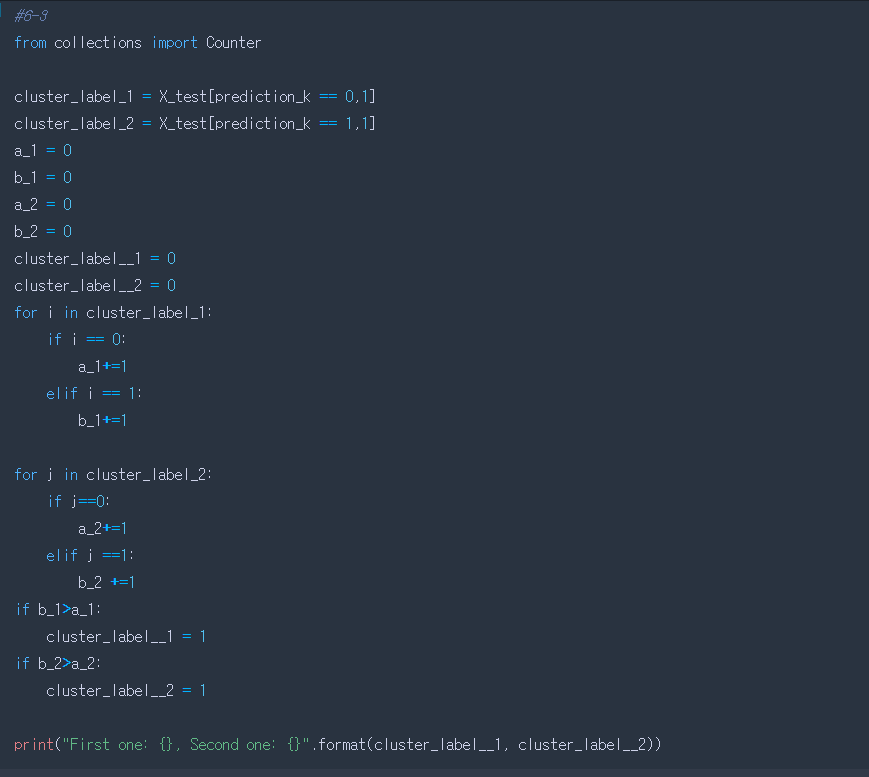


<Result>



1. For each cluster, calculate majority (the most frequent) value of class/target value. (let’s call this ‘cluster label’)

