

# Practical 7

## 2CS501

### 19BCE248

**AIM:** Support Vector Classification and Regression with Grid Search for Hyper-parameter tuning using sklearn

#### About Dataset (Iris):

Iris is a dataset about flowers which is most popular for classification problems. It consists of four columns namely [sepal length (cm), sepal width (cm), petal length (cm), petal width (cm)]. The target value is species of flower which are of three types:

1. Setosa
2. Versicolor
3. Virginica

Rows: 150

Columns: 4

The target column describes the target variable i.e. species of flower according to parameters.

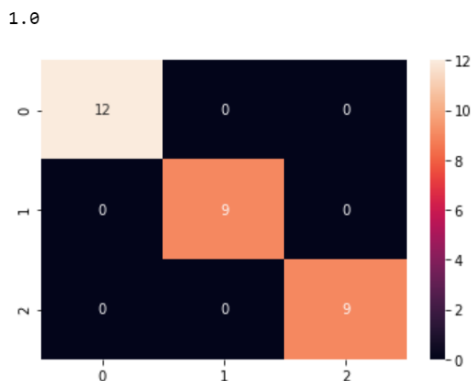
#### Preprocessing:

- Firstly dividing the entire dataset into Training and Testing dataset.

No such preprocessing needed.

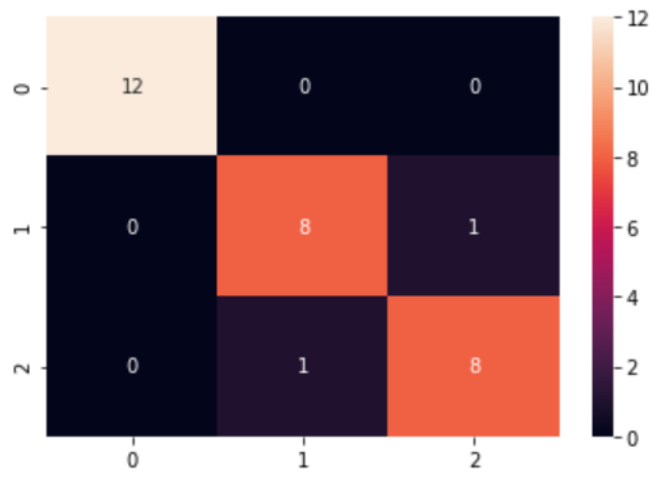
#### Model Implementation:

##### Pure SVC: (rbf)



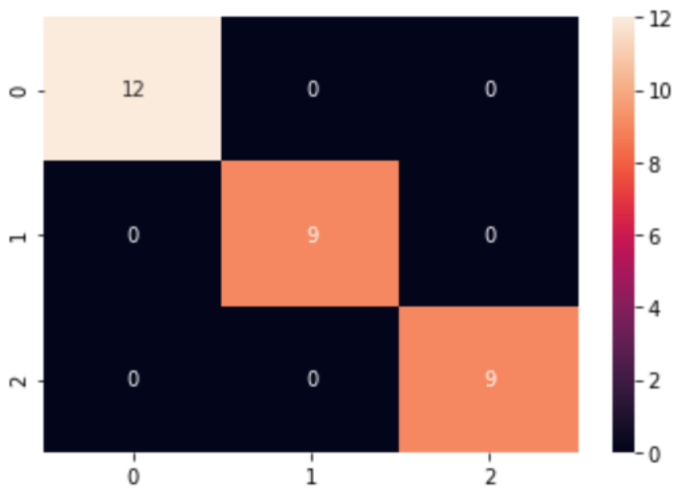
##### Linear SVC:

0.9333333333333333



**NuSVC:**

1.0



**Using GridSearchCV:**

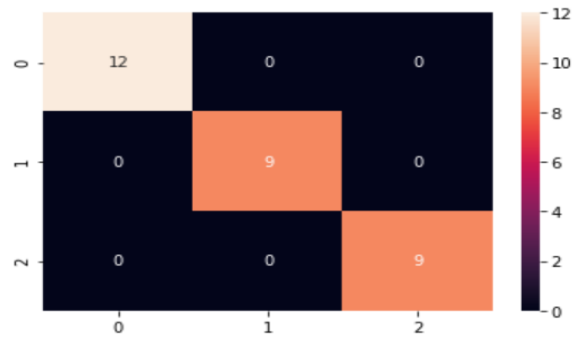
```

gsv.best_params_

{'C': 0.1, 'kernel': 'linear'}

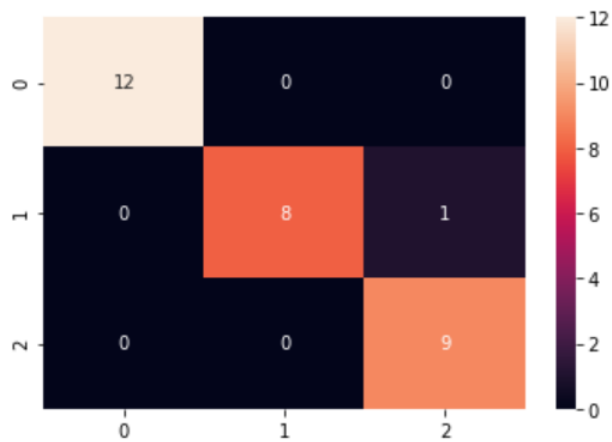
y_pred=gsv.predict(X_test)
cm=confusion_matrix(y_test,y_pred)
acc=accuracy_score(y_test,y_pred)
sns.heatmap(cm,annot=True)
acc
1.0

```

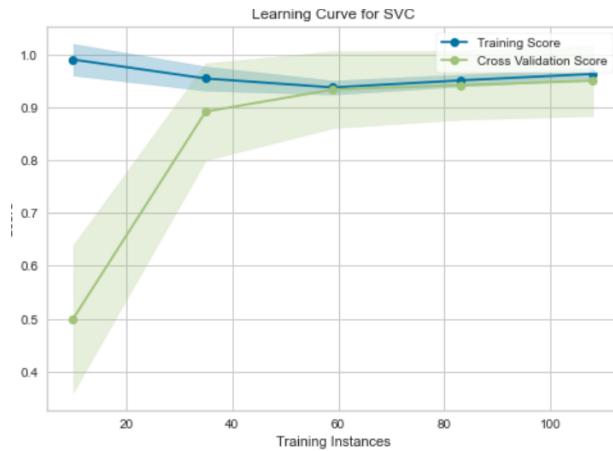


**Doing Separately for Polynomial as degree parameter also needs to be included.**

```
0.9666666666666667
```



**Learning Curve:**



### Conclusion:

So to conclude SVM can be very beneficial in cases like:

- 1) When the number of features (variables) and number of training data is very large.
- 2) When sparsity in the problem is very high, i.e., most of the features have zero value.
- 3) It is the best for document classification problems.
- 4) It also performs very well for problems like image classification, genes classification, drug disambiguation etc. where the number of features are high.