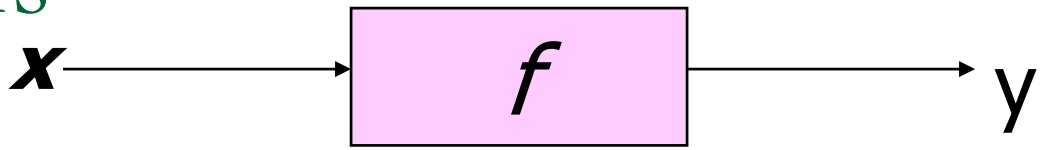
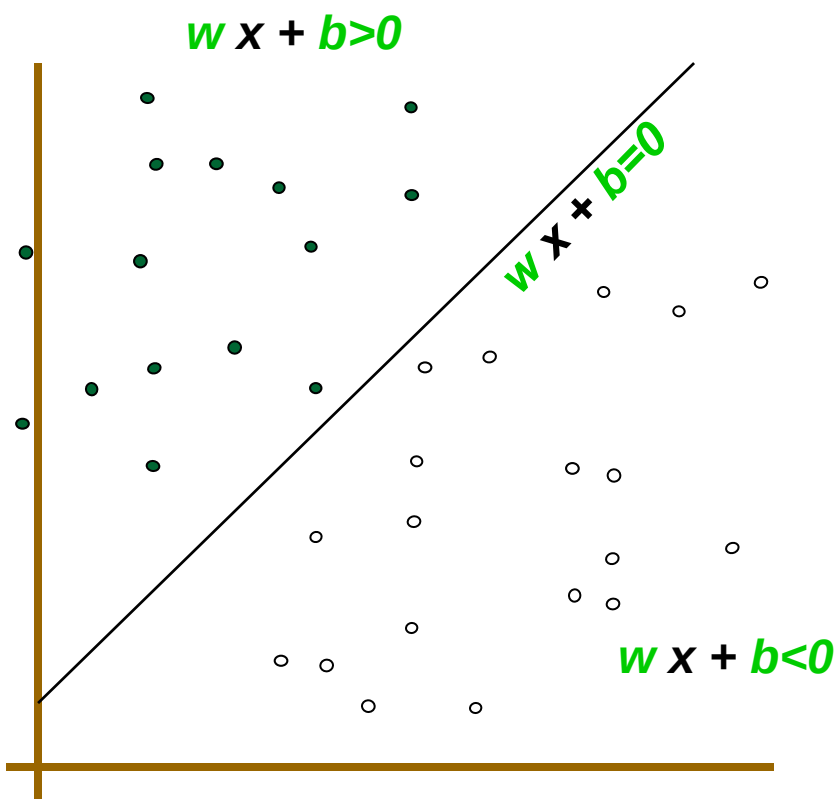
Support Vector Machine

