Group9

SOFT MARGIN SVM

AND REGULARIZATION

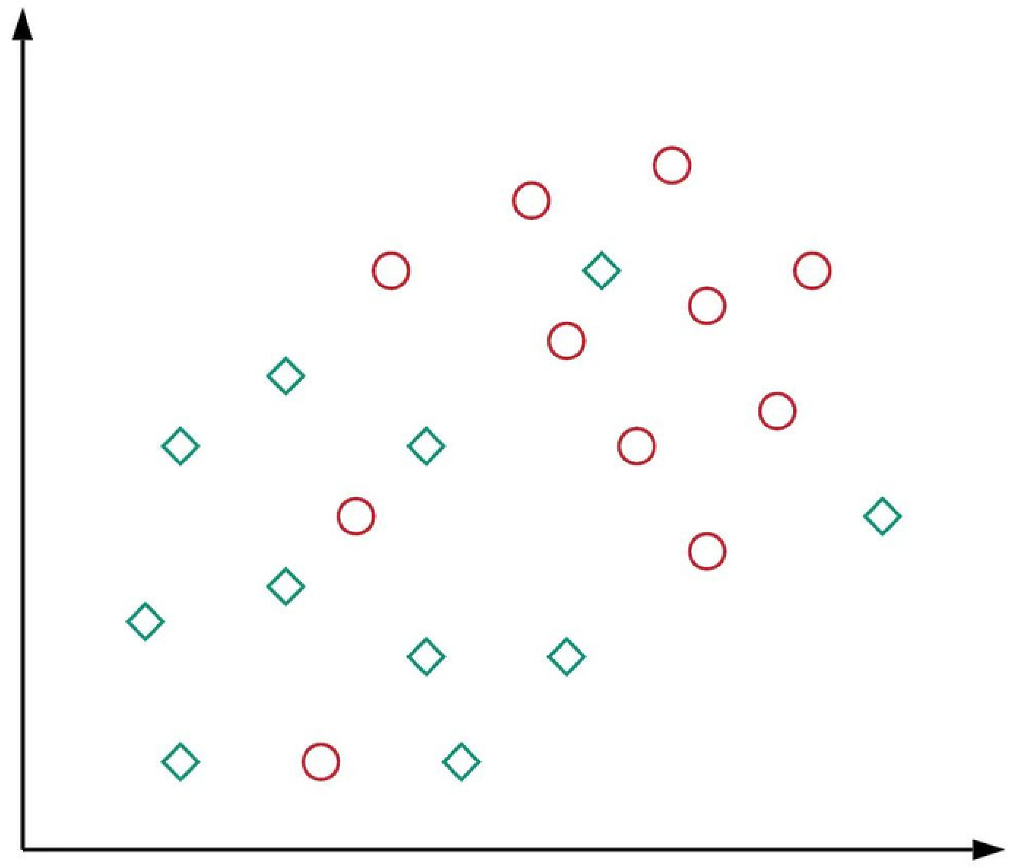
OUR TEAM

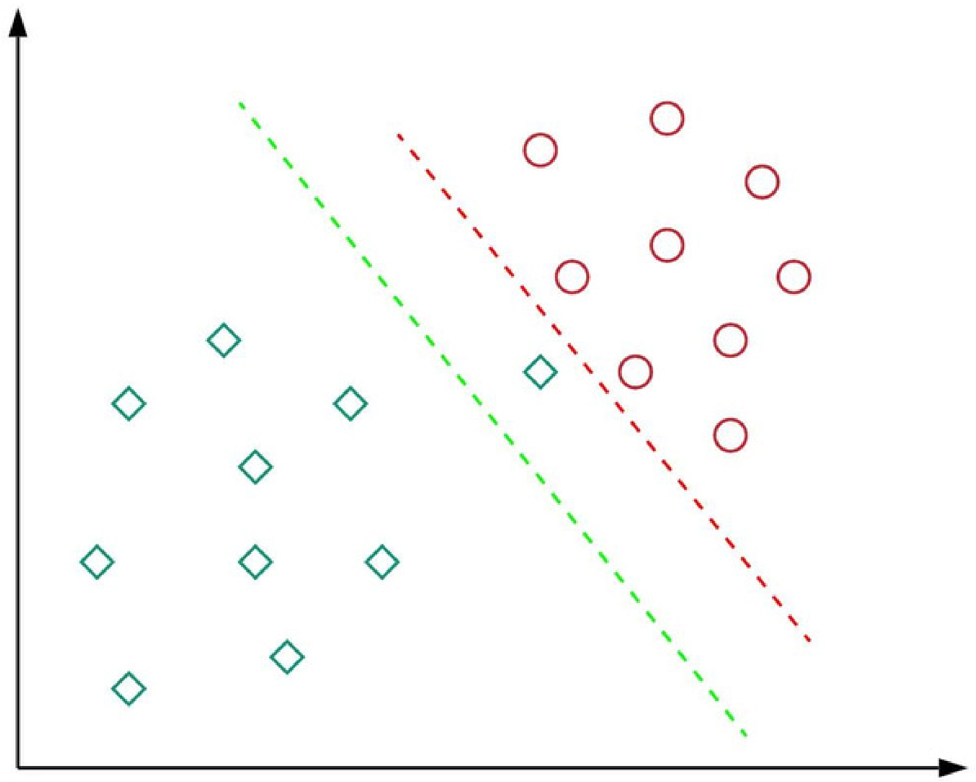
Ek Vong Panharith	e20200877
Run Savin	e20200897
Vann Visal	e20200537
CHHON CHAINA	e20200934
Khon Yin Sakal	e20200425
Pean Chhinger	e20201339

CONTENT

1. Introduction to soft margin SVM and its role in handling non-linearly separable data

Soft margin SVM is a type of support vector machine (SVM) that allows for some misclassifications. Soft margin SVM works by minimizing a cost function that includes both the margin and the number of misclassifications. The margin is a measure of how well the data is separated by the decision boundary. The number of misclassifications is a measure of how well the model predicts the labels of the data.





