Support Vector Machines

NUMERICAL ANALYSIS FOR MACHINE LEARNING

MIGUEL PLANAS DÍAZ (ERASMUS +)

PERSON CODE: 11071870

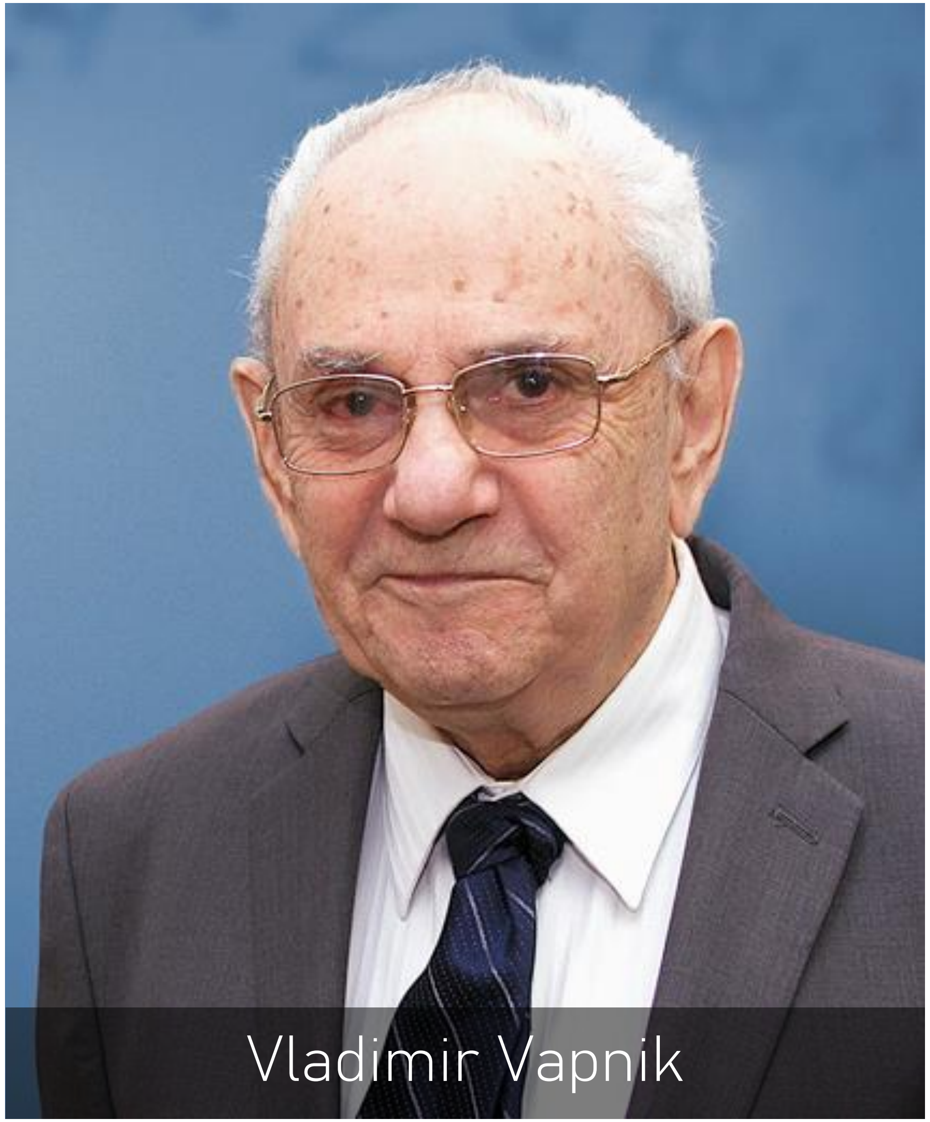
2024-2025

Support Vector Machines

- Introduction to Support Vector Machines
- Functionality of SVM
- Classification with SVM
- SVM for regression
- Real-world SVM applications