Linear Classifiers

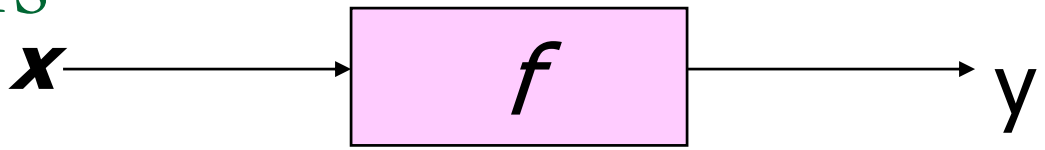


- denotes +1
- denotes -1

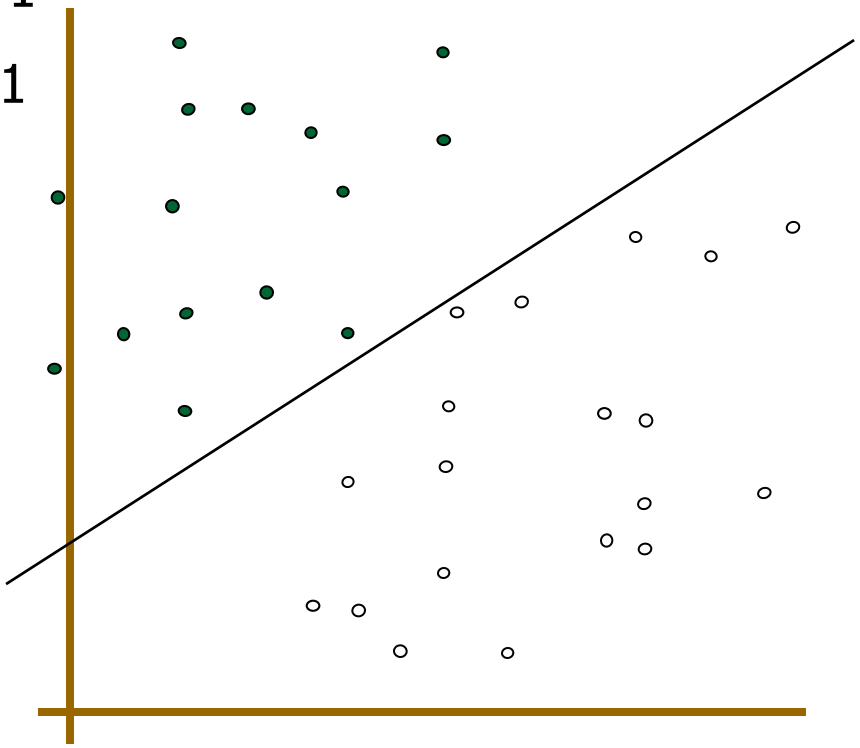


How would you classify this data?

Linear Classifiers



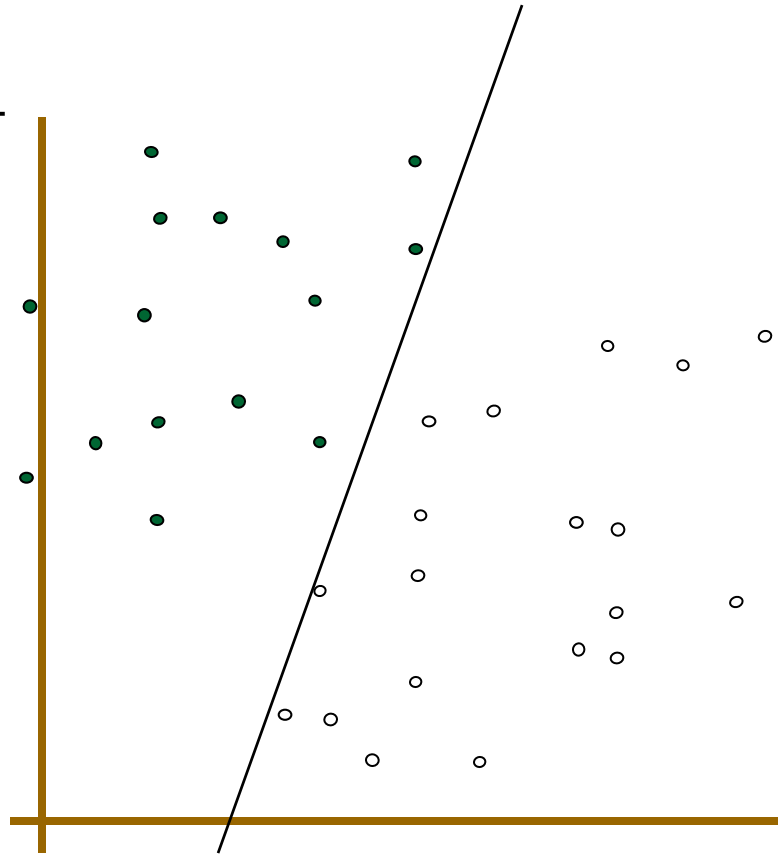
- denotes +1
- denotes -1



How would you classify this data?

