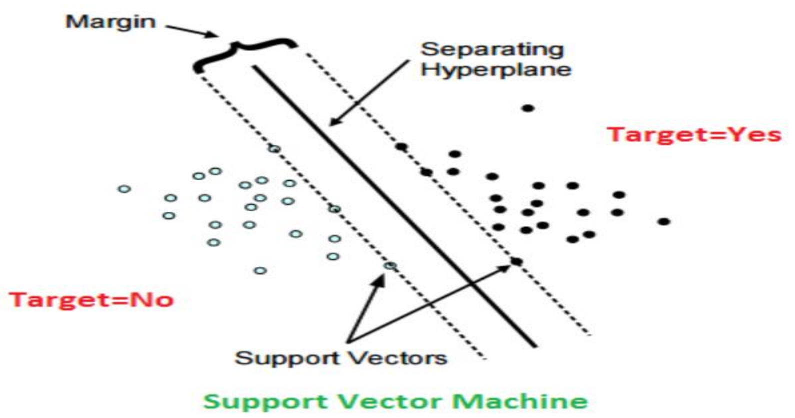
SUPPORT VECTOR MACHINES

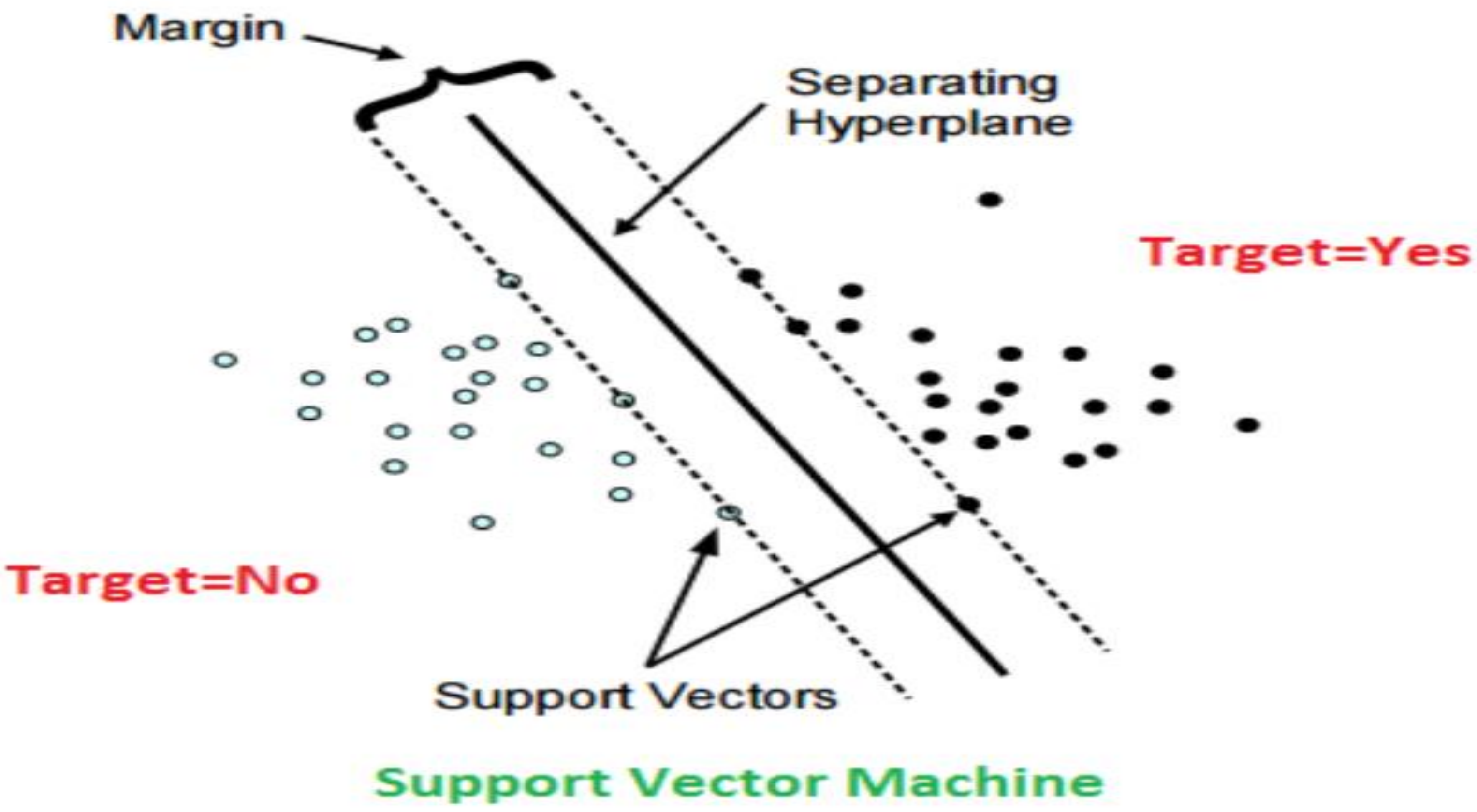
Lab5

What is SVM?

- SVM is a linear model for classification and regression problems. It can solve linear and non-linear problems.
- The algorithm creates a line or a hyperplane which separates the data into classes.
- **Support Vectors:** Support vectors are special data points in the dataset. They