The origins of Support Vector Machines

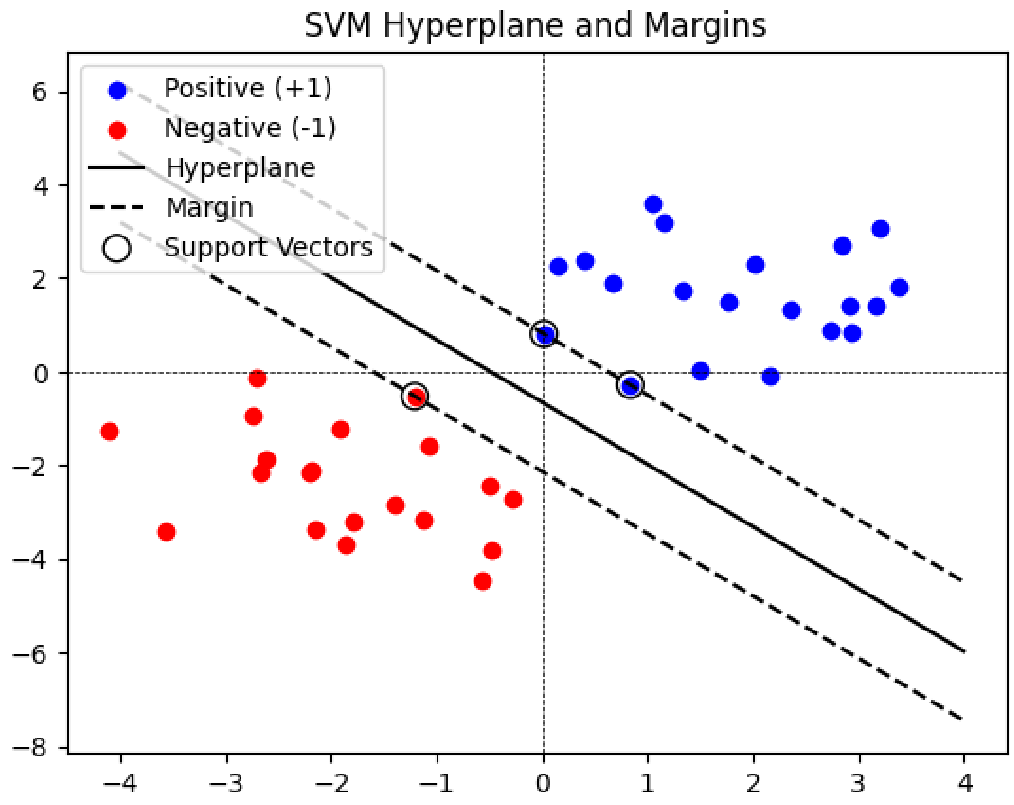
- Introduced by Boser, Guyon and Vapnik in 1992.
- Rooted in Statistical Learning Theory, developed by Vladimir Vapnik and Alexey Chervonenkis.
- Emerged as a robust classifier for handling high-dimensional data managing the balance between complexity and generalization.



Vladimir Vapnik

Functionality of SVM

- Supervised learning technique primarly used for classification tasks.
- Fundamental goal: Determine a hyperplane that effectively separates observations into two classes, by maximizing the margin between them.



Hyperplane definition

A hyperplane is a linear decision boundary that separates the space into two distinct half-spaces. It is mathematically defined as:

$$w^t x + b \ = 0$$

- Points lying on the hyperplane satisfy:

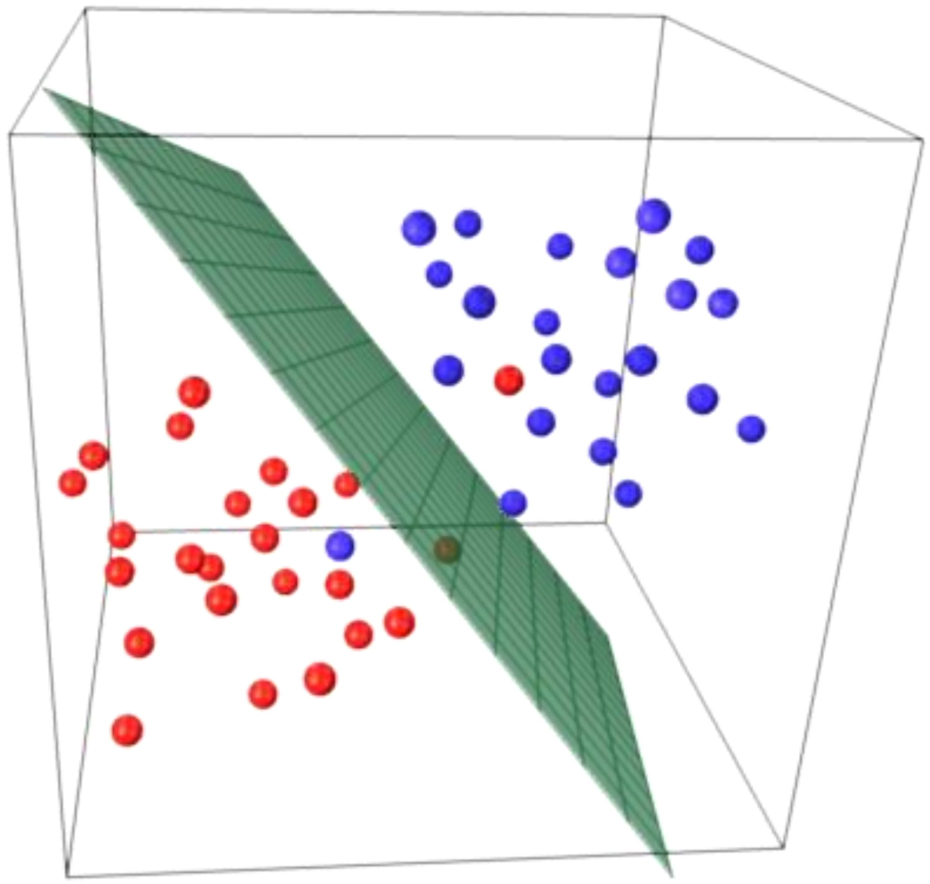
$$w^t x_i + b \ = 0$$

- Points on the positive side of the hyperplane fulfill:

$$w^t x_i + b > 0$$

- Points on the negative side of the hyperplane meet:

$$w^t x_i + b < 0$$



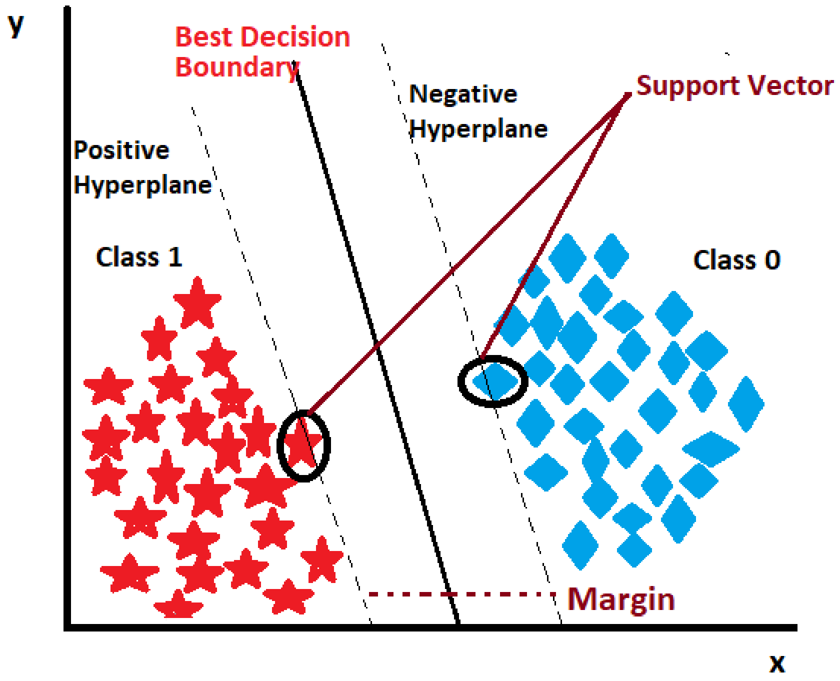
Support Vectors

The support vectors are the data points that lie closest to the hyperplane. They are critical because:

- The position of the hyperplane is entirely determined by the support vectors. Changing or removing any other data point has no effect on the decision boundary.
- These points define the margins, which are the boundaries:

$$w^t x + b = \pm 1$$

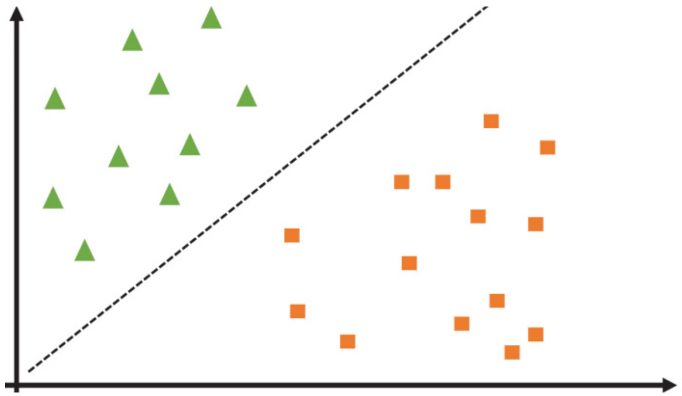
- They help SVM to achieve a balance between model complexity and generalization, as the margin maximization reduces overfitting.