Linear Classifiers

- denotes +1
- denotes -1

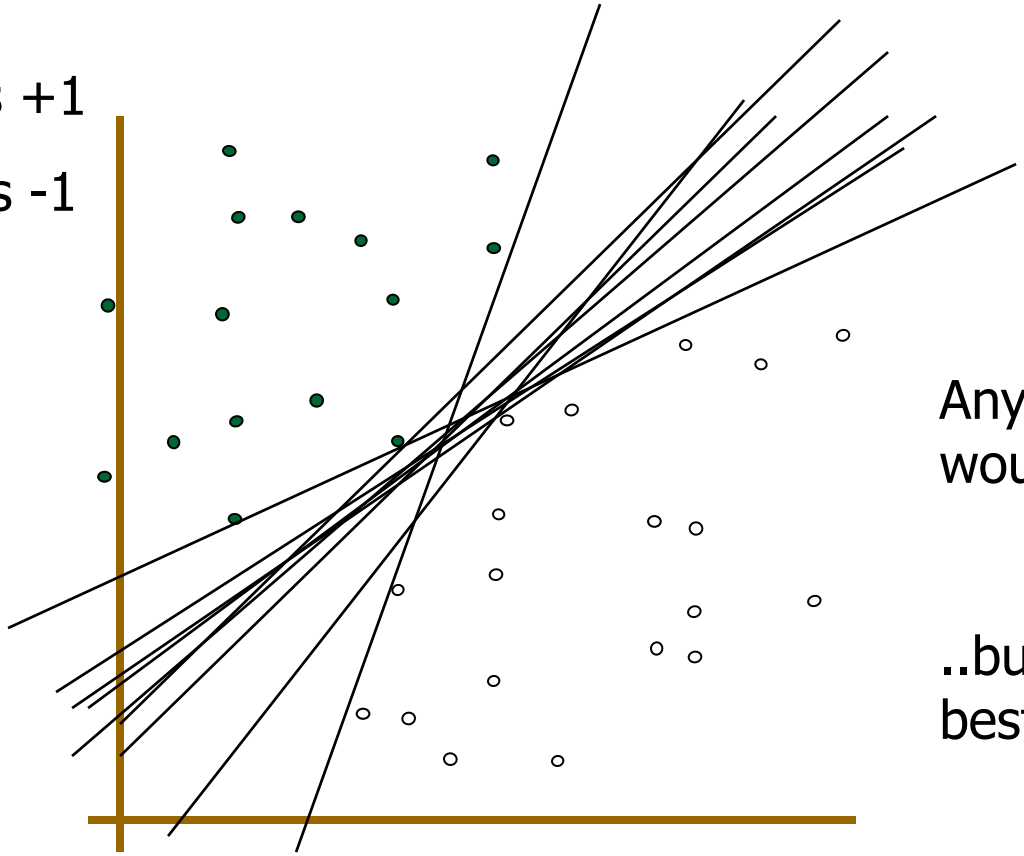


How would you classify this data?

Linear Classifiers

$$f(\mathbf{x},\mathbf{w},b) = sign(\mathbf{w}\,\mathbf{x} + b)$$

- denotes +1
- denotes -1



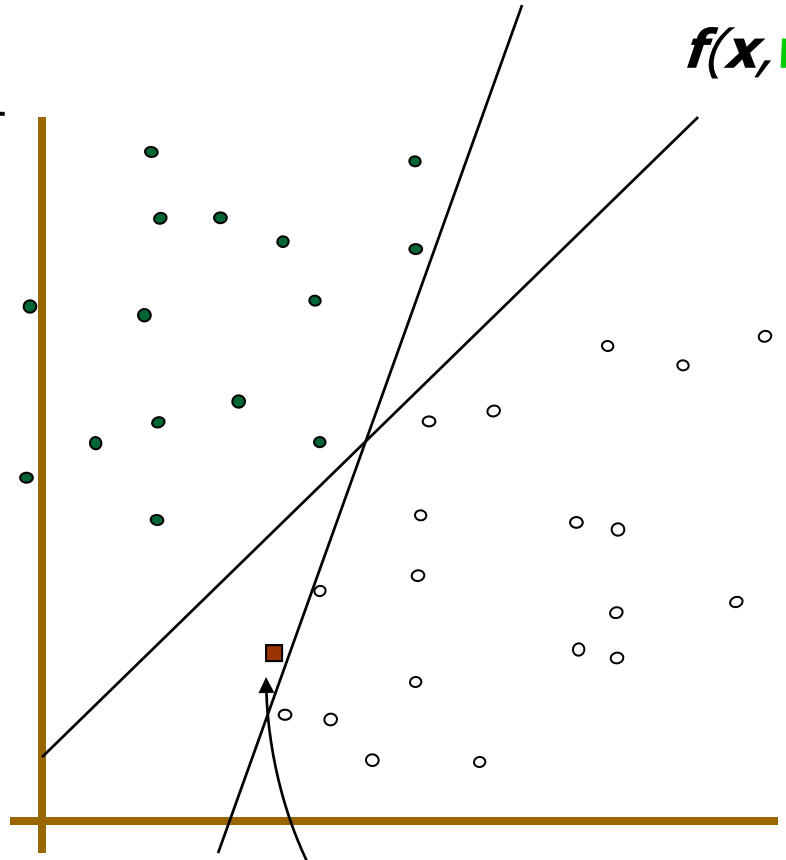
Any of these would be fine..