are responsible for the construction of the hyperplane and are the closest points to the hyperplane.
- **Margins:** The distance of the vectors from the hyperplane are called the ***margins***.



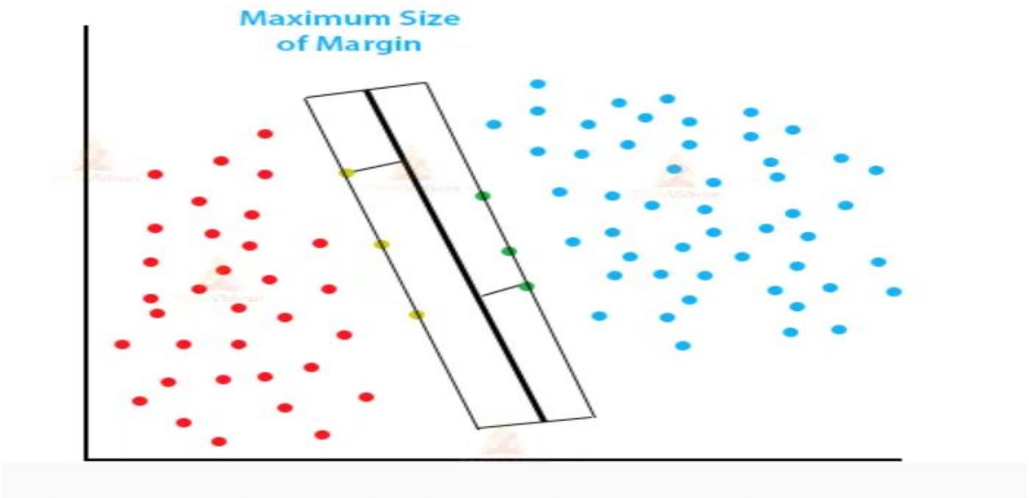
What is SVM?



Margin
Separating
Hyperplane
Target=Yes
Target=No
Support Vectors
Support Vector Machine

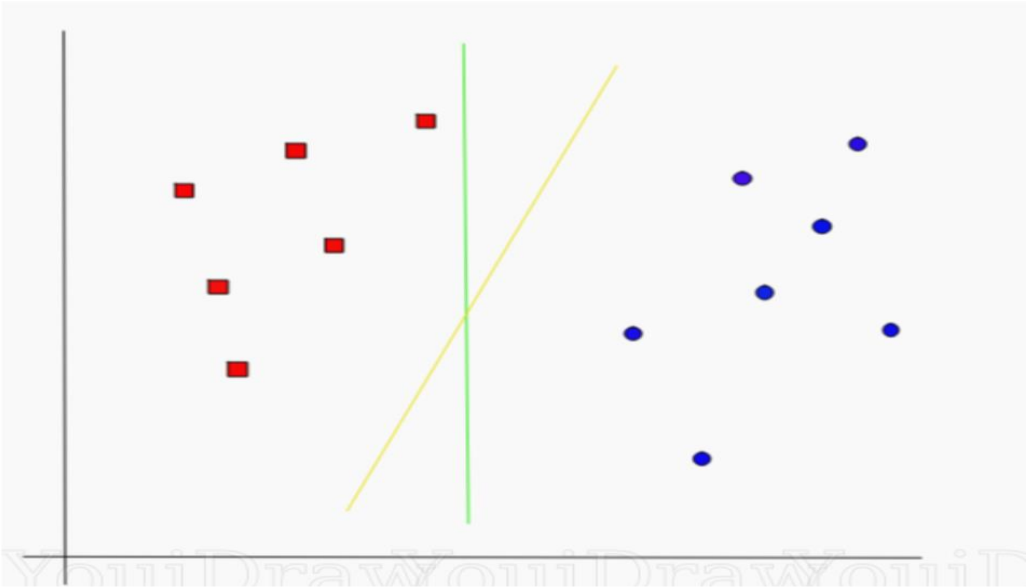
What is SVM?

