DALARNA UNIVERSITY

Machine Learning

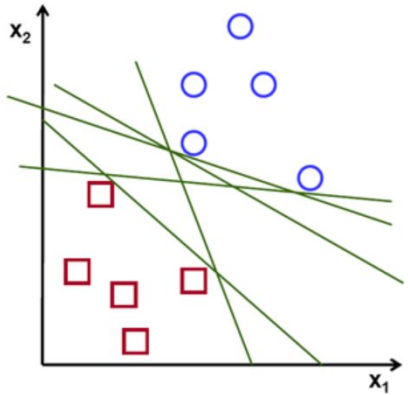
Supervised Learning
Support Vector Machines
for **Classification**

Roger G. Nyberg

Support Vector Machines – SVM

Our Goal: Classification! Separating data into different groups (classes).

- Imagine we have two distinct groups of data points. We need to find a line (or surface) that effectively divides them.



Why not just any line?

- In many cases, there are **many** lines that can separate the two classes. Which one is the best?

The Core Idea of SVM: Find the *optimal* separating boundary!

- The "best" boundary is the one that has the largest possible distance (the margin) to the nearest data points from both classes.
- Like a "motorway" between the classes – we want the widest possible motorway.

DALARNA UNIVERSITY

The Margin – Key to Robust Classification

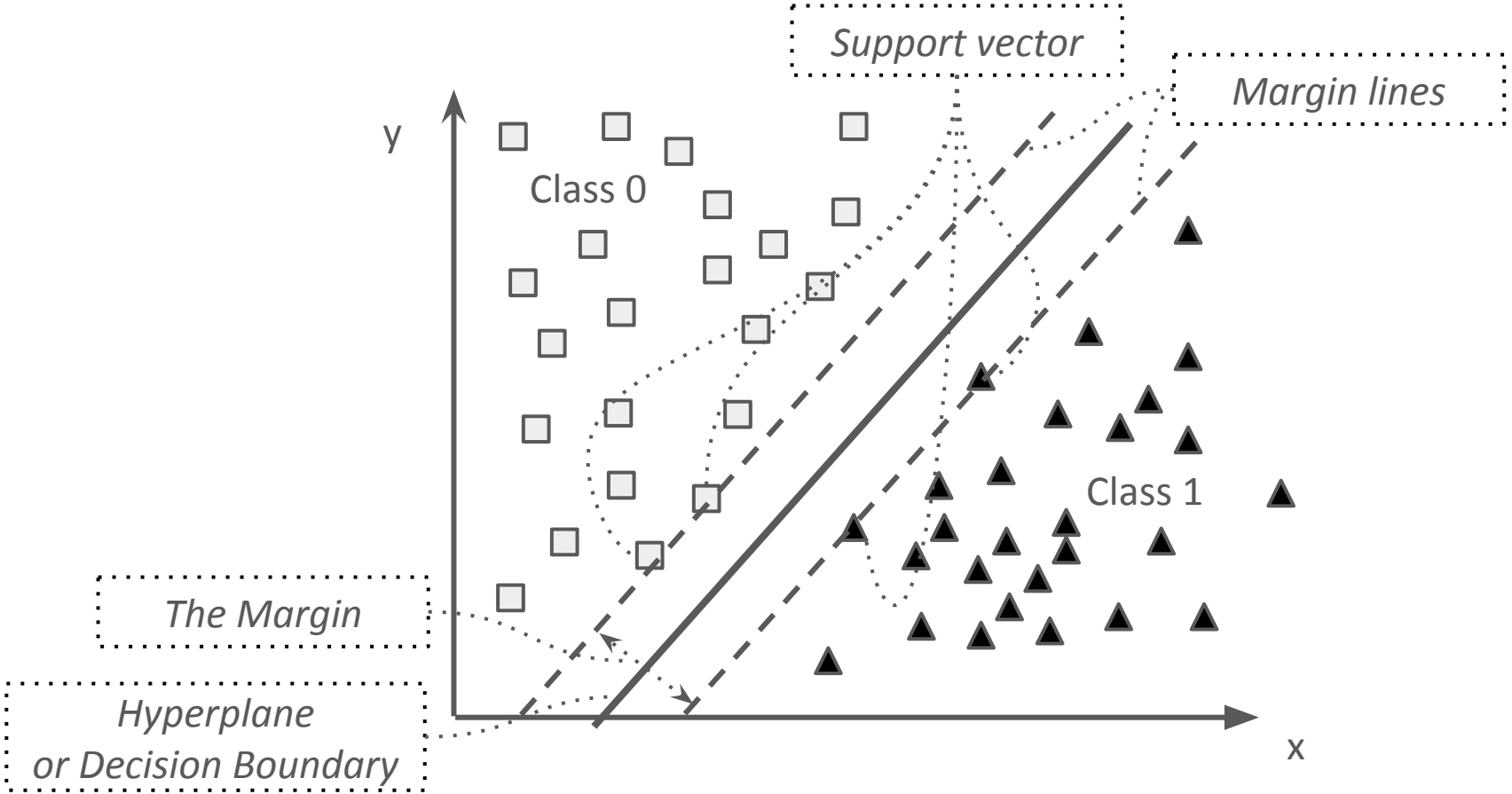
The Margin: The distance that the SVM algorithm seeks to **maximize**.