..but which is best?

Linear Classifiers

- denotes +1
- denotes -1

$f(\mathbf{x}, \mathbf{w}, b) = sign(\mathbf{w}\,\mathbf{x} + b)$

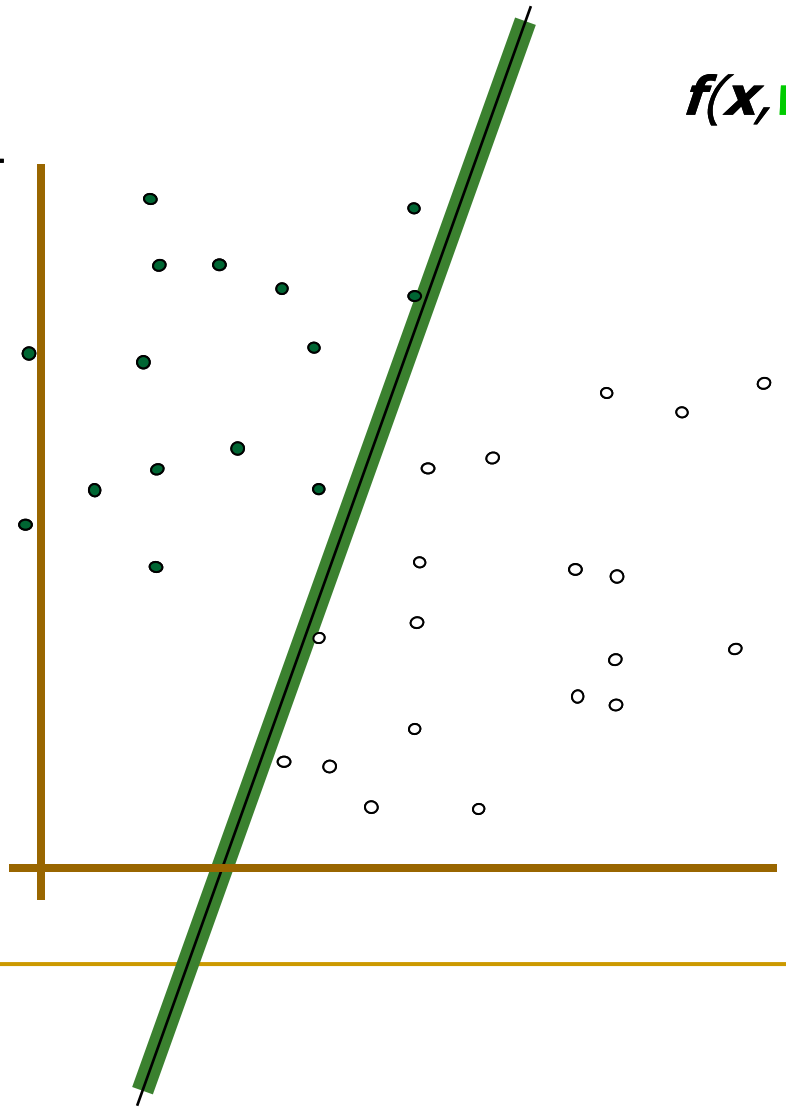


How would you classify this data?