<Program Code>

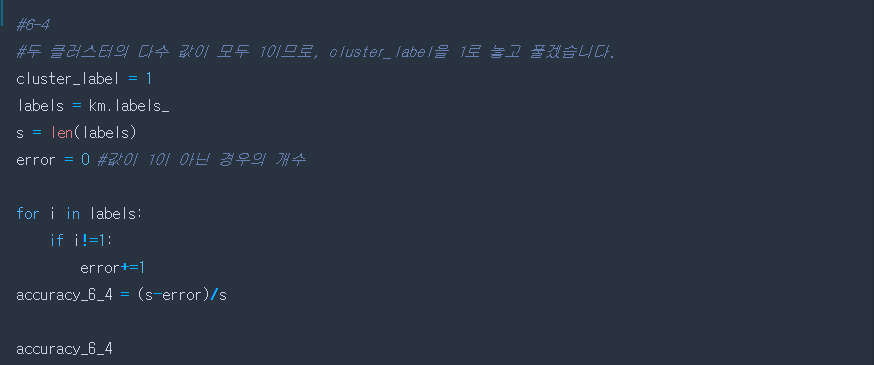


<Result>



1. Suppose each of X\_test is classified based on ‘cluster labels’, calculate the accuracy.

<Program Code>

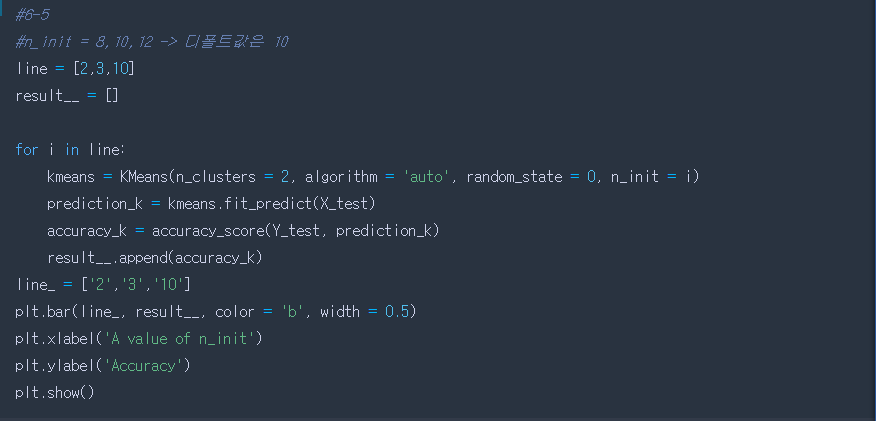


<Result>

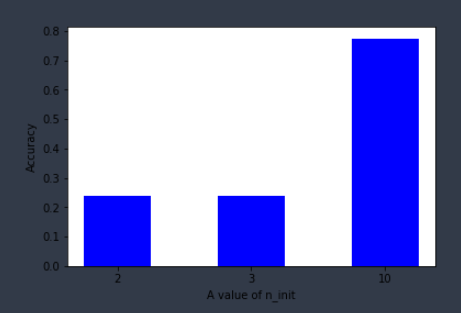


1. Run KMeans 3 times by changing n\_init values (your own choice of n\_init). Compare the performance of each.

<Program Code>

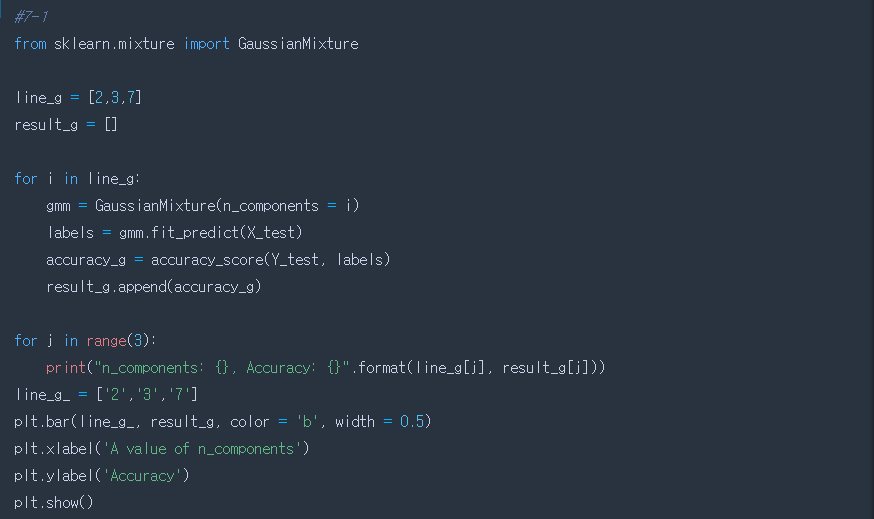


<Result>

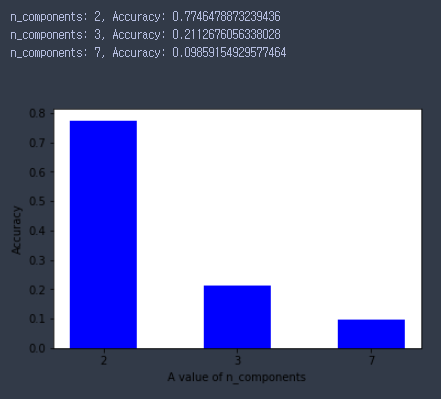


1. Clustering(EM)
2. Run GaussianMixture 3 times by changing n\_clusters = 2, 3, 7 respectively.

<Program Code>

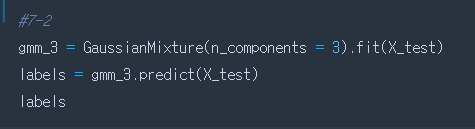


<Result>

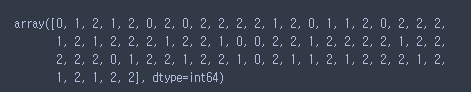


1. For the clustering of k=3, show the predicted labels for the input data.

<Program Code>

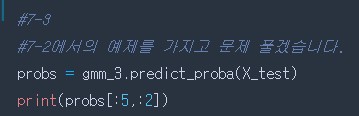


<Result>



1. Show the probabilistic cluster assignments. This returns a matrix of size [n\_samples, n\_clusters].

<Program Code>



<Result>