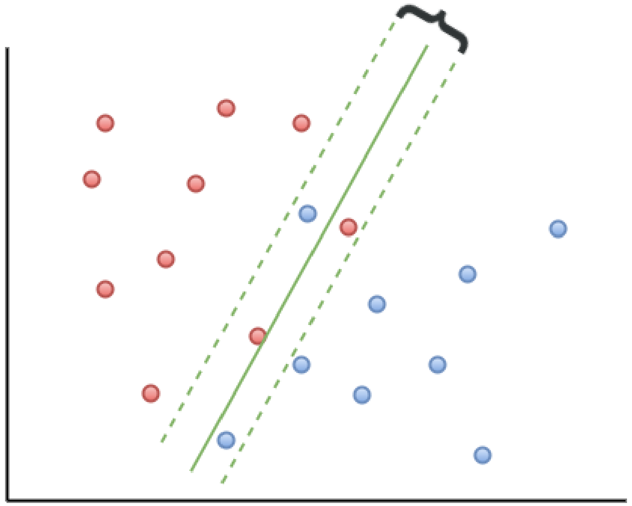


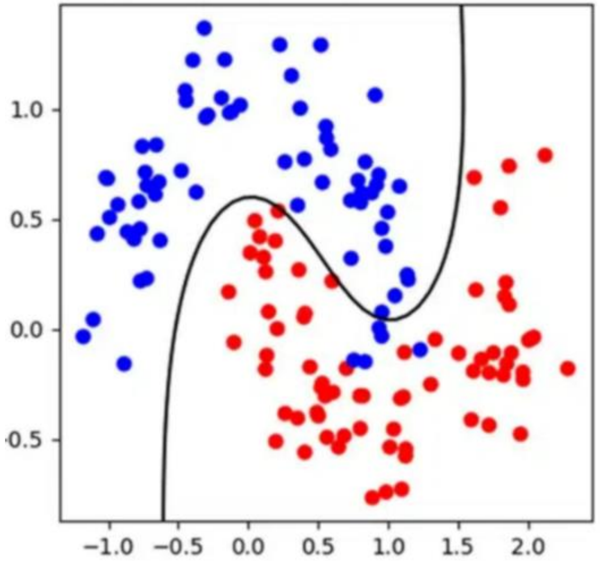
SVM for classification

Types of SVM classification:

- Linearly separable problems.
- Non-linearly separable problems.
- Non-linear SVMs.



