

# CA04 – Ensemble Models

## 1. Data Source and Contents

Same as CA03

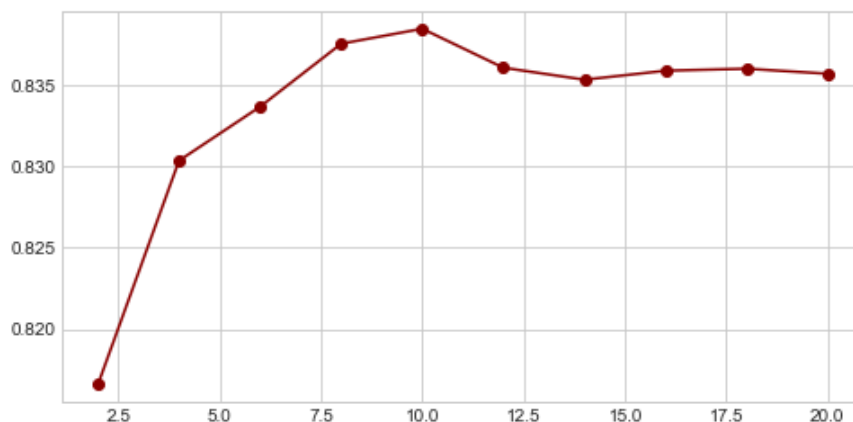
## 2. Finding Optimal Value of a key Hyper-parameter

For Ensemble Models, one of the key hyper-parameter is number of “estimators”. Following is an example of finding the optimal value of “Maximum Depth”. Review the following code snippet to understand how the optimal value of this hyper-parameter (Max Depth = 10) is found by plotting a graph.

```
In [17]: results = []
max_depth_options = [2,4,6,8,10,12,14,16,18,20]
for trees in max_depth_options:
    model = DecisionTreeClassifier(max_depth=trees, random_state=101)
    model.fit(x_train, y_train)
    y_pred = model.predict(x_test)
    accuracy = np.mean(y_test==y_pred)
    results.append(accuracy)

plt.figure(figsize=(8,4))
pd.Series(results, max_depth_options).plot(color="darkred",marker="o")
```

Out[17]: <matplotlib.axes.\_subplots.AxesSubplot at 0x20703c75630>



## 3. Building a Random Forest Model

Using Notebook, and the same data source from CA03, train a Random Forest Model. Using similar approach of Section 2 above, plot a graph of Accuracy vs. n\_estimator. Use n\_estimator values as [50,100,150,200,250,300,350,400,450,500].

Answer the following question:

1. Write your observations about the Classifier's behavior with respect to the number of estimators
2. Is there an optimal value of the estimator within the given range?

## 4. Building AdaBoost, Gradient Boost (classifier) and XGB Model

Repeat the process of Section 3 above for all three models here

## 5. Compare Performance

Keep all common Hyper-parameters same for four models (Random Forest, AdaBoost, Gradient Boost, XGB), run them again and create a performance comparison table within your code and print the same as follows:

	Random Forest	AdaBoost	Gradient Boost	XGB
Accuracy				
AUC				

Also, print the values of all COMMON Hyper-parameters below this table (from code)

## 6. Deliverables

Your assignment outputs will have the following components:

- (1) Fully functional Notebook, Data, Readme file in a single folder at GitHub
- (2) Word / PDF document with answers to all questions highlighted in yellow.

### \*\*\* Useful Material \*\*\*

Here is a great list of all Model Evaluation Functions in SKlearn library grouped by the Model Type:

[https://scikit-learn.org/stable/modules/model\\_evaluation.html](https://scikit-learn.org/stable/modules/model_evaluation.html)