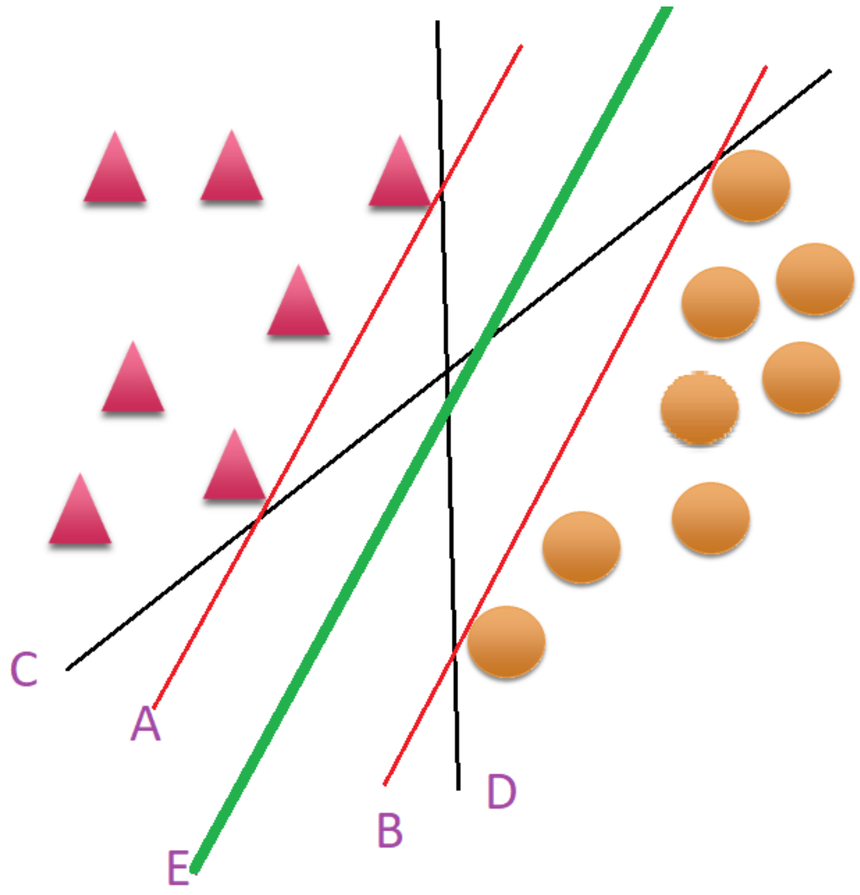
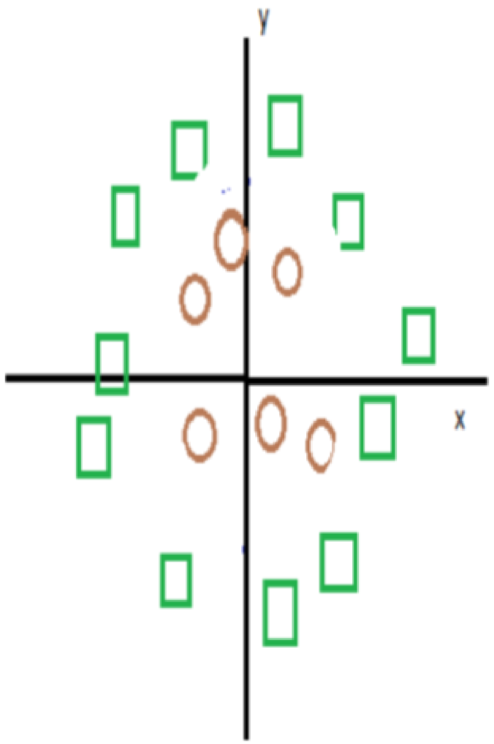
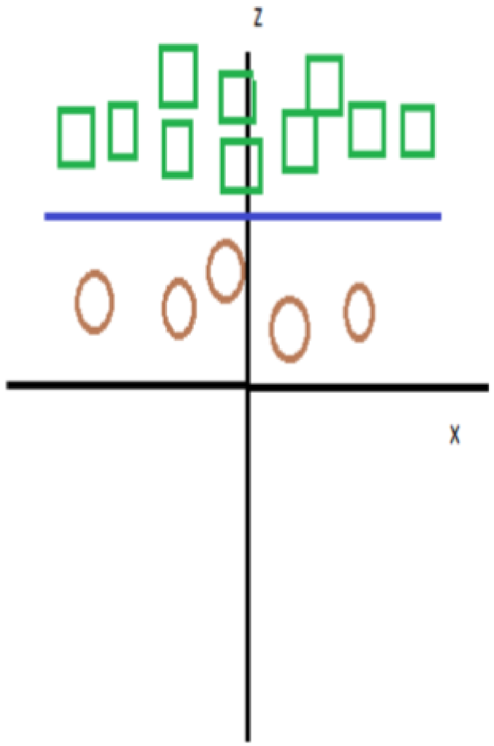
Support Vector Machines