- It is the distance between the separation boundary (the **hyperplane**) and the nearest data point from either class.
- A large margin means the model is more **robust** and generalizes better to new, unseen data. It reduces the risk of **overfitting**.

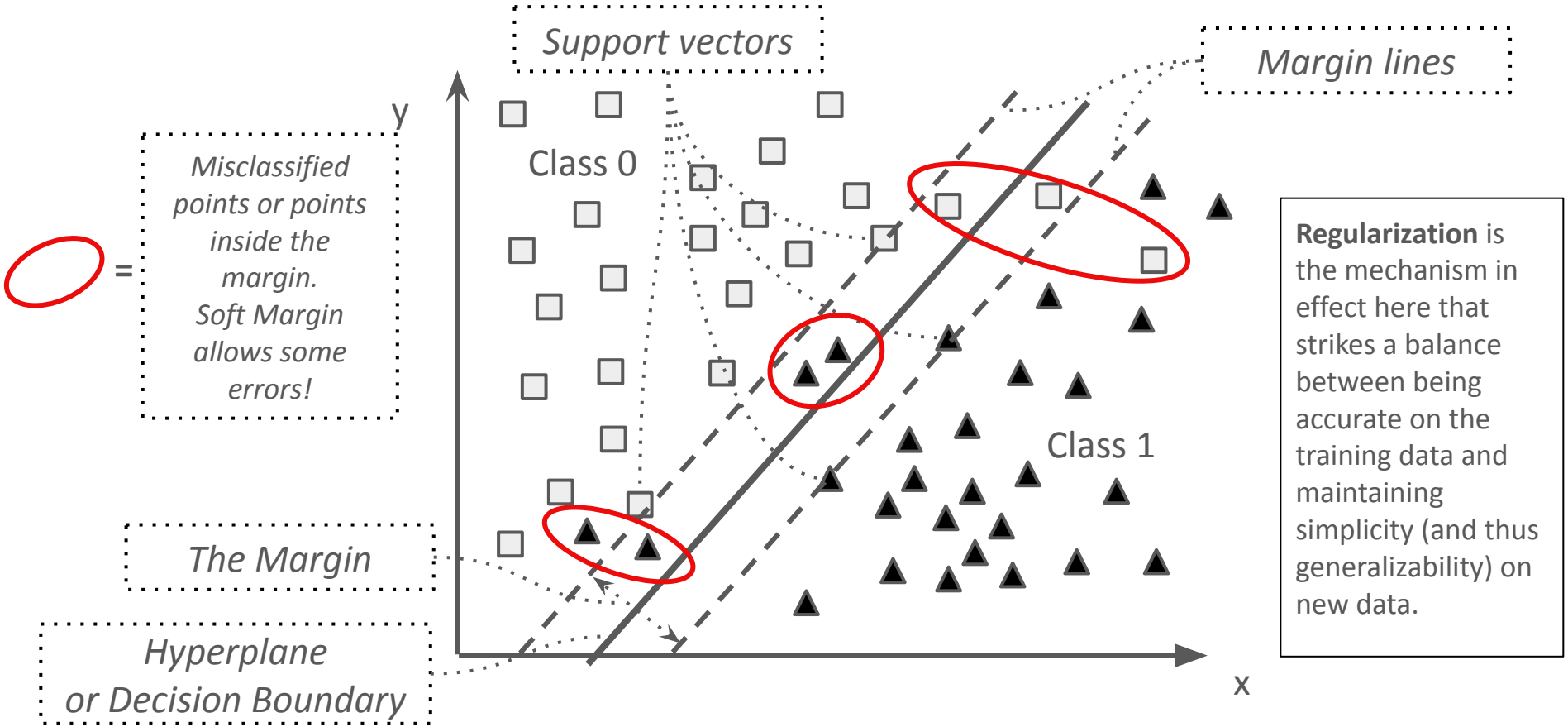
Support Vectors (SVs): The critical data points that "support" the margin.

- These are the points that lie closest to the hyperplane on either side.
- They are the **only** points required to define the position and orientation of the optimal hyperplane. If you move any other non-SV point, the boundary remains unchanged.

Hard Margin SVM



Soft Margin SVM - Handling non-linearly separable data or noisy data



Regularization is the mechanism in effect here that strikes a balance between being accurate on the training data and maintaining simplicity (and thus generalizability) on new data.

When a Straight Line Isn't Enough – The Kernel Trick

The Problem: Often, our data is **not linearly separable** in its original dimensions.

- Example: Data shaped like a circle, where one class is in the center and the other forms a ring around it. A straight line simply won't work.

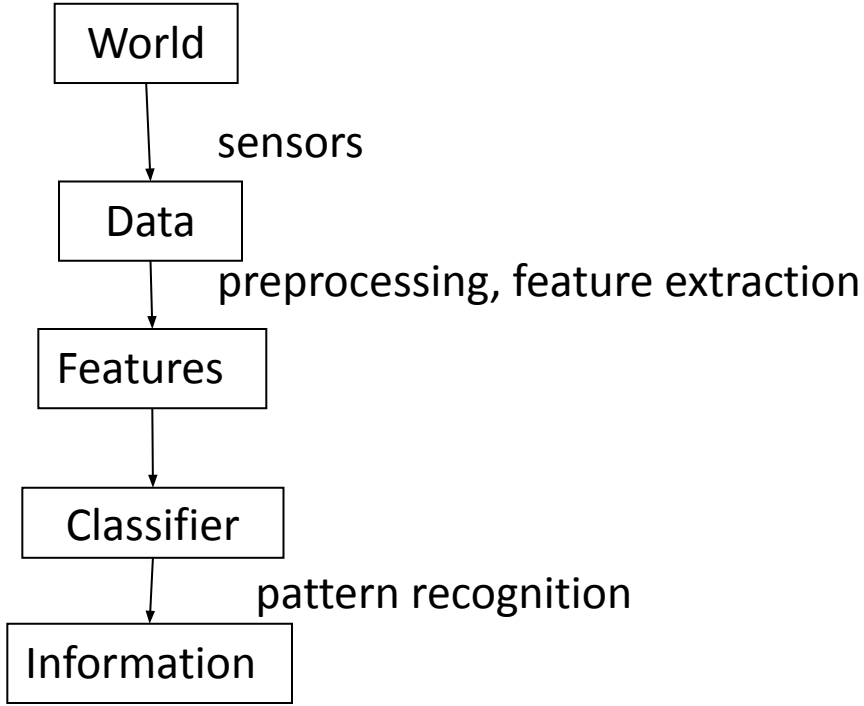
The Solution – The Kernel Trick:

- The SVM implicitly "maps" the data into a **higher-dimensional space** where it *becomes* linearly separable.

Common Kernels: Radial Basis Function (RBF) / Gaussian Kernel, Polynomial Kernel.

From the world to information

Overview, with example, of the whole sequence as 7 steps:

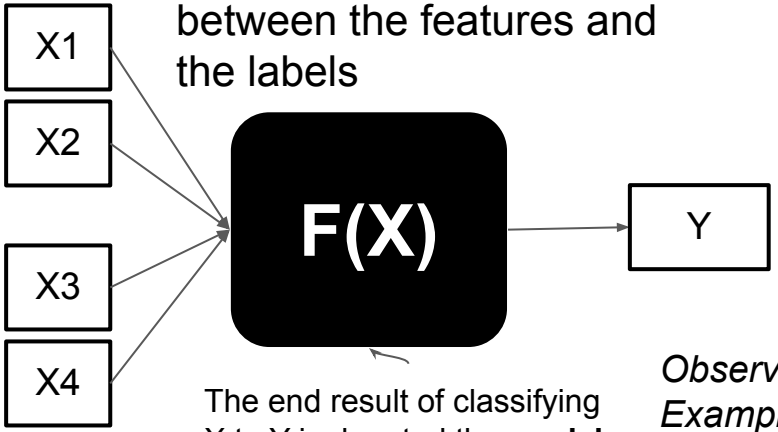


Supervised Learning

You want to find out if it is beer is beer or wine

	Alcohol [%] (feature 1)	Bitterness (feature 2)	Beverage (Labels)
Observation 1 Example 1 	20.8	29.9	Wine
Observation 2 Example 2 	1.7	7.6	Beer
	9.4	15.3	Beer
	14.0	17.2	Wine
	...	...	...
Observation n Example n	4.15	14.4	Beer

The classification algorithm (i.e. classifier) you have chosen, tries to find the ***relationship*** in between the features and the labels



The end result of classifying X to Y is denoted the **model**

DALARNA UNIVERSITY

THANK YOU! :)