- **Hyperplane:** The hyperplane is the central line in the diagram below In this case, the hyperplane is a line because the dimension is 2-D. If we had a 3-D plane, the hyperplane would have been a 2-D plane itself.
- **Decision Boundaries:** Decision boundaries in SVM are the two lines that we see alongside the hyperplane.
- The distance between the two light-toned lines is called the margin. **An optimal or best hyperplane form when the margin size is maximum.**



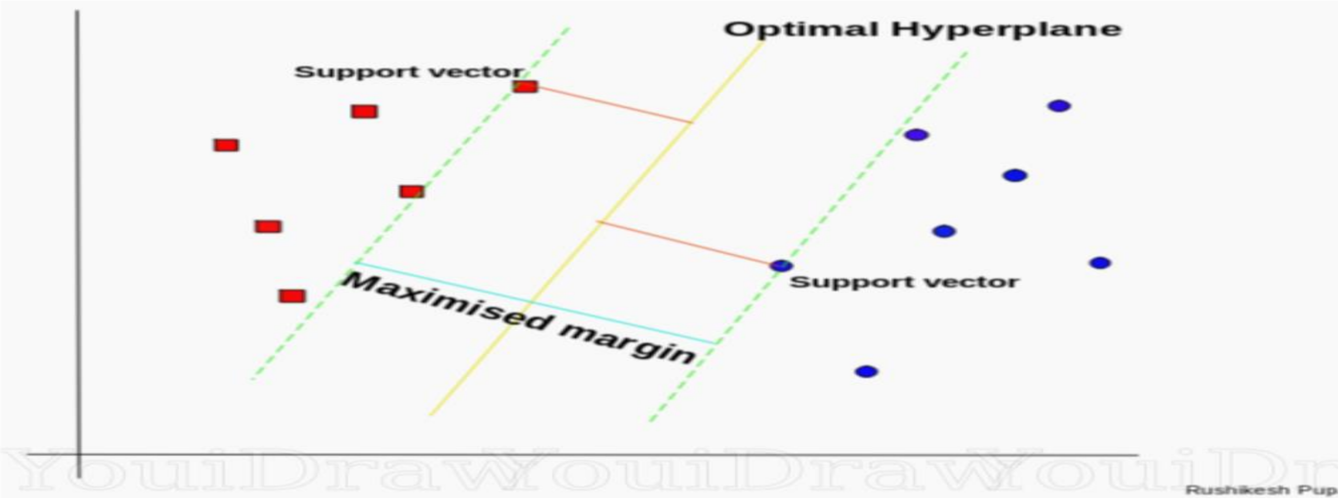
What is SVM?

- At first approximation what SVMs does is to find a separating line(or hyperplane) between data of two classes. SVM is an algorithm that takes the data as an input and outputs a line that separates those classes if possible.
- In the example below, which line is a better separator?
- Our goal is to find a generalized separator.