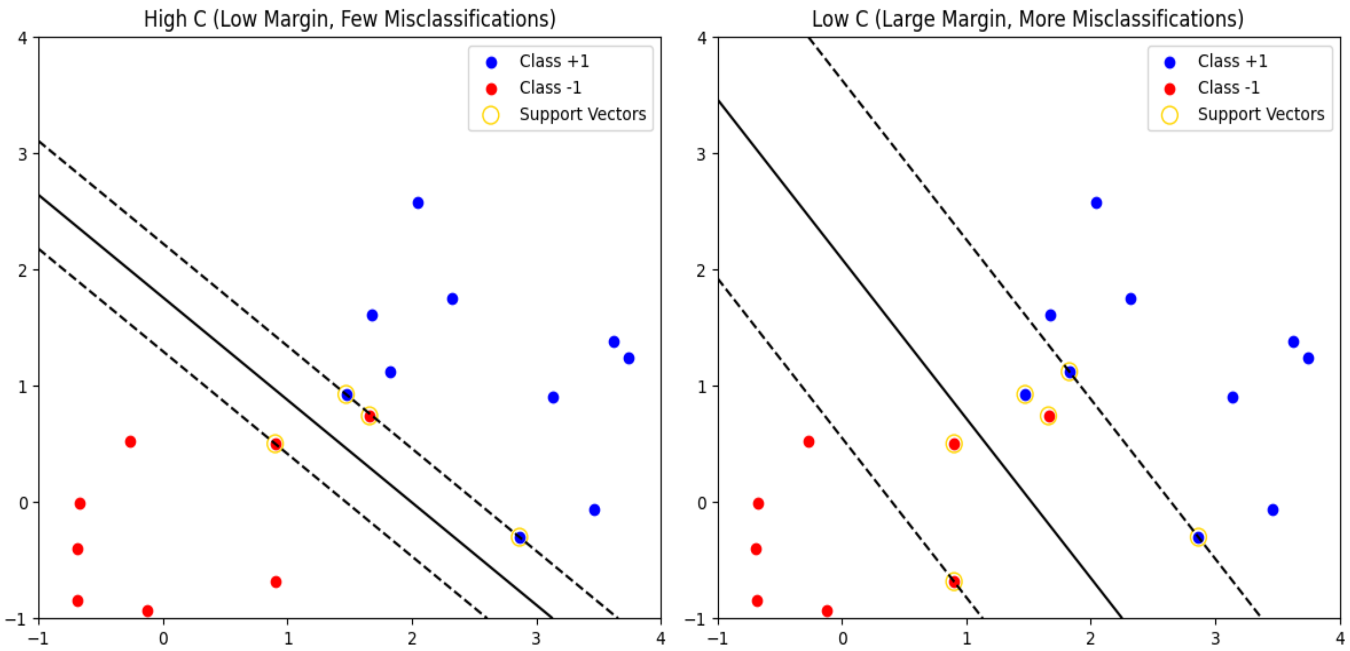




Non-Linearly Separable Problems

In real-world applications, data is often not perfectly linearly separable. To address this, SVMs introduce the concept of soft margins, which allow for some degree of misclassification. This is made by adding C, a regularization parameter that controls the trade-off between maximizing the margin and minimizing classification errors.

- A large C places a higher penalty on misclassification, resulting in a smaller margin and fewer errors, but it may lead to overfitting.
- A small C allows for a larger margin and tolerates more misclassifications, improving generalization but at the cost of training accuracy.



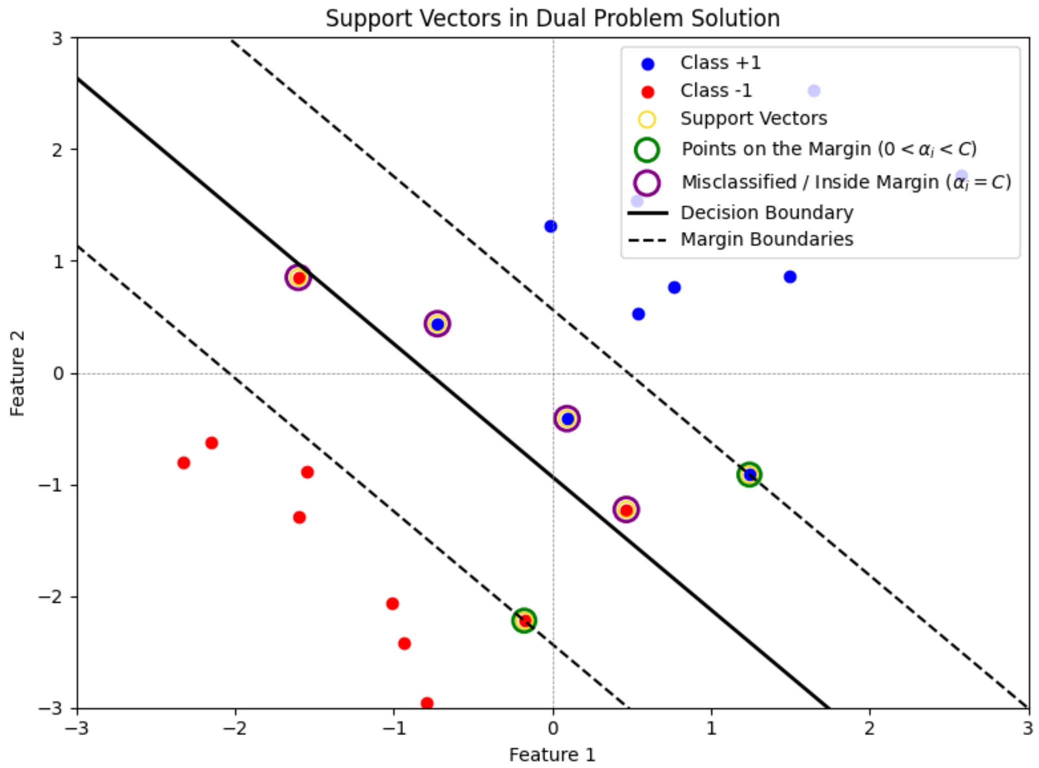
Types of Support Vectors

In the solution of the dual problem,

$$L_p = \frac{1}{2}\|w\|^2 - C\sum_{i=1}^{I}\xi_i - \sum_{i=1}^{I}\alpha_i\left[y_i(w^T x_i + b) - 1 + \xi_i\right] - \sum_{i=1}^{I}\beta_i\xi_i,$$

the Lagrange multipliers α_i help identify the support vectors, which are critical points for defining the decision boundary:

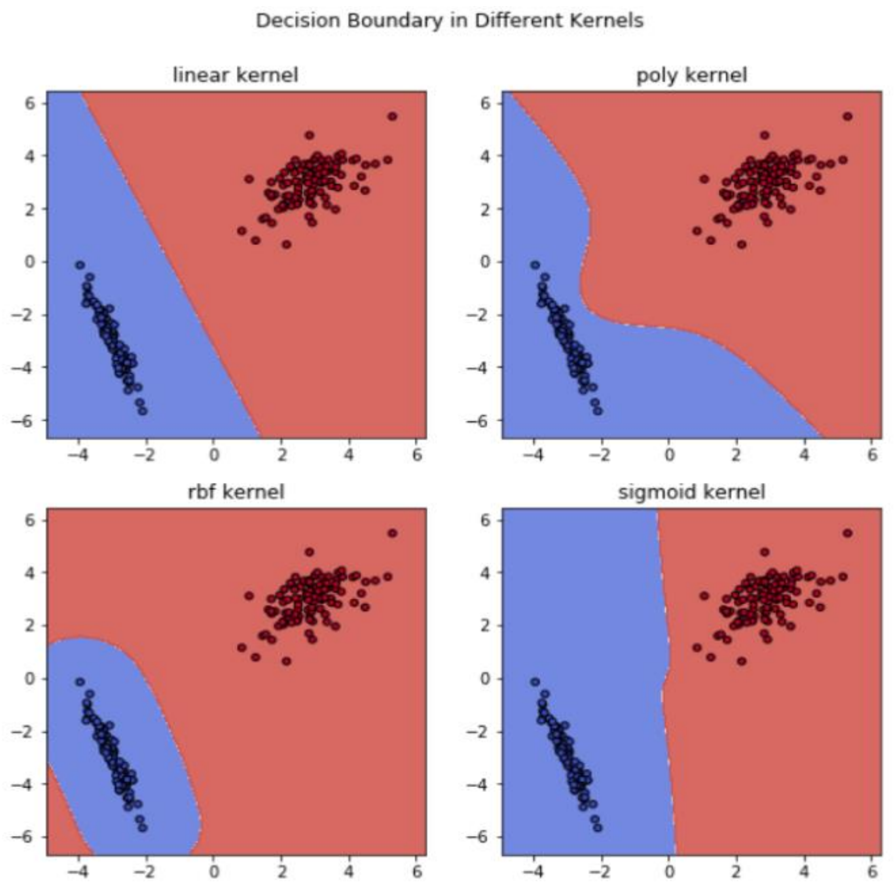
- If $0 < \alpha_i < C$: The corresponding data point lies exactly on the margin, satisfying $y_i(w^t x_i + b) = 1$ with $\xi_i = 0$.
- If $\alpha_i = C$: The corresponding data point lies inside the margin or is misclassified. In this case, $\xi_i > 0$, and the condition $y_i(w^t x_i + b) = 1 - \xi_i$ holds.



Non-Linear SVM Techniques

The key idea is to map the data into a higher-dimensional space where a linear hyperplane can separate the classes effectively using kernel methods.