BY MG ANALYTICS

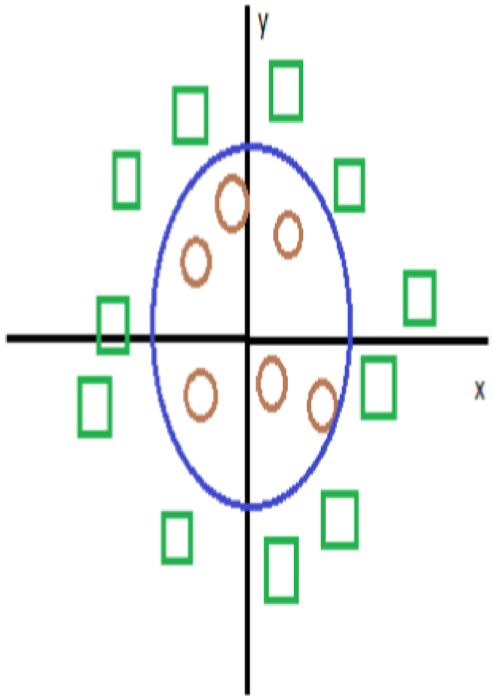

C
A
E
B
D



No linear boundary separates the two classes



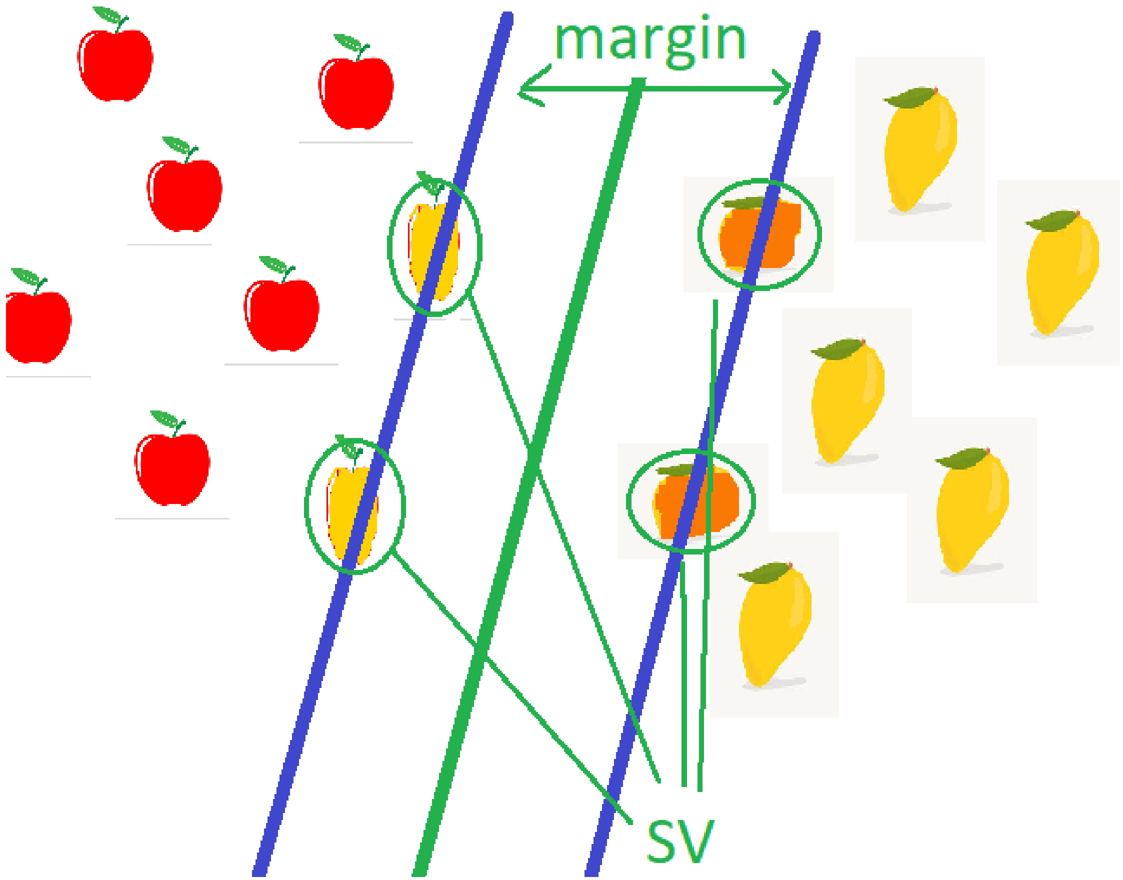
Transformation applied and one more dimension 'z' added; data now linearly separable