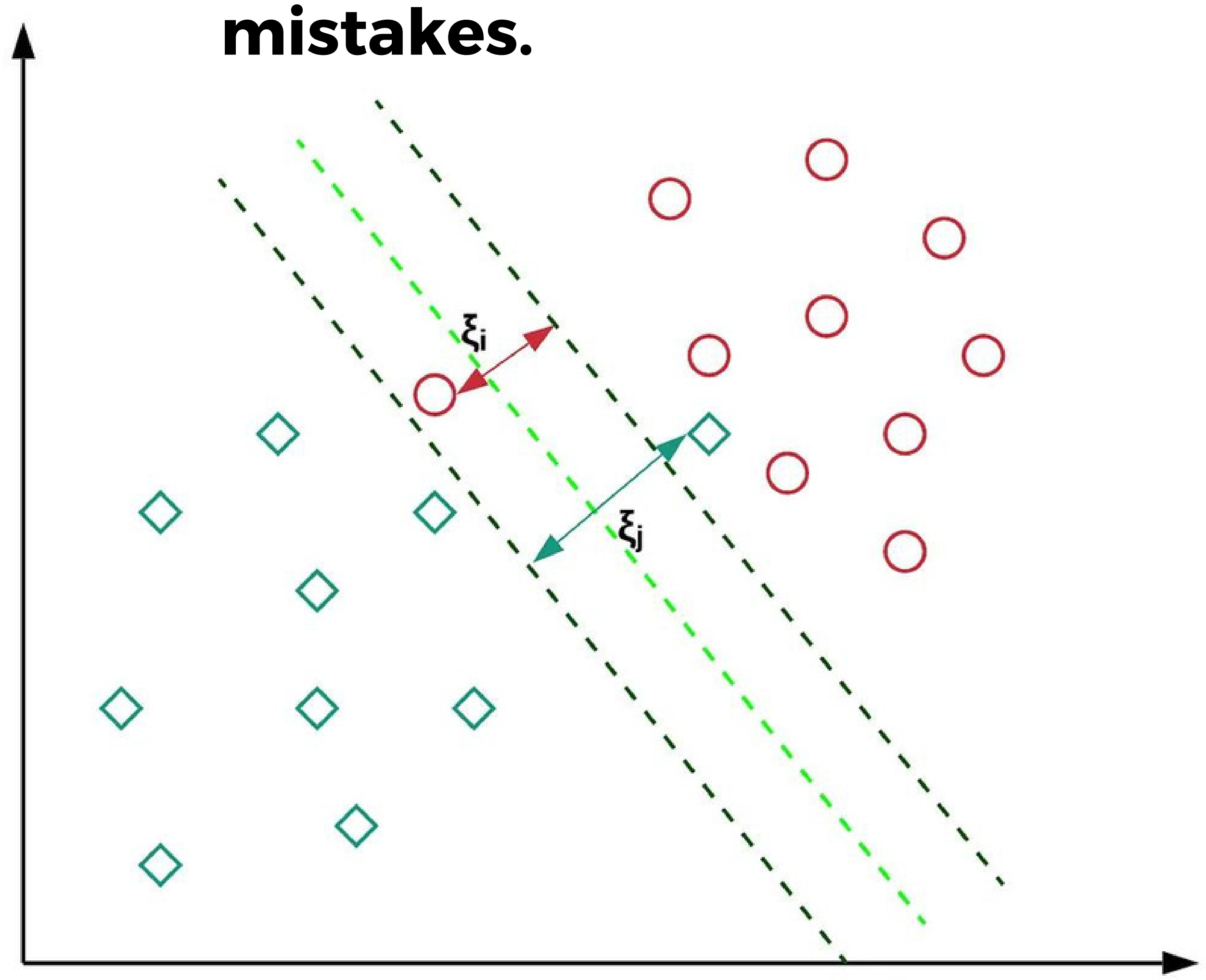
We would aim to minimize the following objective:

$$L = \frac{1}{2}\|w\|^2 + C(\#\ of\ mistakes)$$

Here, C is a hyperparameter that decides the trade-off between maximizing the margin and minimizing the mistakes.

Confidence score

$$y_i(\vec{w}.\vec{x}_i + b) \geq 1 - \xi_i$$



Our objective is to minimize the following function:

$$L = \frac{1}{2}\|\vec{w}\|^2 + C\sum_i \xi_i + \sum_i \lambda_i(y_i(\vec{w}\ .\ \vec{x}_i + b) - 1 + \xi_i)$$

- w is the weight vector
- b is the bias term
- C is a hyperparameter that controls the trade-off between the margin and the number of misclassifications
- λ is a hyperparameter that controls the regularization of the model
- y_i is the label of the ith data point
- x_i is the features of the ith data point