SVM IN LINEARLY SEPARABLE CASES

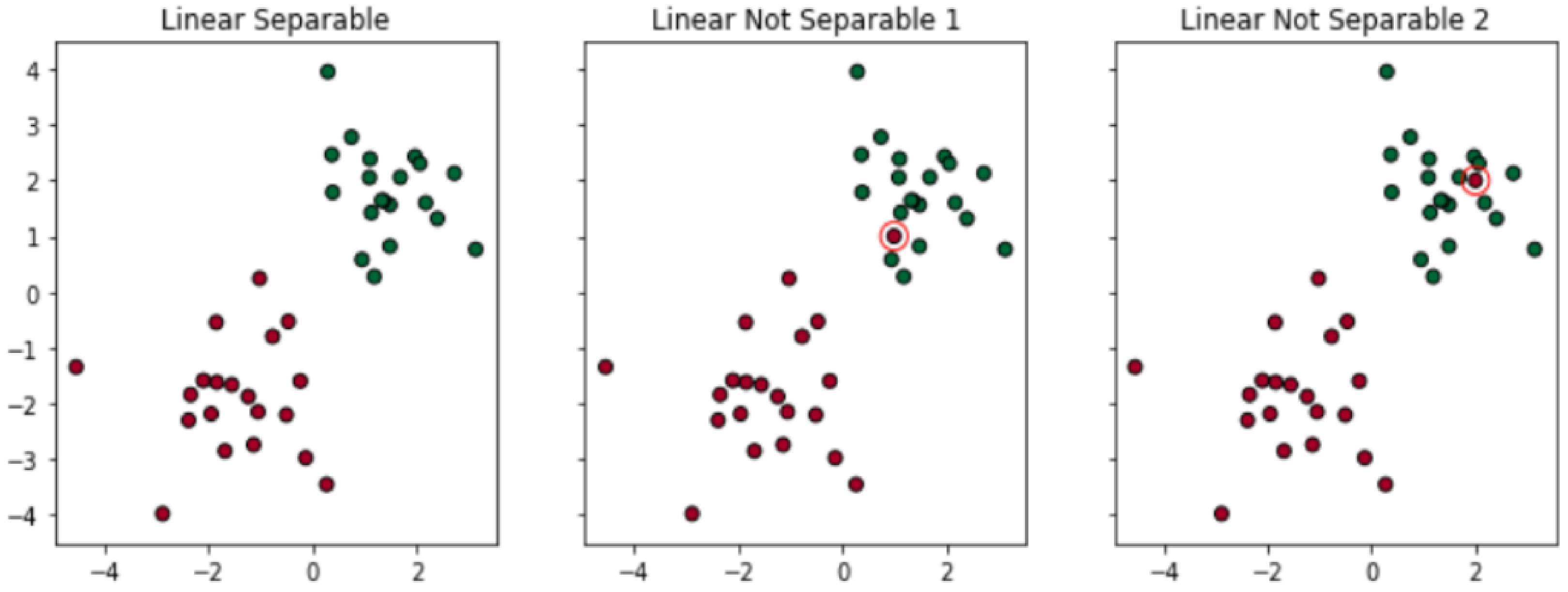
- According to the SVM algorithm we find the points closest to the line from both the classes.
- These points are called **support vectors.**
- Now, we compute the distance between the line and the support vectors. This distance is called the **margin.**
- Our goal is to **maximize the margin.** The hyperplane for which the margin is maximum is the **optimal hyperplane**.
- Thus SVM tries to make a decision boundary in such a way that the separation between the two classes is as wide as possible.
- If the points are linearly separable then only our hyperplane is able to distinguish between them and if any outlier is introduced then it is not able to separate them. So these type of SVM is called ***as hard margin SVM***



SVM IN NON-LINEARLY SEPARABLE CASES

- Two ways:
- **Soft Margin:** try to find a line to separate, but tolerate one or few misclassified dots (e.g. the dots circled in red)
- **Kernel Tricks:** try to find a non-linear decision boundary