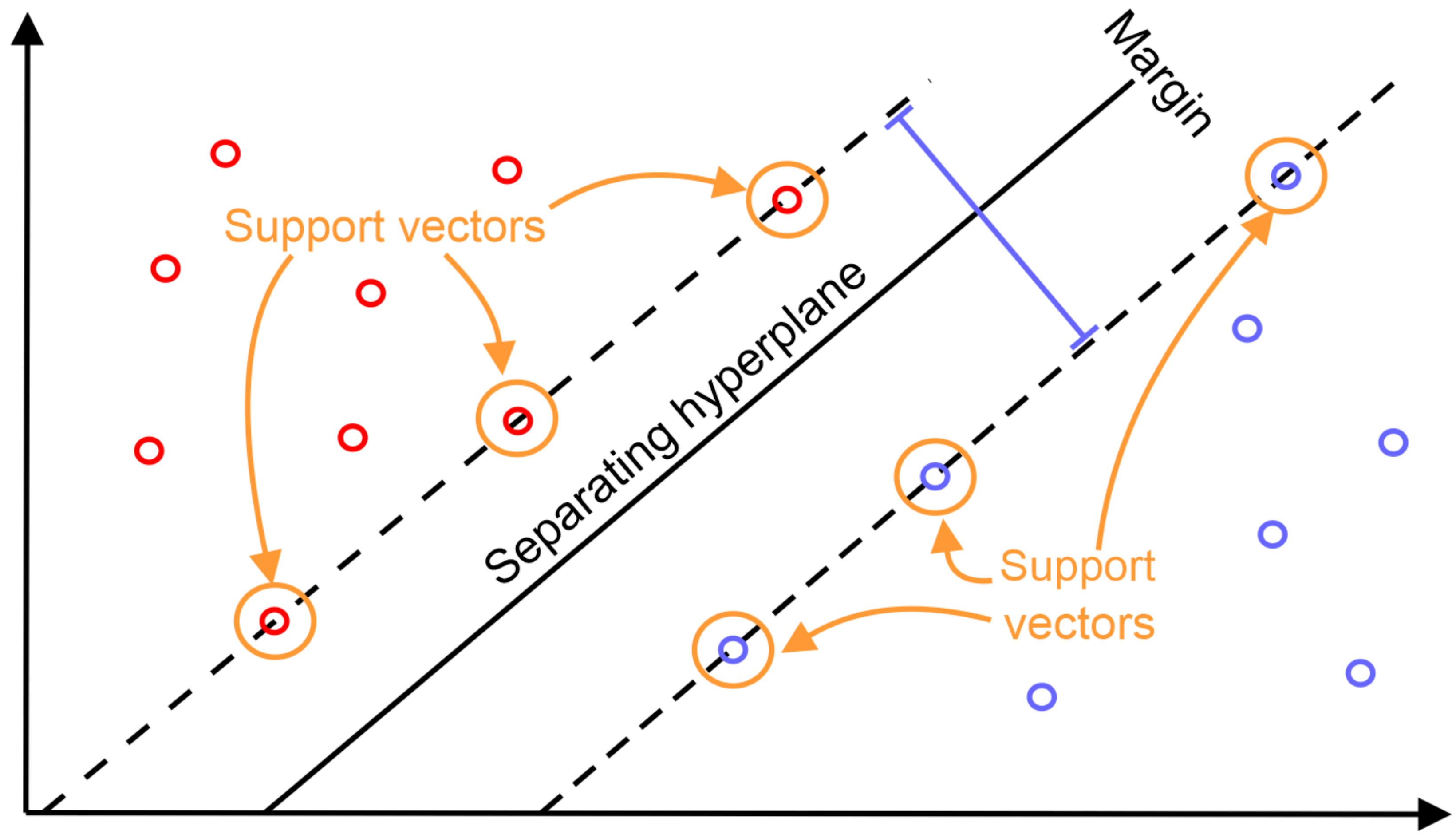


Transformed back to the xy plane; linear separator becomes a circle

SVM

- **finds** the **most similar examples** between classes. Those will be the **support vectors**.
- For example Mango vs apple, other algos will try to find differences between mango and apple i.e.
  - Mango : elliptical , yellow
  - Apple : round , red
- SVM will try to find
  - Mango that looks like apple : red and round.
  - Apple that looks like mango : yellow and elliptical
  - And use these as support vectors.



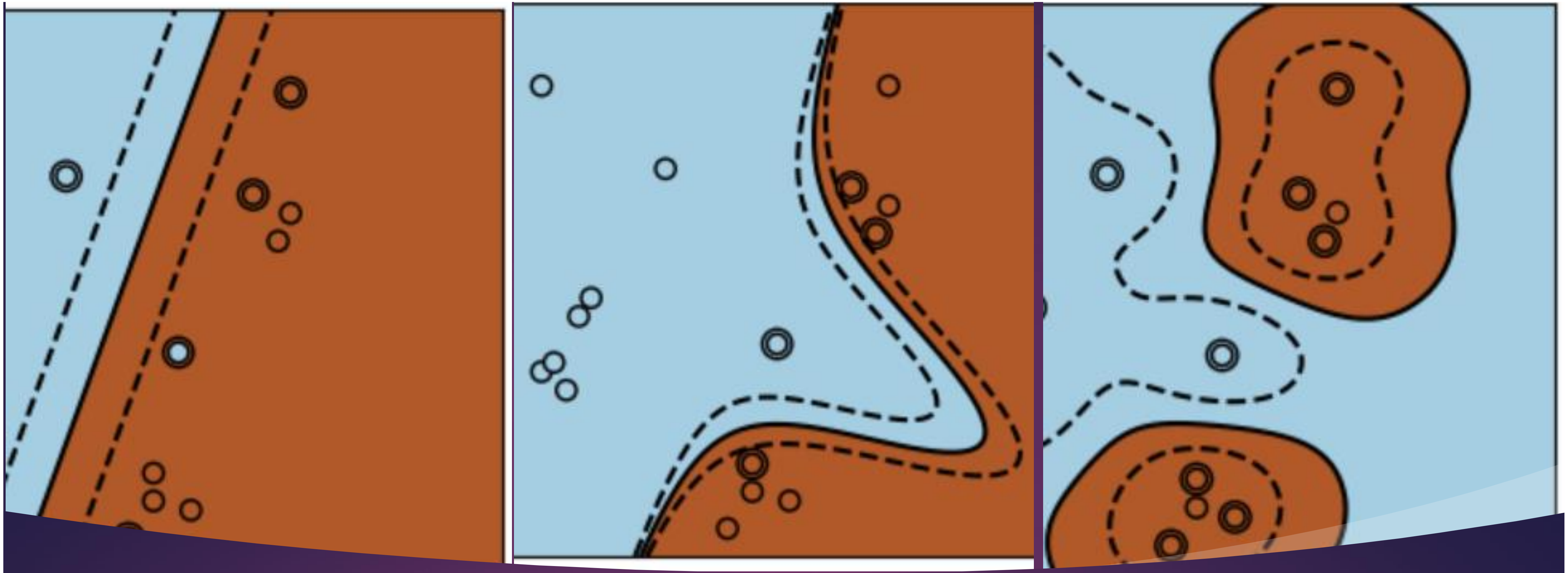

Margin
Support vectors
Separating hyperplane
Support
vectors

Steps

- select **two hyperplanes** (in 2D) which separates the data **with no points between them** (red lines)
- **maximize their distance** (the margin)
- the **average line** (here the line halfway between the two red lines) will be the **decision boundary**

Kernel :

- Mathematical **functions**
- The **function** of **kernel** is to take data as input and transform it into the required form.
- Kernel defines the distance measure between new data and the support vectors i.e. observations closest to the hyperplane.
- **Higher dimensions** kernels : Polynomial Kernel and a Radial Kernel transform the input space into higher dimensions.