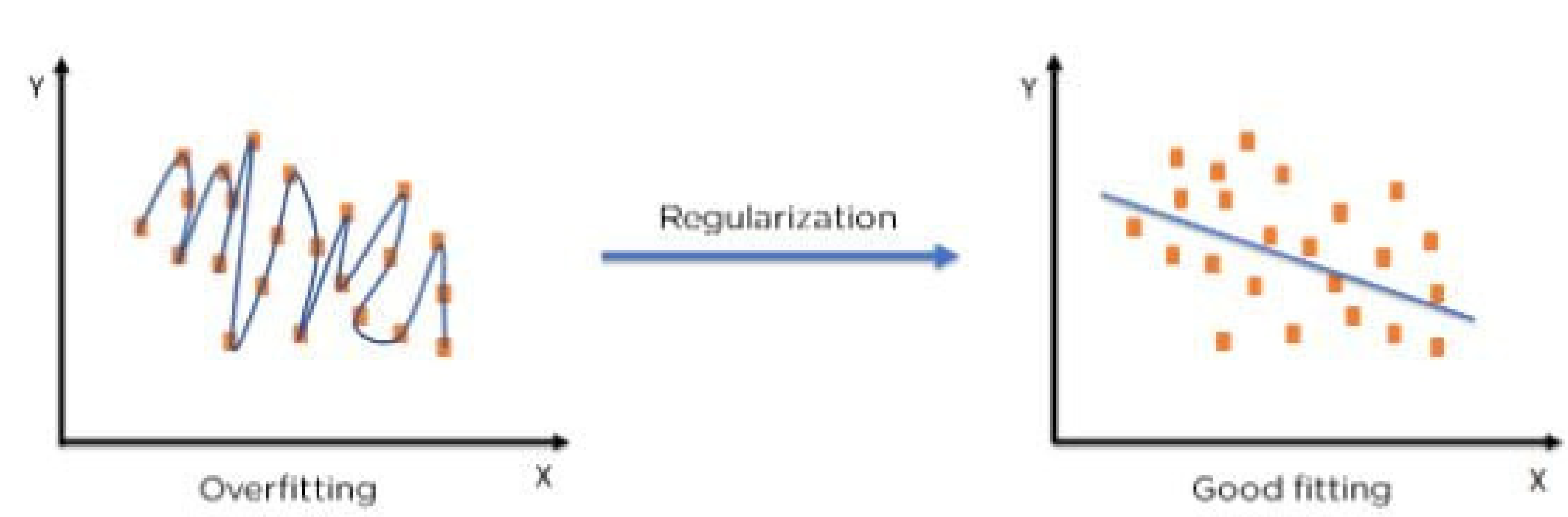
2. Understanding the concept of regularization in SVM

What is Regularization?

Regularization is a technique used to reduce errors by fitting the function appropriately on the given training set and avoiding overfitting.

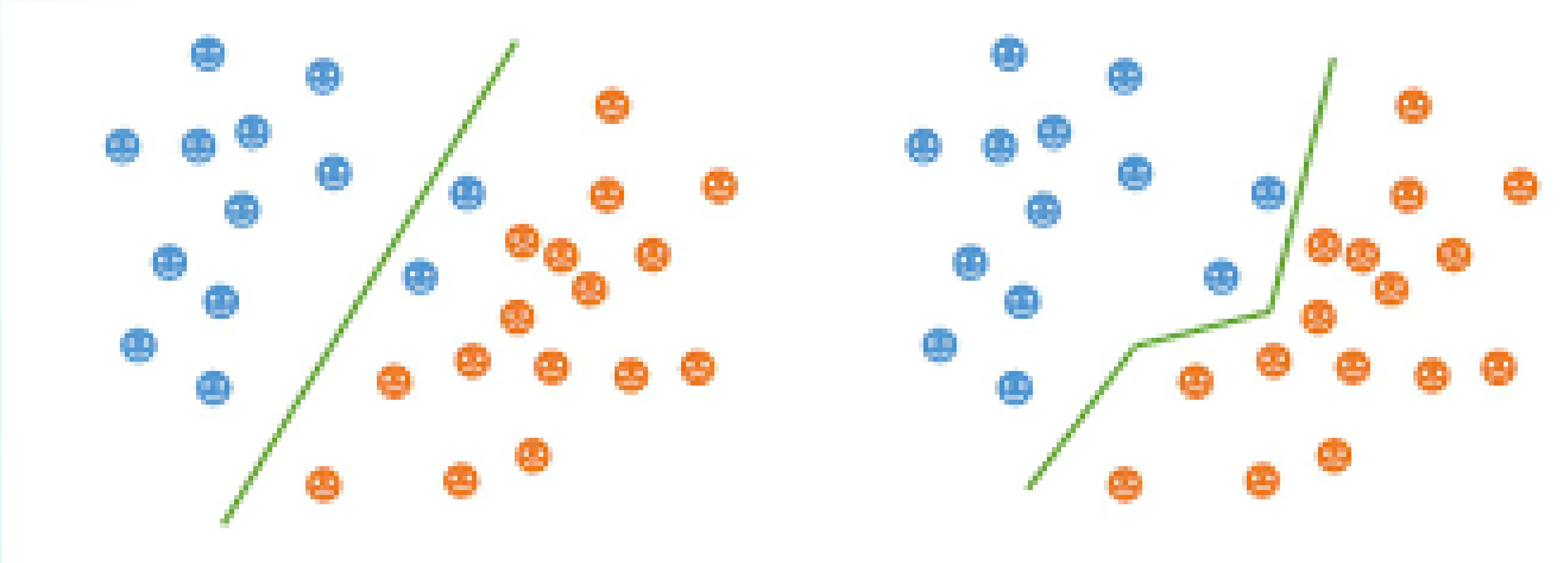
Using Regularization, we can fit our machine learning model appropriately on a given test set and hence reduce the errors in it.