SVM IN NON-LINEARLY SEPARABLE CASES

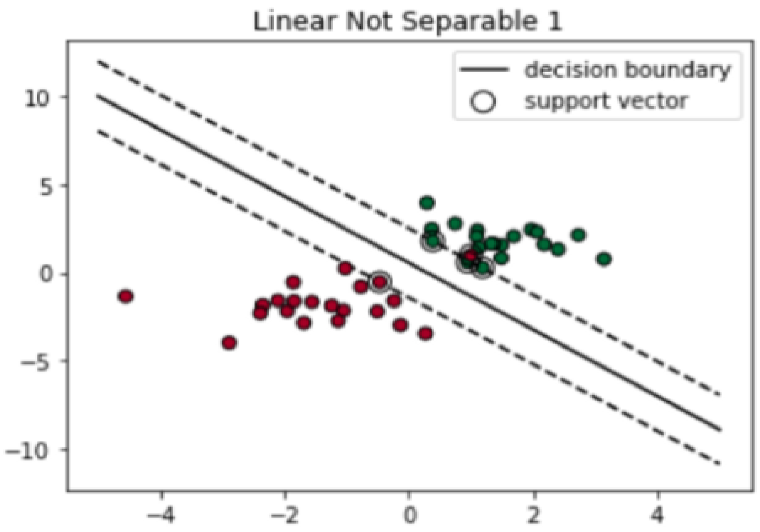


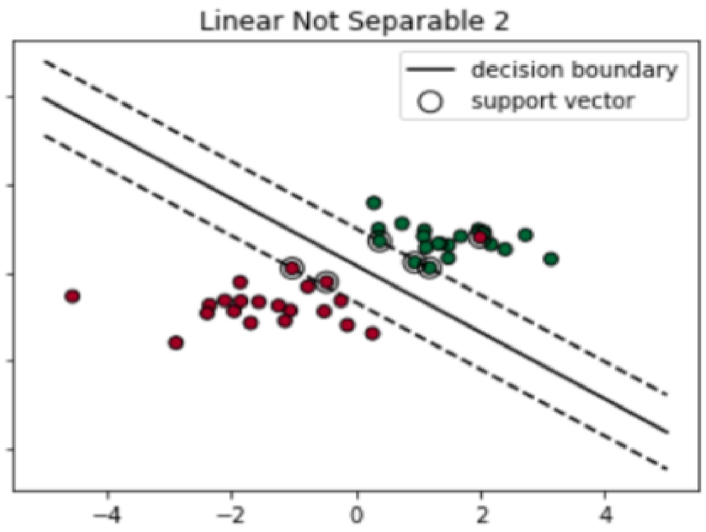
Soft Margin SVMs

Two types of misclassifications are tolerated by SVM under soft margin:

1.The dot is on the wrong side of the decision boundary but on the correct side/ on the margin (shown in left)

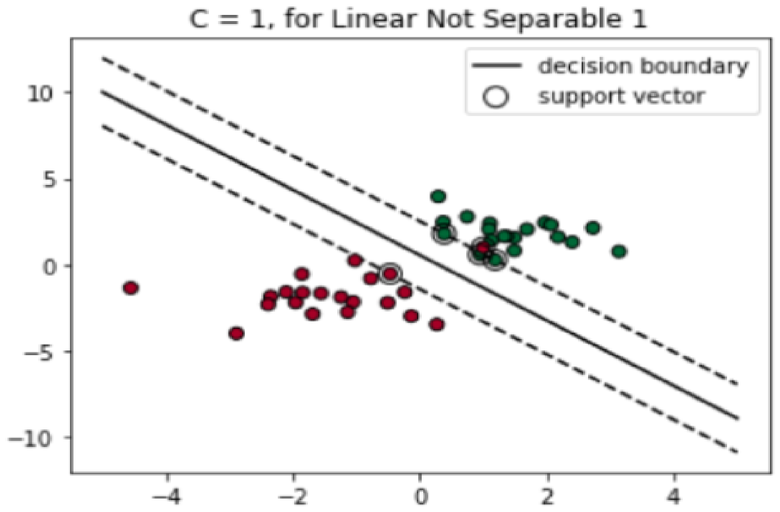
2.The dot is on the wrong side of the decision boundary and on the wrong side of the margin (shown in right)

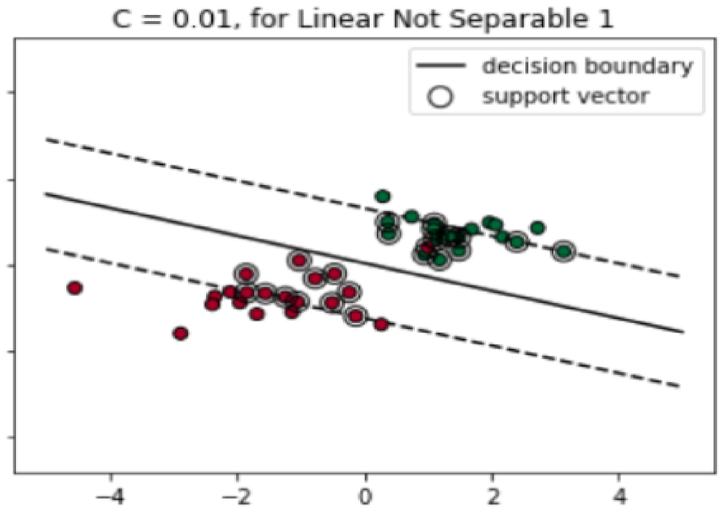




Soft Margin-Degree of Tolerance Term-C

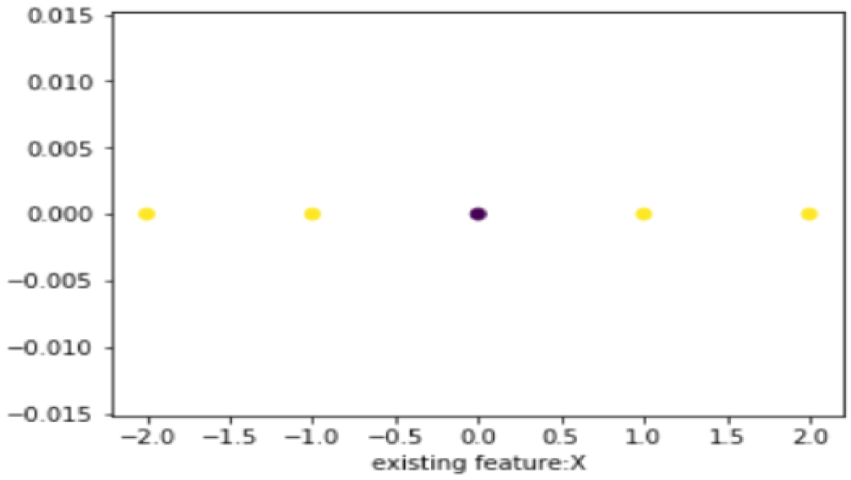
- Applying Soft Margin, SVM tolerates a few dots to get misclassified and tries to balance the trade-off between finding a line that maximizes the margin and minimizes the misclassification.
- **Degree of Tolerance:** How much tolerance(soft) we want to give when finding the decision boundary is an important hyper-parameter .
- The bigger the C, the more penalty SVM gets when it makes misclassification. Therefore, the narrower the margin is and fewer support vectors the decision boundary will depend on.

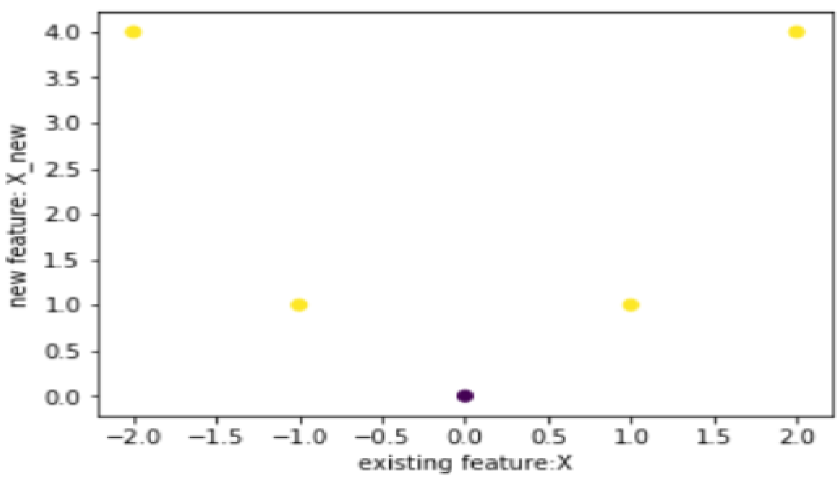




Kernel Trick

- Kernel Trick utilizes existing features, applies some transformations, and creates new features.
- **Polynomial kernel :** Think of the polynomial kernel as a transformer/processor to generate new features by applying the polynomial combination of all the existing features.
- If we apply transformation of degree 2, we get new features.
- Using Polynomial Kernel , SVM can generate non-linear decision boundary to separate the two sets of points.





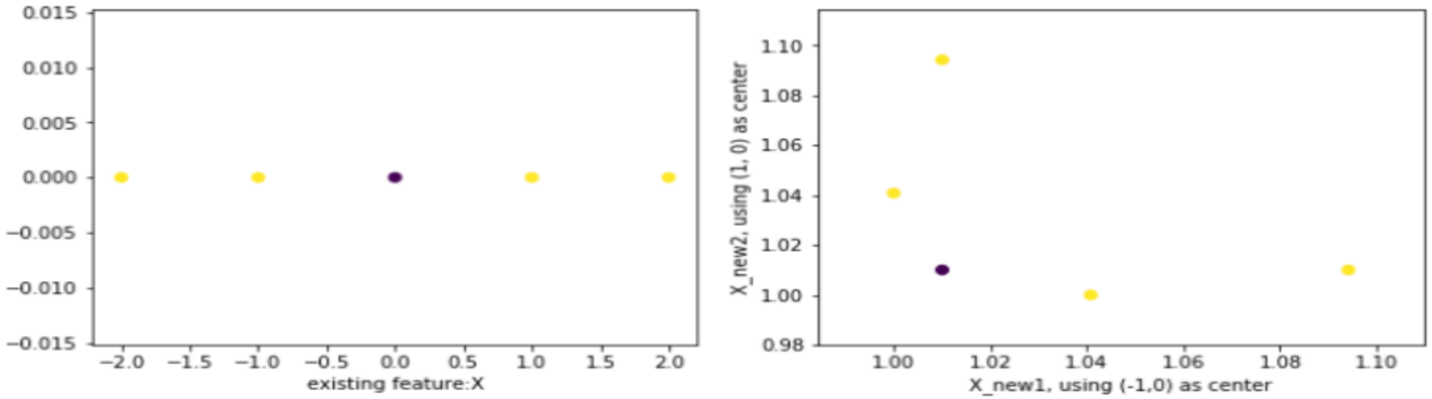
Kernel Trick

- **Radial Basis Function Kernel :** Think of the polynomial kernel as a transformer/processor to generate new features by applying the polynomial combination of all the existing features.
- **Gaussian RBF**:

$$K(X_1, X_2) = exponent(-\gamma||X_1 - X_2||^2)$$

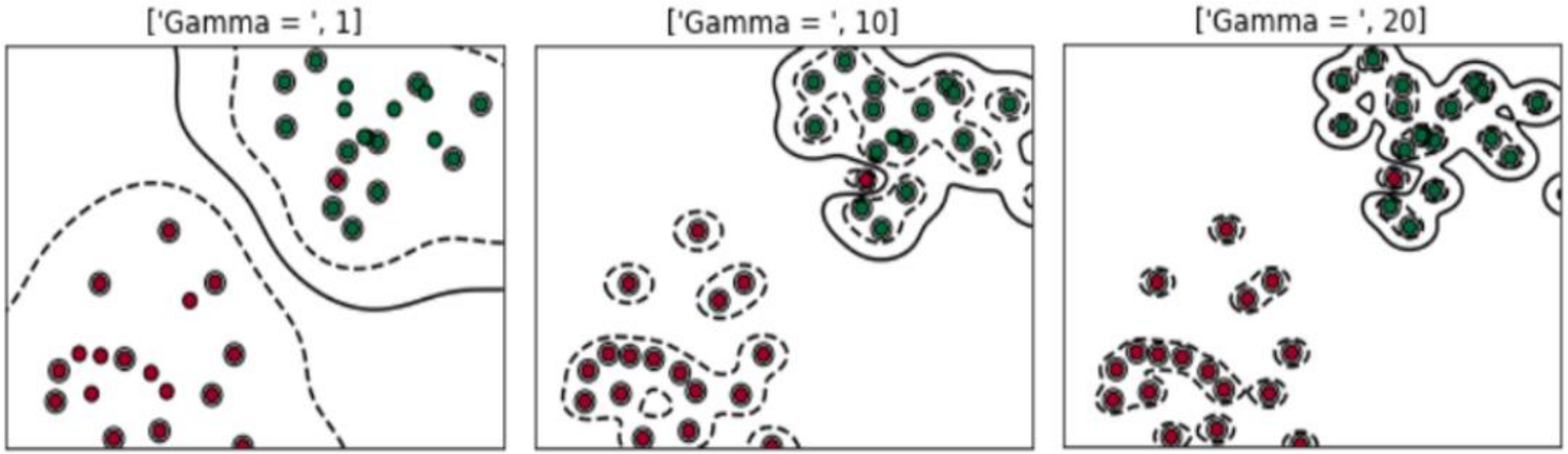
||X1 — X2 || = Euclidean distance between X1 & X2

- **gamma (γ)** controls the influence of new features on the decision boundary. The higher the gamma, the more influence of the features will have on the decision boundary, more wiggling the boundary will be.
- Applying Gaussian RBF with gamma=0.1 we get new features, which are non-linearly separable.



Kernel Trick-Gamma Hyperparameter

- Gamma is a hyperparameter that we can tune for when we use SVM with kernel.
- Gamma decides that how much curvature we want in a decision boundary.
- Gamma high means more curvature.
- Gamma low means less curvature.



Demo

https://cs.stanford.edu/~karpathy/svmjs/demo/