Linear

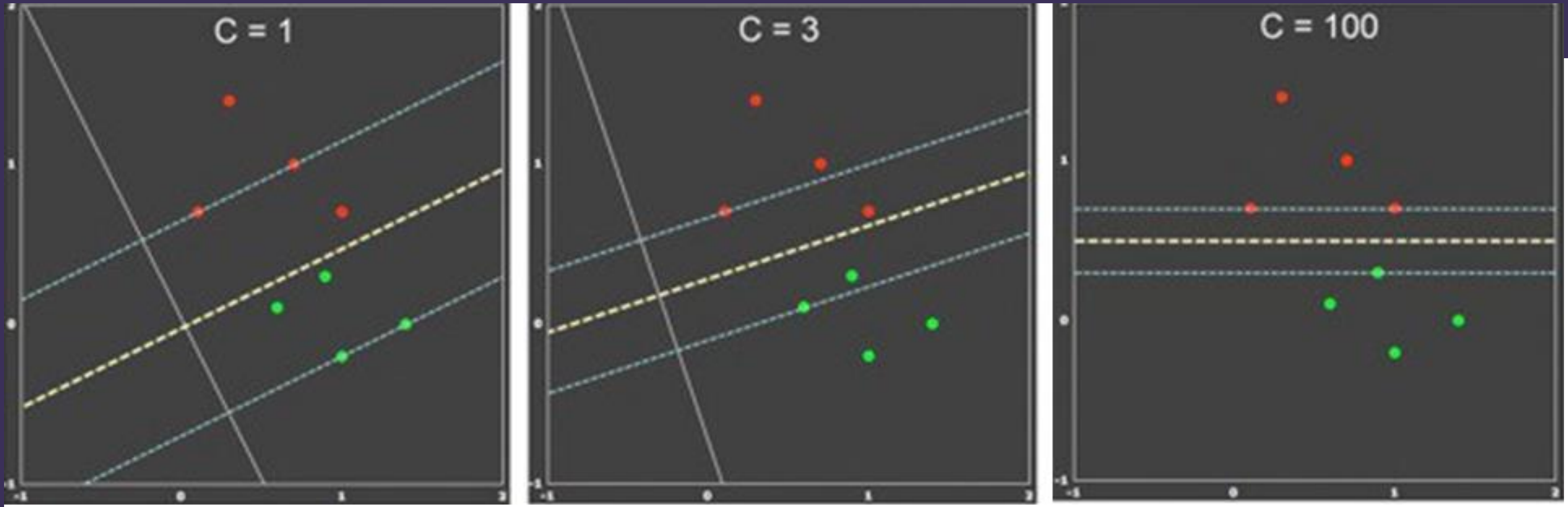
Polynomial

Radial Basis Function(RBF)