.



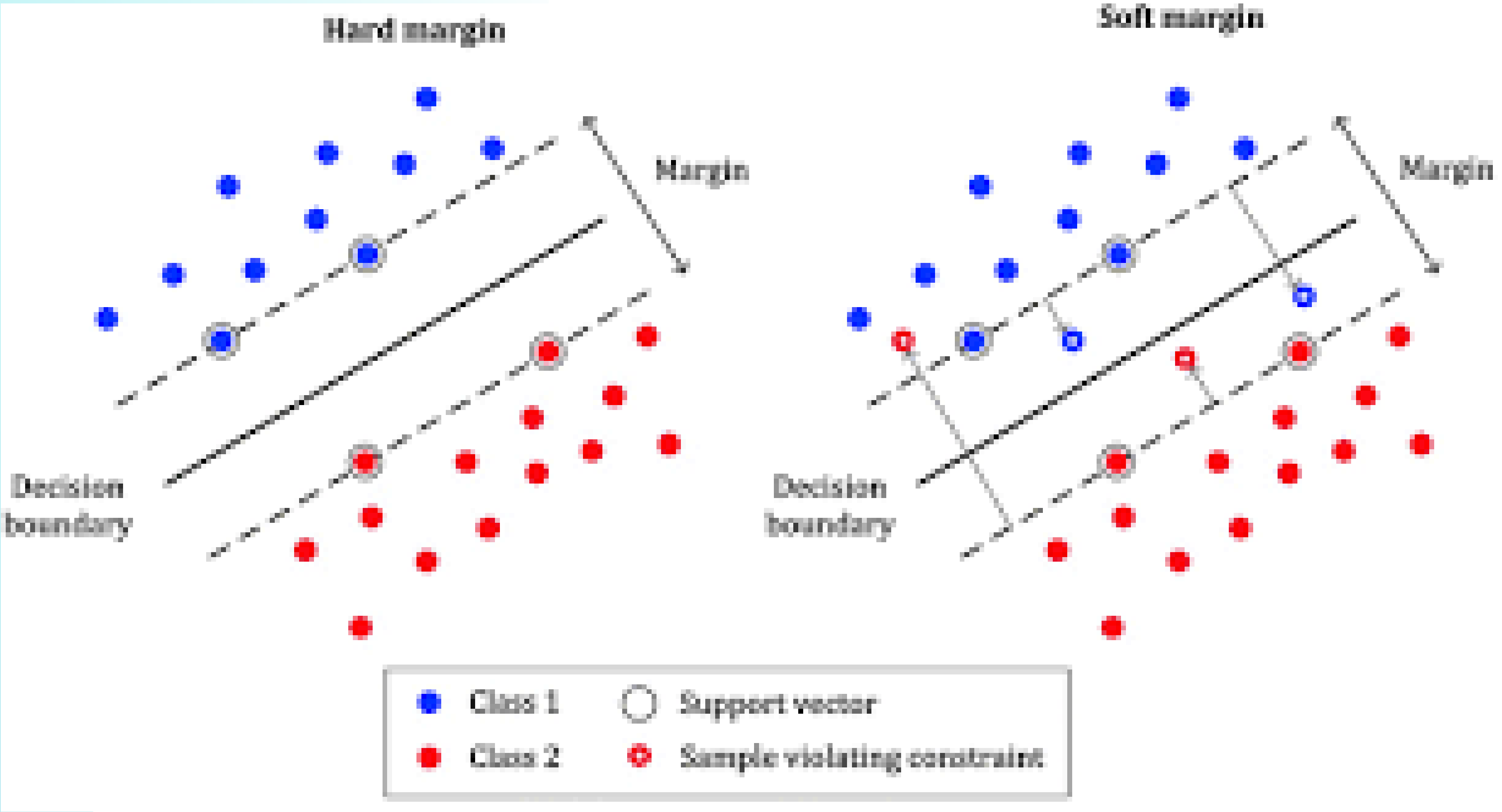
Regularization in SVM

In Support Vector Machines (SVMs): Regularization is a technique used to prevent overfitting and improve the generalization performance of the model.



