# Support Vector Machines

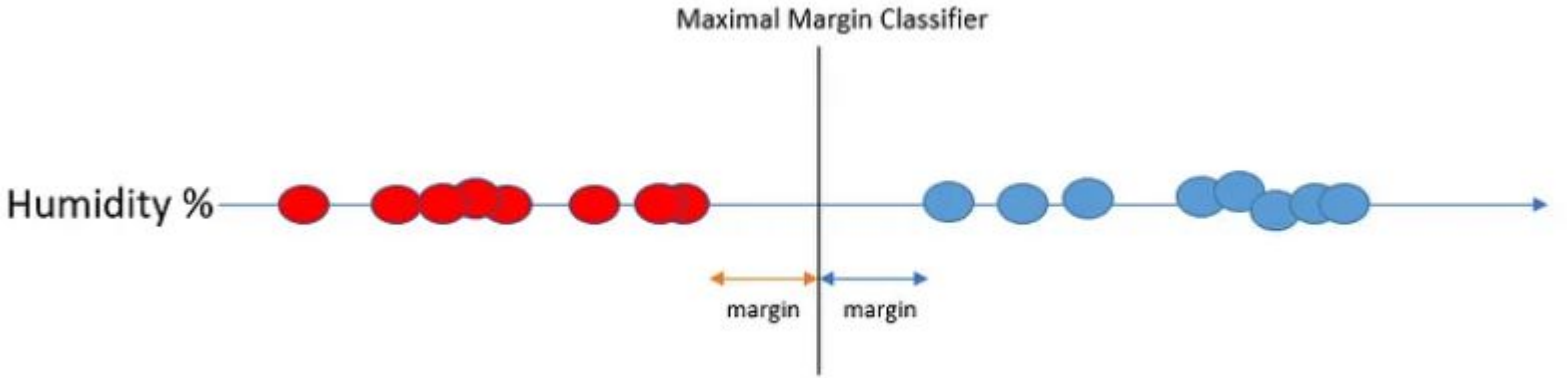
Agastya Thoppur

# Contents


- What are Margins?
- Hard Margins vs Soft Margins
- Soft Margin Classifier
- Limitations of Soft Margin Classifiers
- Support Vector Machines
- Kernel Functions
- Different Types of Kernel Functions

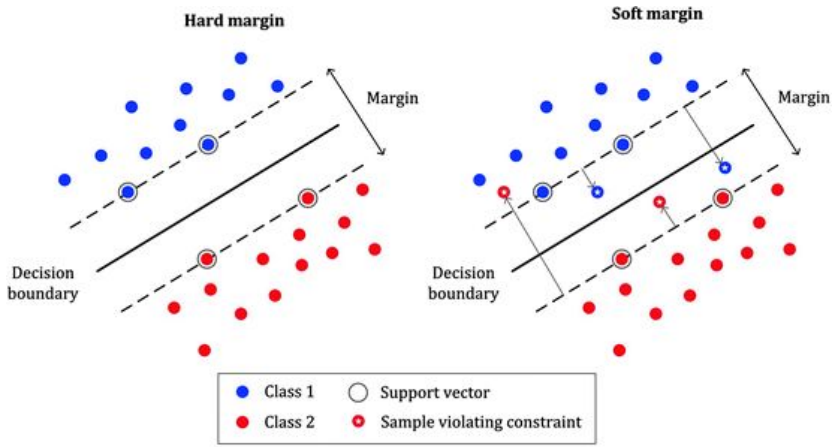
# What are Margins?

- A margin is the distance between a decision threshold and the data of a class.
- In this binary classification example, the decision threshold is exactly at the midpoint between the data of both classes. Hence, the margins are equal and maximum, making this a Maximal Margin Classifier.



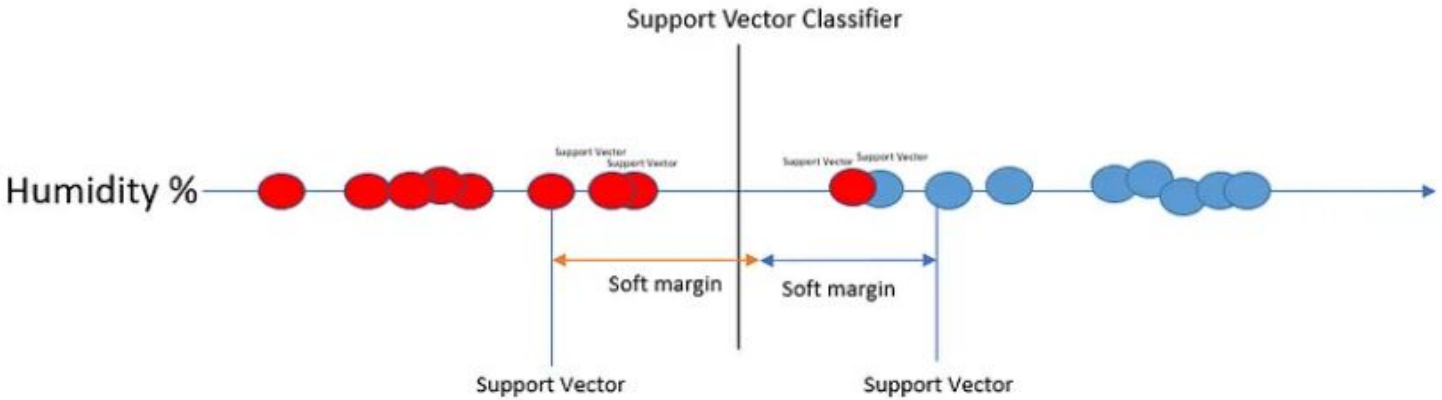
# Hard Margins vs Soft Margins

- Hard margins do not have any data points of the incorrect class within them. Since they do not allow misclassifications, they are extremely sensitive to outliers.
- Soft margins allow data points of the incorrect class to lie within it. Setting a decision threshold with soft margins make the classifier less sensitive to outliers by allowing misclassifications. (Higher bias, lower variance)



# Soft Margin Classifier

- A Soft Margin Classifier, or a Support Vector Classifier uses soft margins to determine the best decision threshold.



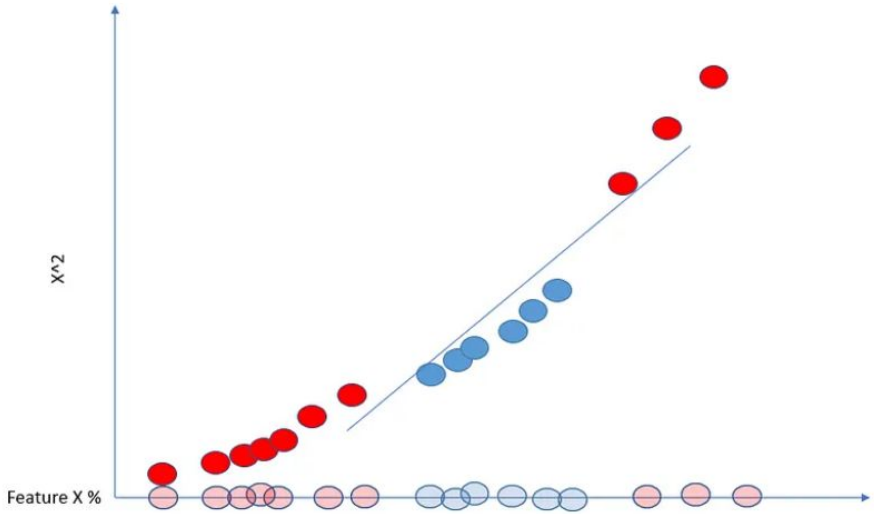
# Limitations of Soft Margin Classifiers

- The problem with Soft Margin Classifiers is that they don’t perform well on data with a lot of overlap.

- Support Vector Machines “transform” data to find accurate Support Vector Classifiers.

# Support Vector Machines

- Support vector machines move low-dimensional data into a higher dimension and then finds a soft margin classifier.
- SVMs use kernel functions to find support vector classifiers in higher dimensions.



# Kernel functions

- Kernel functions transform data columns so that Soft Margin Classifiers can find an accurate decision threshold.
- Transforming each and every data point in a large dataset is computationally expensive, especially if the kernel function conducts a complex transformation.
- Kernel functions use **The Kernel Trick** to allow SVMs to operate in higher dimensions without actually calculating the transformed data points.

# Different types of Kernel functions

- Linear Kernels are used when the data is linearly separable (can be separated with a single line without having to transform the data).
- Polynomial Kernels are used when the data is not linearly separable, i.e the data points of both classes overlap a lot.
- RBF Kernels finds soft margin classifiers in infinite dimensions. It behaves like a weighted Nearest Neighbors Model, since it assigns higher weights to points closest to the test point and lower weights to points farther from it.

- linear $k(\mathbf{x}_1, \mathbf{x}_2) = \mathbf{x}_1 \cdot \mathbf{x}_2$
- polynomial $k(\mathbf{x}_1, \mathbf{x}_2) = (\gamma\, \mathbf{x}_1 \cdot \mathbf{x}_2 + c)^d$
- Gaussian or radial basis $k(\mathbf{x}_1, \mathbf{x}_2) = \exp\left(-\gamma \left\|\mathbf{x}_1 - \mathbf{x}_2\right\|^2\right)$

# Thank you!