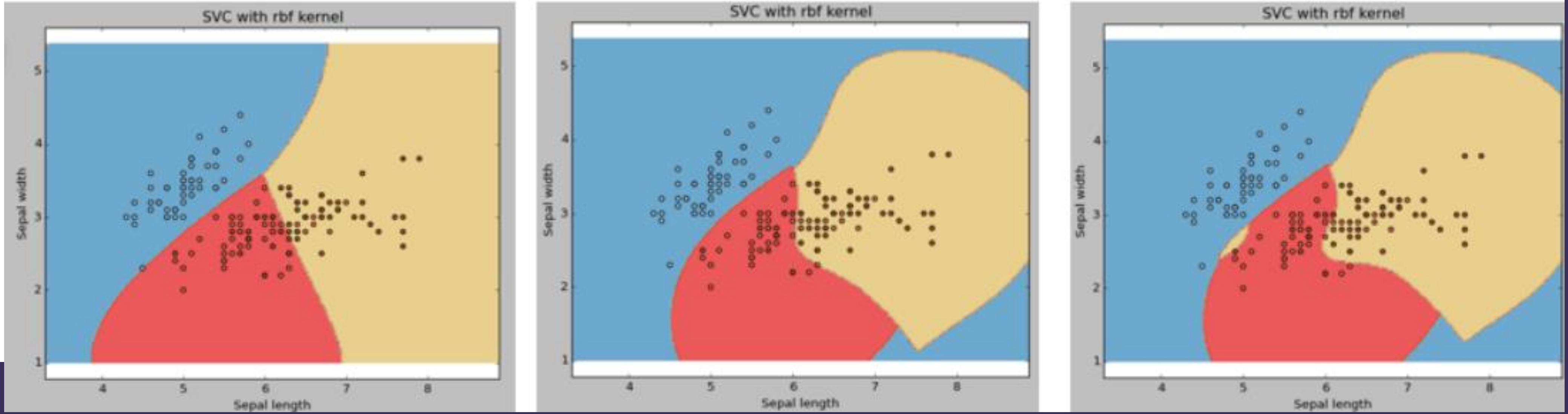
Kernels

C value

- Regularization parameter
- When the value of C is **large**, smaller-margin hyperplane will be considered since it stresses on getting all the training points classified correctly.
- a **small** value of C will consider a larger margin hyperplane, even if some points are misclassified by the hyperplane.



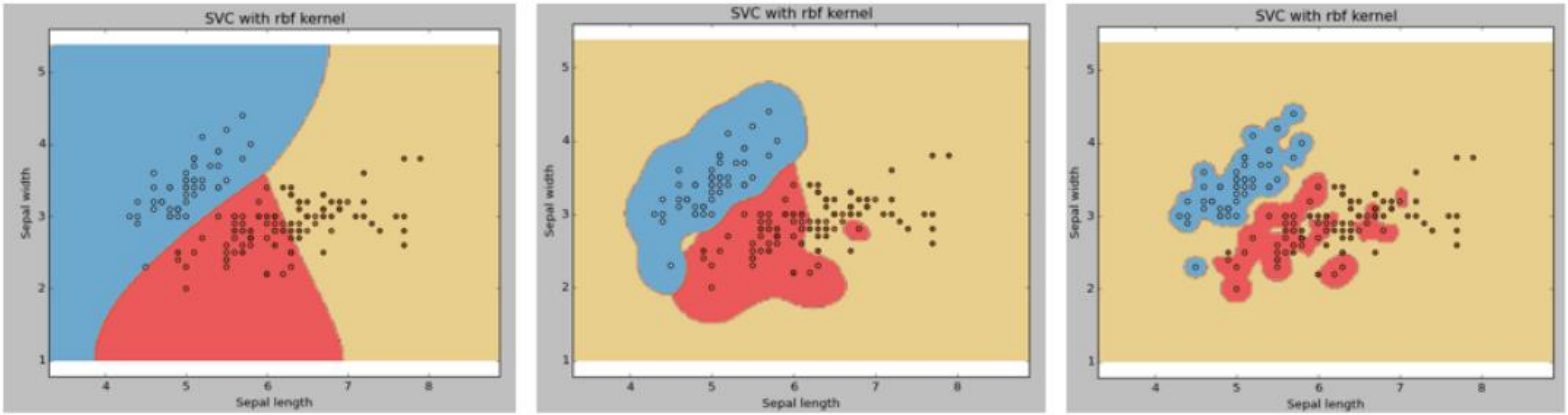
c =1 C =100 c =1000



Gamma Value

- The gamma parameter defines how far the influence of each training observation affects the calculation of the optimal hyperplane.
- defines how far the influence of a single training example reaches
- **low** values meaning 'far'.
- **high** values meaning 'close'.
- The gamma parameters is the inverse of the radius of influence of samples selected by the model as support vectors.

gamma =0 gamma =10 gamma =100



Pros:

- It works really well with a clear margin of separation
- It is effective in high dimensional spaces.
- It is effective in cases where the number of dimensions is greater than the number of samples.
- It uses a subset of training points in the decision function (called support vectors), so it is also memory efficient.

Cons:

- It doesn’t perform well when we have large data set because the required training time is higher
- It also doesn’t perform very well, when the data set has more noise i.e. target classes are overlapping
- SVM doesn’t directly provide probability estimates, these are calculated using an expensive five-fold cross-validation. It is included in the related SVC method of Python scikit-learn library.