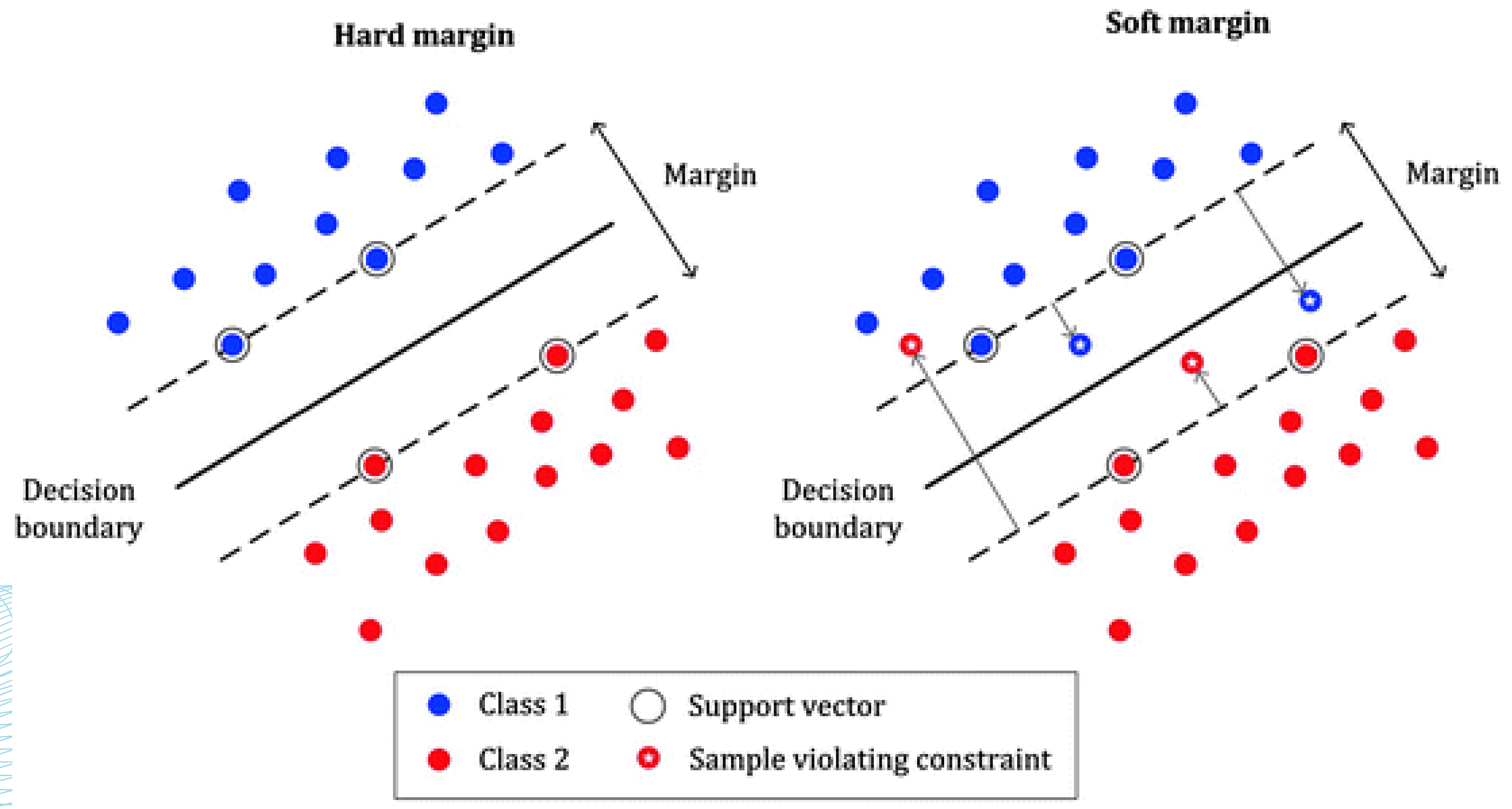
Purpose of Regularization in SVM

The regularization parameter (c) serves as a degree of importance that is given to misclassifications. SVM pose a quadratic optimization problem that looks for maximizing the margin between both classes and minimizing the amount of misclassifications.



3. Trade-offs

The trade-off between maximizing the margin and minimizing the classification error in SVM (Support Vector Machine) is controlled by the parameter C. This parameter determines the balance between achieving a larger margin and allowing misclassification errors in the training data.



Larger Margin :

- **Advantage :**
  - Robustness to outlier
  - Better Generalization
  - Increased Tolerance for Misclassification
- **Disadvantage :**
  - Reduce Training Accuracy
  - Sensitivity to Class Imbalance

Smaller Margin :

- **Advantage :**
  - High Training Accuracy
  - Potentially better on Balanced Dataset
- **Disadvantage :**
  - Sensitivity to Noise
  - Higher Risk of Overfitting

4.Discussing the impact of regularization parameter (C)

Small C Value

When the regularization parameter, C, is small, the SVM algorithm places more emphasis on achieving a wider margin rather than minimizing misclassifications. is result is a more flexible and tolerant model. C allows more misclassifications on the training set. The algorithm prioritizes finding a more significant margin and will sacrifice correctly classifying a few data points.

Larger C value

A larger C places more importance on minimizing misclassifications, leading to a narrower margin model that becomes more sensitive to individual data points and tries to classify them correctly resulting in a more complex decision boundary that adapts closely to the training data C value useful when the cost of misclassification is high, and the goal is to reduce training errors even if it means sacrificing some margin width.

Here are the key impacts of the regularization parameter C

Misclassification

The value of C determines the penalty associated with misclassification:

- A larger C value imposes a higher penalty, leading the SVM model to prioritize reducing misclassifications.
- A smaller C value allows more misclassifications and focuses on achieving a wider margin.

Model complexity

The value of C influences the complexity of the SVM model:

- A larger C value allows for more complex decision boundaries with potentially more support vectors.
- A smaller C value leads to simpler models with fewer support vectors.

overfitting and underfitting

- overfitting is often a result of a high model complexity or a large value of the regularization parameter C. The model may have high accuracy on the training set but performs poorly on new, unseen data.
- Underfitting results in a decision boundary that is overly generalized and has a wider margin. The model may have low accuracy on both the training set and new data

Impact on margin width

The regularization parameter C directly affects the width of the margin in SVM:

- A smaller C value emphasizes a wider margin, allowing more tolerance for misclassifications.
- A larger C value leads to a narrower margin, prioritizing accurate classification at the cost of margin width

CONCLUSION

The regularization parameter C in SVM allows control over the trade-off between maximizing the margin and minimizing classification errors. Selecting an appropriate C value depends on the specific problem and the relative importance of achieving a wider margin versus minimizing misclassifications.

THAN YOU !