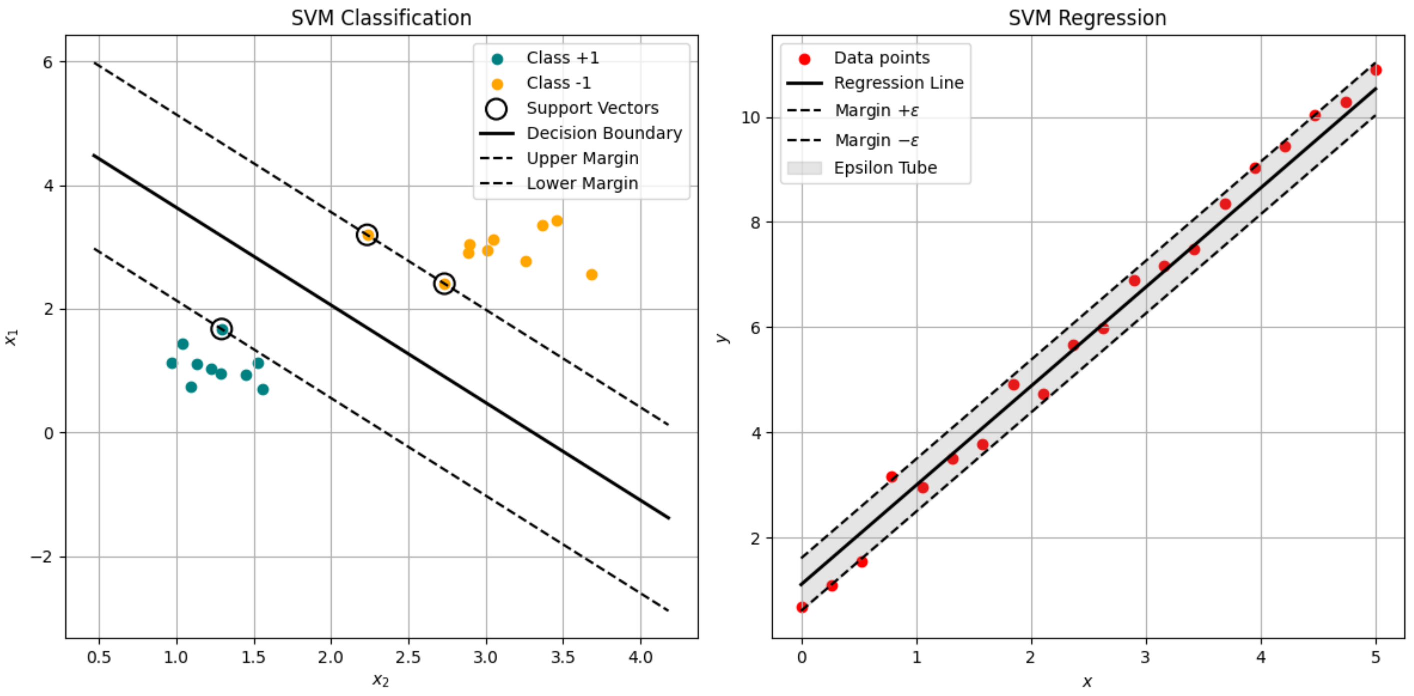
- The most common kernel functions are Linear, Polynomial, Radial Basis function (RBF) and Sigmoid.
- The main challenges of these techniques are the choice of the kernel, the computational cost and the overfitting risk.



SVM for Regression

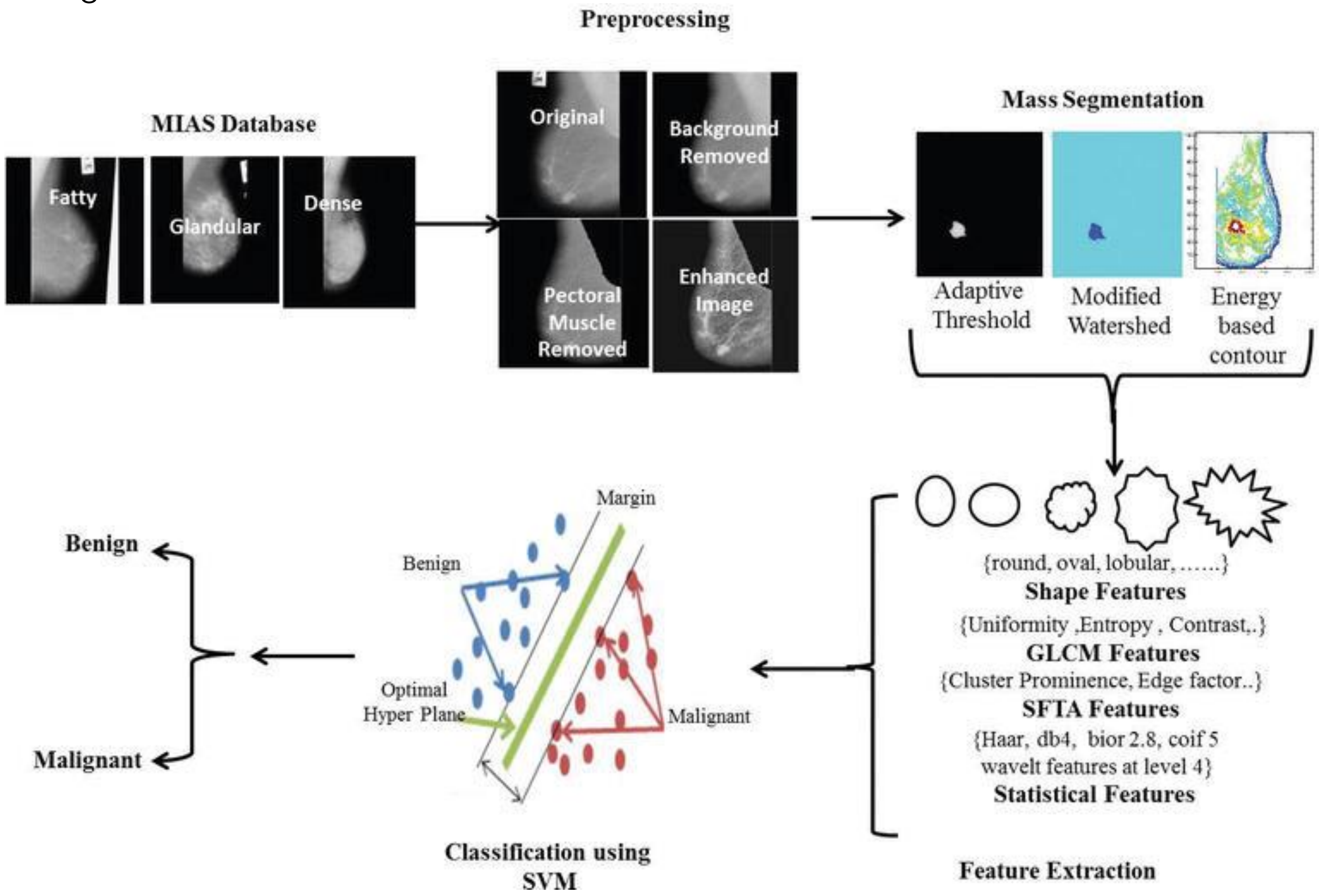
In contrast to classification, regression aims to predict a continuous output value based on input features. In SVMs, the distinction manifests in how data is handled relative to the margin:

- Classification: The data points are placed outside the margin, with the goal of minimizing classification errors and maximizing the separation margin between classes.
- Regression: The data points are allowed to lie within the margin, and the objective is to minimize the deviation from a predicted value, often constrained by a tolerance parameter ε.



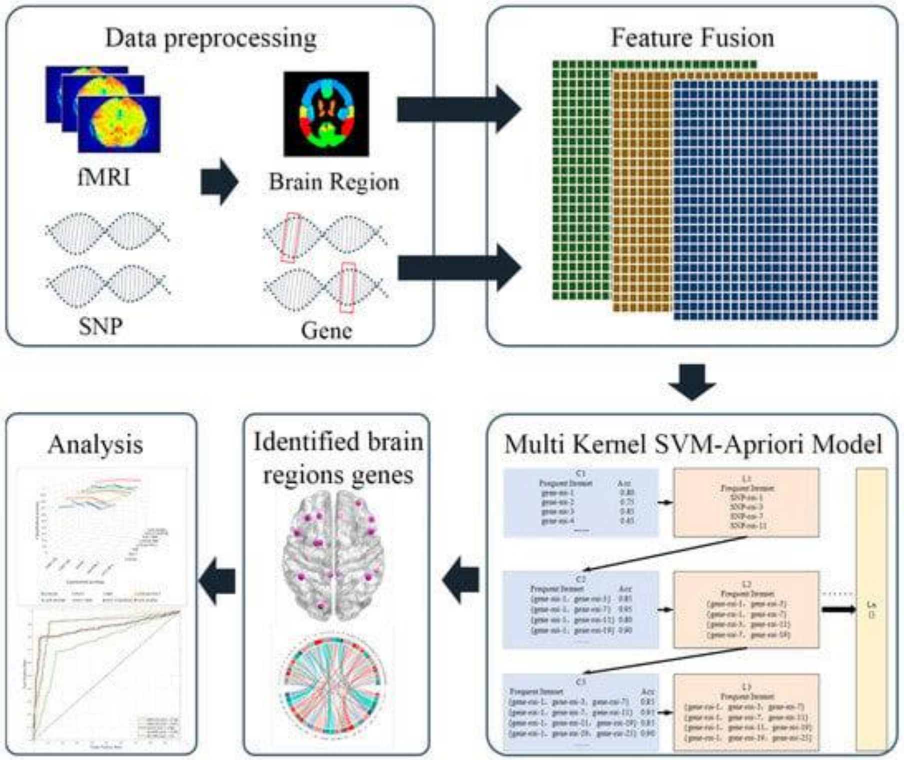
Real World Applications:

Image Classification:



Real world applications

Genetic data analysis:



Finance: