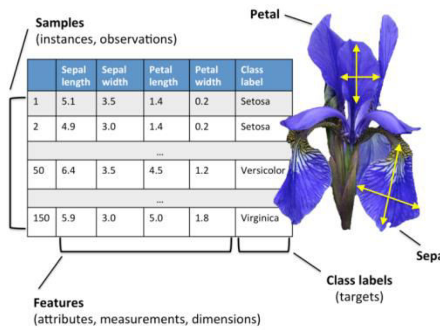
Support Vector Machine Experiment Analysis

CmpE530
Mine Melodi Çalışkan
2015705009

SVM Experiment on Iris Data Set

- The iris dataset contains measurements for 150 iris flowers from three different species.



The three classes in the Iris dataset are:

- *Iris-setosa* (n=50)
- *Iris-versicolor* (n=50)
- *Iris-virginica* (n=50)

And the four features of in Iris dataset are:

- *sepal length* in cm
- *sepal width* in cm
- *petal length* in cm
- *petal width* in cm

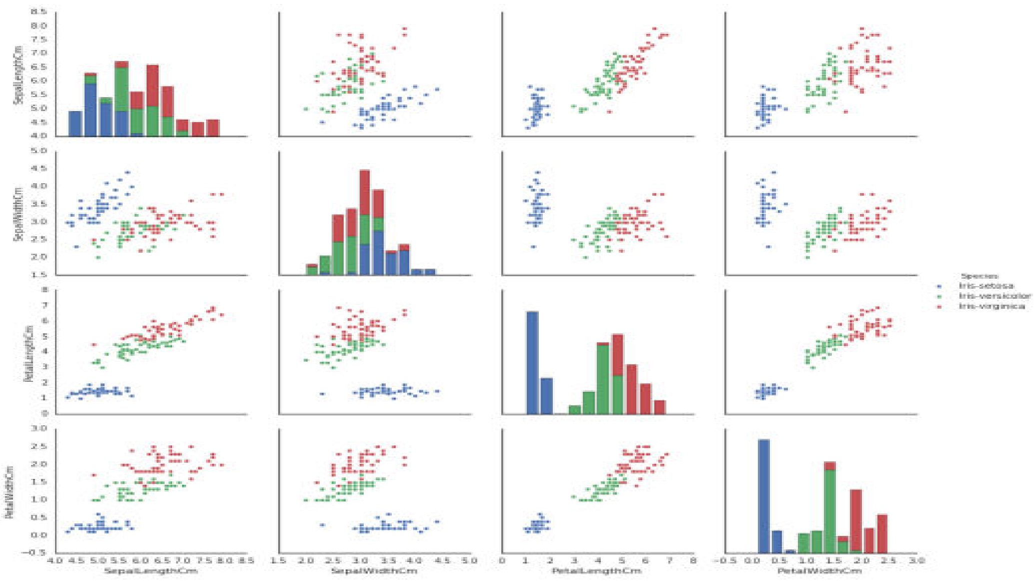
We store iris dataset in form of a 150×4 matrix where the columns are the different features, and every row represents a separate flower sample.

- Each sample row x can be pictured as a 4-dimensional vector.

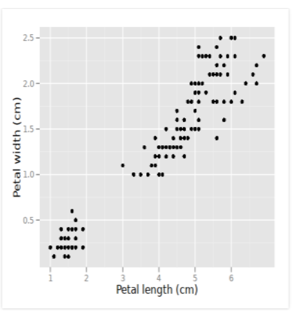
$$\mathbf{x}^{\mathrm{T}} = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} = \begin{pmatrix} \text{sepal length} \\ \text{sepal width} \\ \text{petal length} \\ \text{petal width} \end{pmatrix}$$