Misclassified to +1 class

Classifier Margin

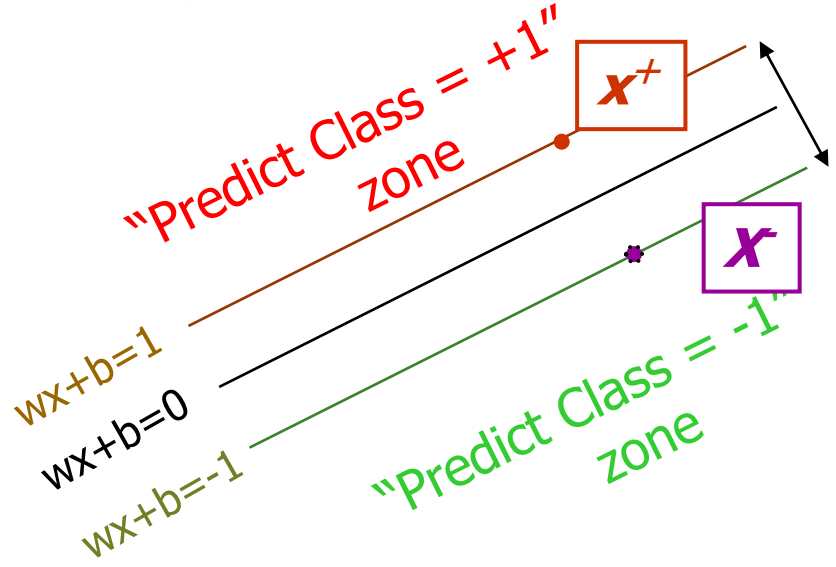
- denotes +1
- denotes -1



$f(\mathbf{x},\mathbf{w},b) = sign(\mathbf{w}\,\mathbf{x} + b)$

Define the margin of a linear classifier as the width that the boundary could be increased by before hitting a datapoint.

Linear SVM Mathematically



M=Margin Width

What we know:

- $\boldsymbol{w} \cdot \boldsymbol{x}^+ + b = +1$
- $\boldsymbol{w} \cdot \boldsymbol{x}^- + b = -1$
- $\boldsymbol{w} \cdot (\boldsymbol{x}^+ - \boldsymbol{x}^-) = 2$

$$M = \frac{(x^+ - x^-) \cdot w}{|w|} = \frac{2}{|w|}$$

Margin and Optimal HyperPlane

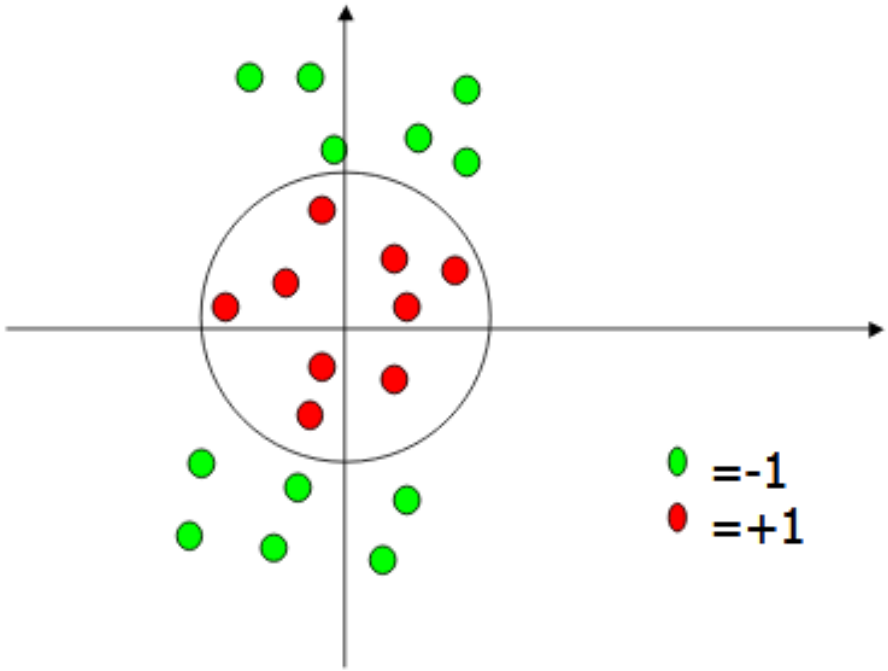
- Margin of Separation (d)

  The separation between the hyperplane and the closest data point for a given vector

- Optimal Hyperplane

  The particular hyperplane for which the margin of separation d is maximized

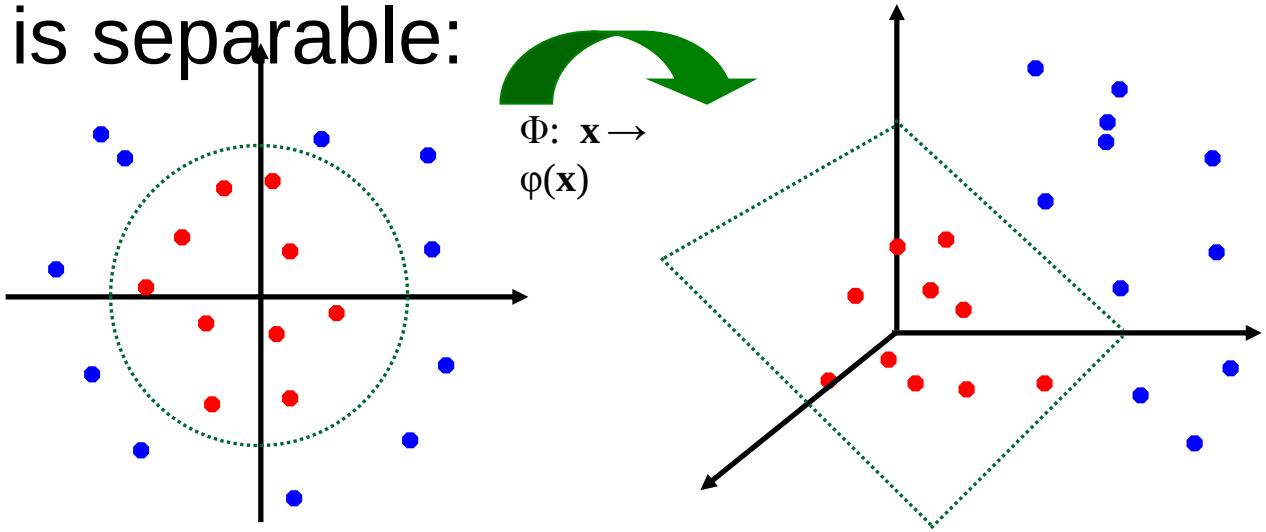
Problems with linear SVM



What if the decision function is not linear? What transform would separate these?

Kernels

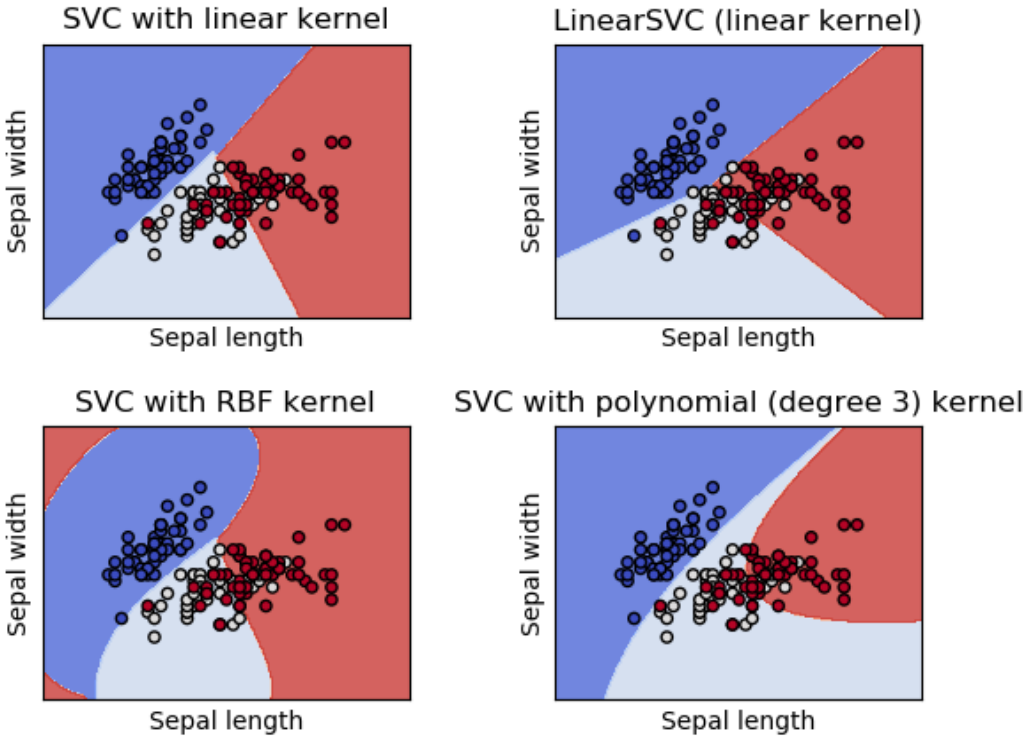
- General idea: the original feature space can always be mapped to some higher-dimensional feature space where the training set is separable:



Doing multi-class classification in SVM

- SVMs can only handle two-class outputs (i.e. a categorical output variable with arity 2).
- What can be done?
- Answer: with output arity N, learn N SVM's
  - SVM 1 learns "Output==1" vs "Output != 1"
  - SVM 2 learns "Output==2" vs "Output != 2"
  - :
  - SVM N learns "Output==N" vs "Output != N"
- Then to predict the output for a new input, just predict with each SVM and find out which one puts the prediction the furthest into the positive region.

Different Kernel Functions of IRIS dataset



What if data plot overlaps? Which line we draw

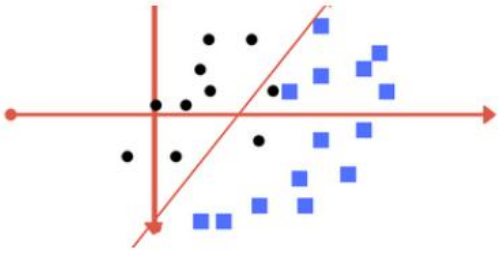


Image 1

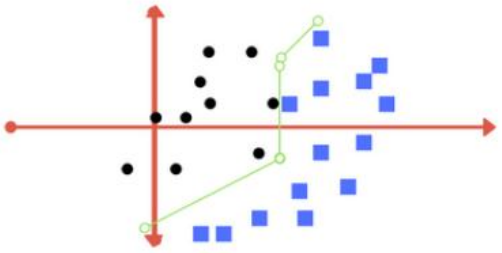


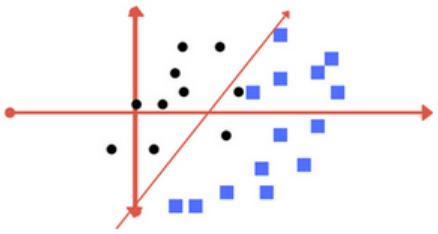
Image 2

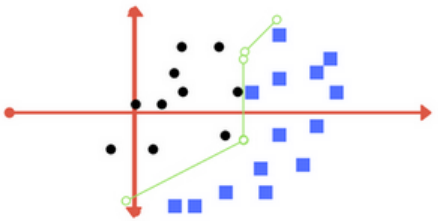
regularization parameter

- we define two terms **regularization parameter** and **gamma.** These are tuning parameters in SVM classifier. Varying those we can achieve considerable non linear classification

Regularization

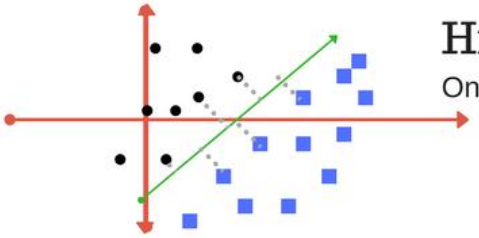
- The Regularization parameter (often termed as C parameter in python's sklearn library) tells the SVM optimization how much you want to avoid misclassifying each training example.





Gamma

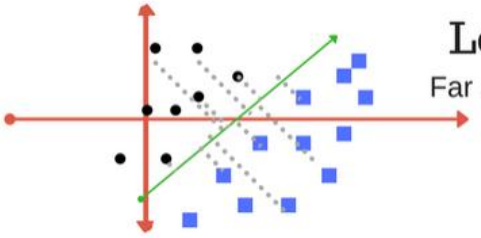
- The gamma parameter defines how far the influence of a single training example reaches
- With low values meaning ‘far’ and high values meaning ‘close’.
- With low gamma, points far away from separation line are considered in calculation for the separation line. Where as high gamma means the points close to the separation line are considered in calculation.



High Gamma

Only nearby points are considered.

High Gamma



Low Gamma

Far away points are also considered.