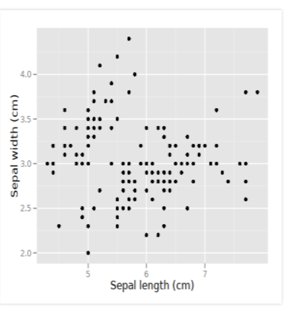
Observations of the data set

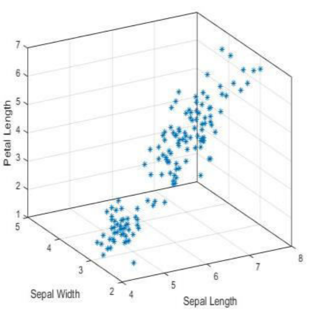
- <u>Pairplot with labels</u>



- Plots of the data without labels with 2 and 3 features







<u>Remark.</u>

- From the above graphs it can be seen that the Iris dataset is not easy to graph for predictive analytics in its original form because we cannot plot all four coordinates (from the features) of the dataset onto a two-dimensional screen.
- We can reduce the dimensions by applying a *dimensionality reduction* algorithm to the features.
- In this case, the algorithm for the data transformation (reducing the dimensions of the features) is called Principal Component Analysis (PCA).

- After PCA, the plot with the reduced feature set which represents four features with two items:

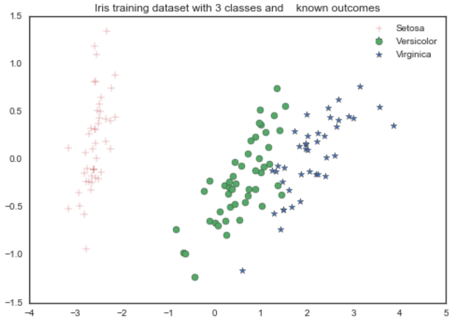


Fig. From this plot you can clearly tell that the Setosa class is linearly separable from the other two classes. While the Versicolor and Virginica classes are not completely separable by a straight line.

- A plot of the Support Vector Machine (SVM) model trained with a dataset that has been dimensionally reduced to two features.

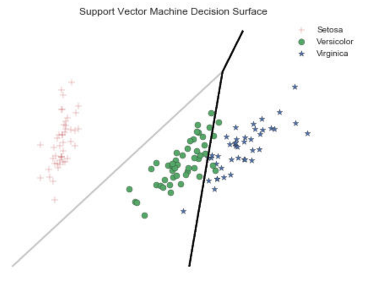
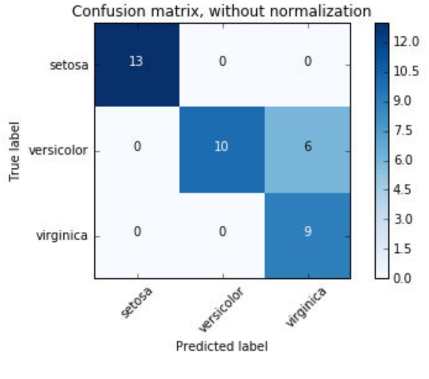
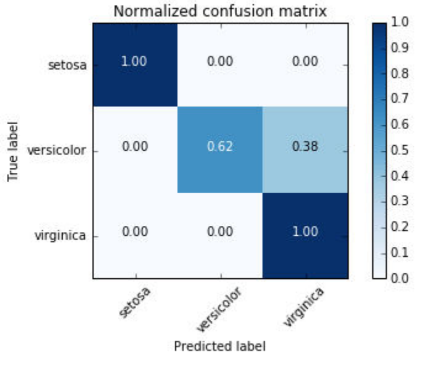


Fig. The *decision surface* for the classifier — the area in the graph that represents the decision function that SVM uses to determine the outcome of new data input.

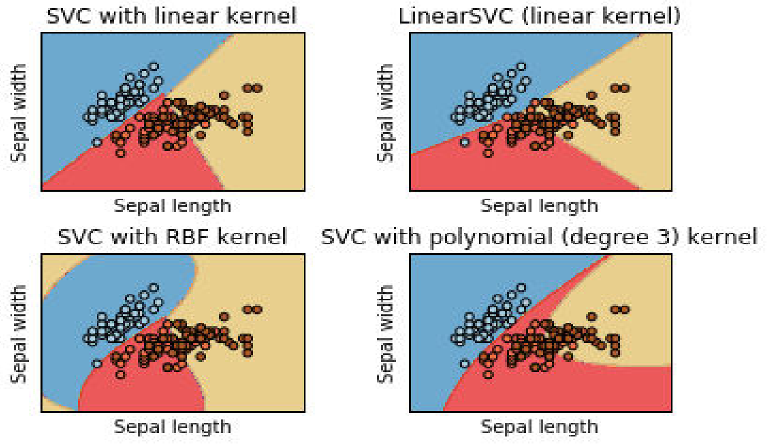
- The lines separate the areas where the model will predict the particular class that a data point belongs to. The left section of the plot will predict the Setosa class, the middle section will predict the Versicolor class, and the right section will predict the Virginica class.

Visualization of The Performance of the algorithm by Confusion Matrix

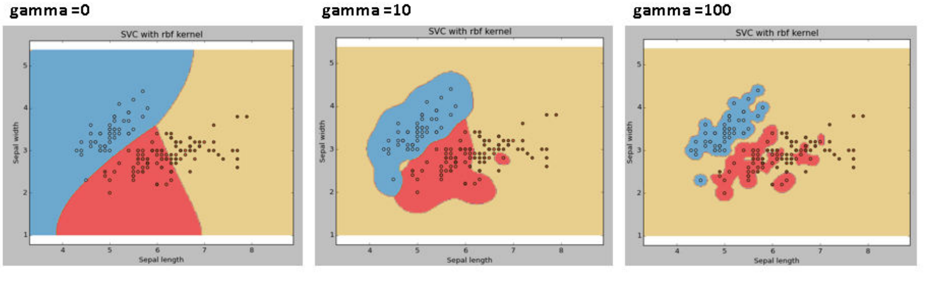




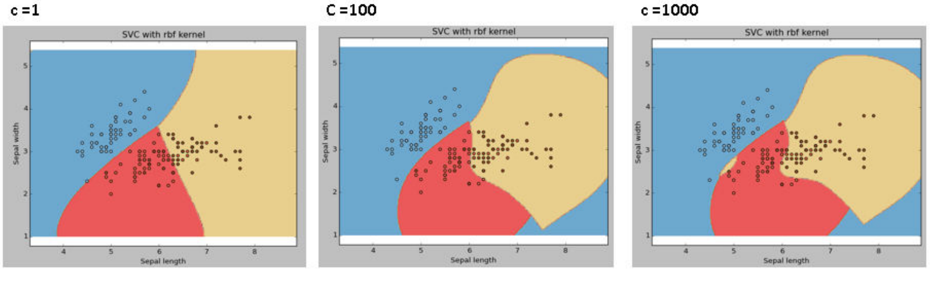
SVM with different Kernels



gamma: Kernel coefficient for ‘rbf’, ‘poly’ and ‘sigmoid’. Higher the value of gamma, will try to exact fit the as per training data set i.e. generalization error and cause over-fitting problem.



C: Penalty parameter C of the error term. It also controls the trade off between smooth decision boundary and classifying the training points correctly.



- We should always look at the cross validation score to have effective combination of these parameters and avoid over-fitting.

Cross-Validation: Accuracy of the predictive model.

1. Results with 10-Folds

- Correctly Classified Instances: %96,667
- Incorrectly Classified Instances %3.3333
- Kappa Statistic 0.95
- Mean Absolute Error 0.0222
- Coverage of cases(0.95 level) %96.667

2. Results with 15-Folds

- Correctly Classified Instances: %98
- Incorrectly Classified Instances %2
- Kappa Statistic 0.97
- Mean Absolute Error 0.0189
- Coverage of cases(0.95 level) %98

Advantages and Disadvantages of SVM

Advantages

- The solution is guaranteed to be the global minimum not a local minimum. (Convex Optimization)
- SVM is useful for both Linearly Separable(hard margin) and Non-linearly Separable(soft margin) data with the optimal penalty variable C .
- Feature mapping is implicitly carried out via simple dot products. (Kernel Trick)
- Memory efficient: It uses a subset of training points in the decision function (support vectors)

Disadvantages

- SVM doesn't perform very well, when the data set has more noise i.e. target classes are overlapping.
- It doesn't directly provide probability estimates, it requires expensive fold cross-validations.

References:

1) Wolfgang Karl Härdle, Dedy Dwi Prastyo, Christian Hafner, «Support Vector Machines with Evolutionary Feature Selection for Default Prediction, 2012»
2) Tan, Steinbach, Kumar «Introduction to Data Mining , 2004»
3) Sunil Ray, «Analytics Vidhya, 2015»