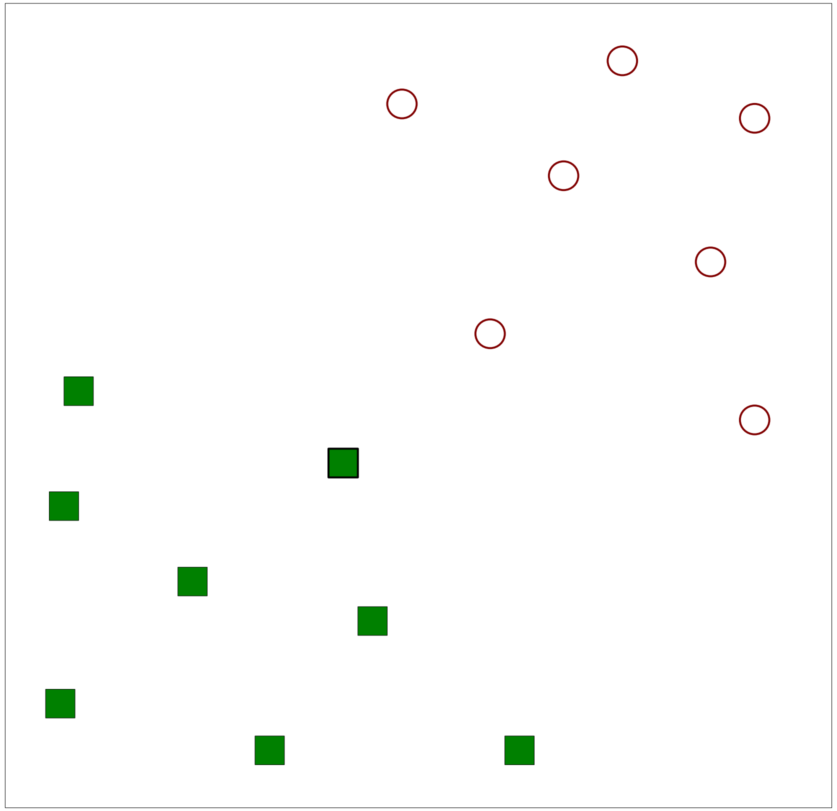
Support Vector Machine

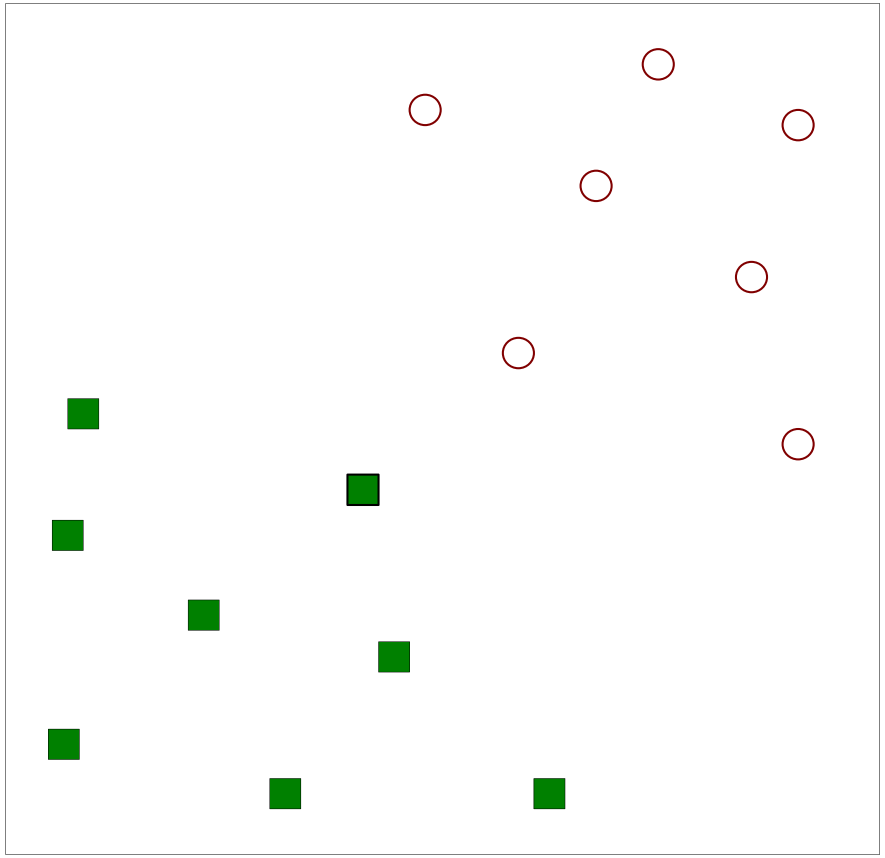
SVM

Support vector machines (SVMs) are supervised learning models with associated learning algorithms that analyze data and recognize patterns, used for classification and regression analysis.

How to separate these points?

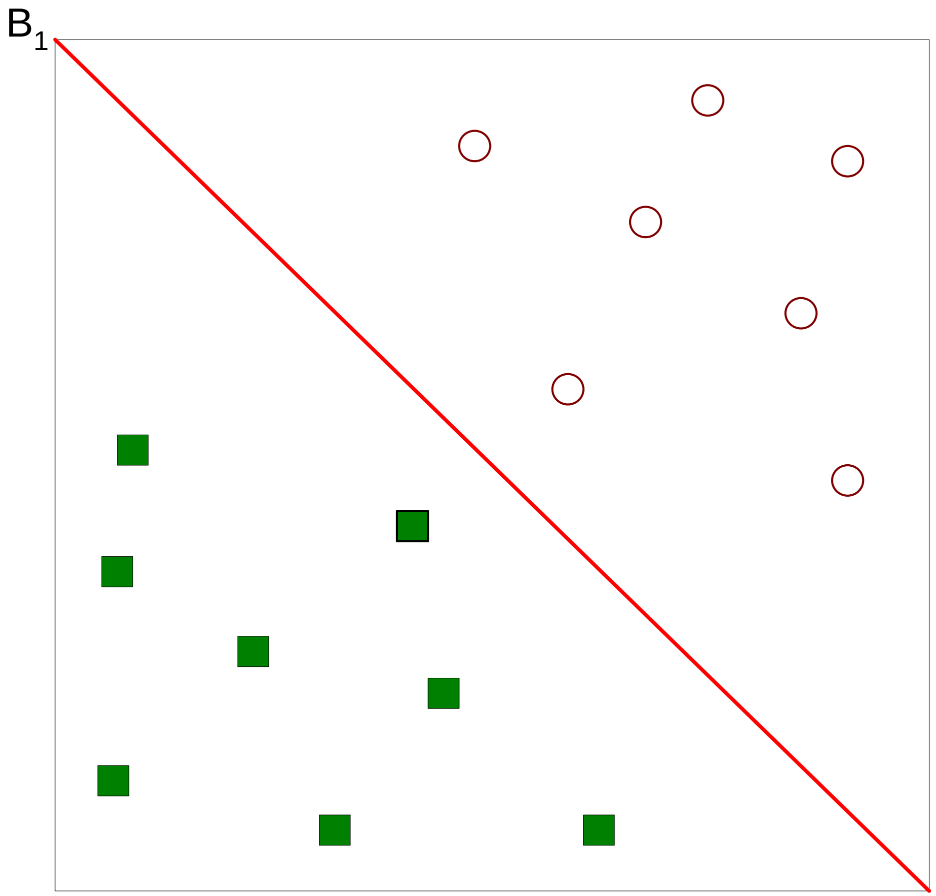


Support Vector Machines



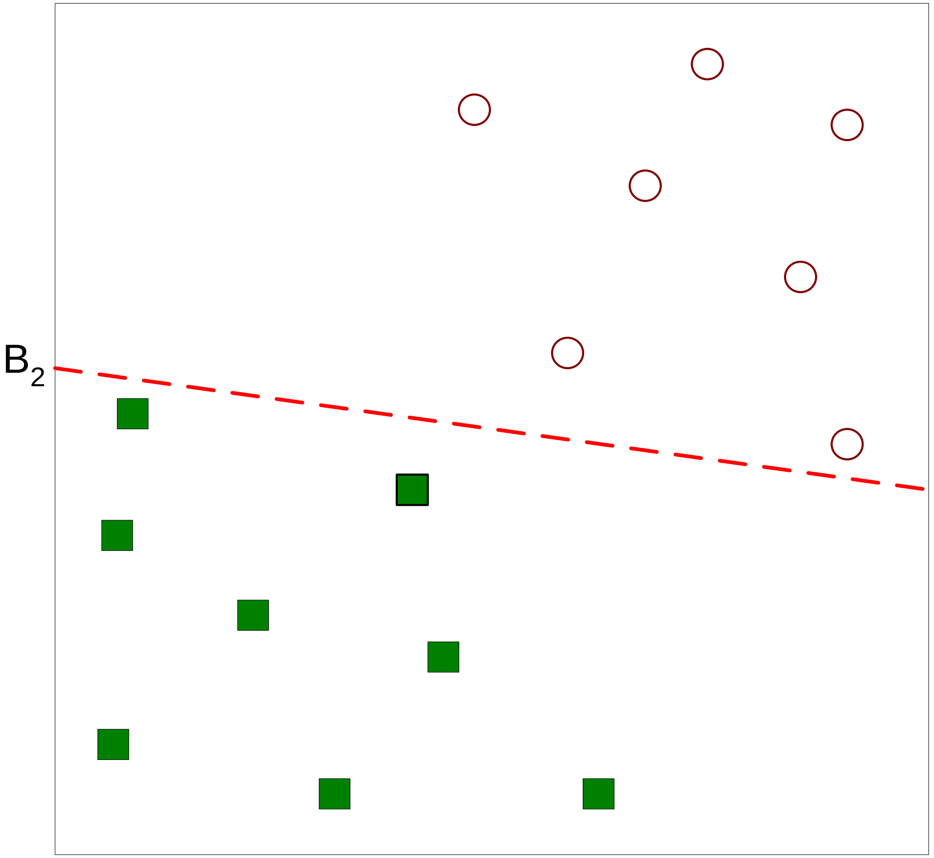
- Find a linear hyperplane (decision boundary) that will separate the data

Support Vector Machines



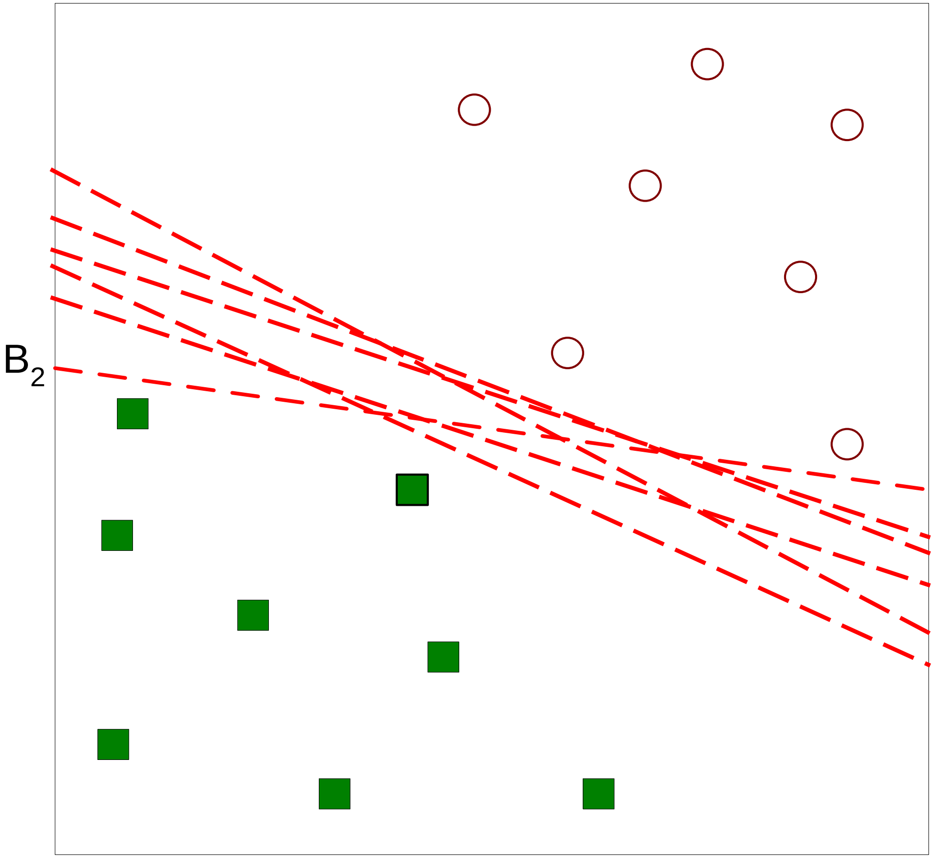
- One Possible Solution

Support Vector Machines



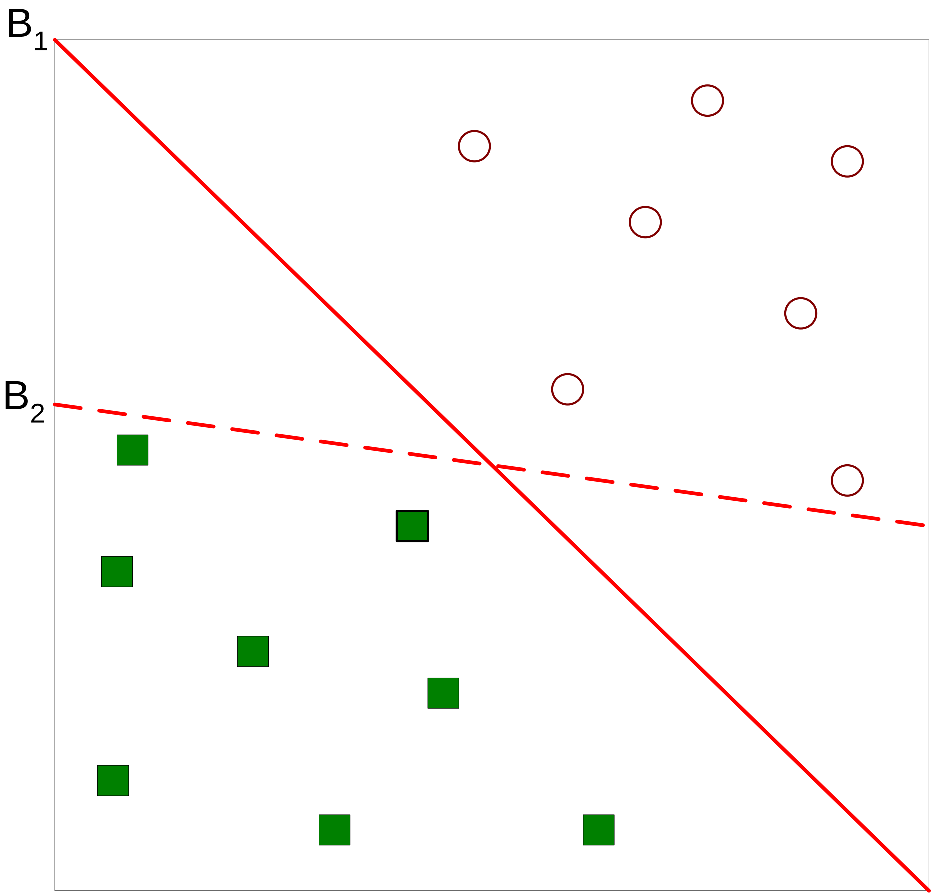
- Another possible solution

Support Vector Machines



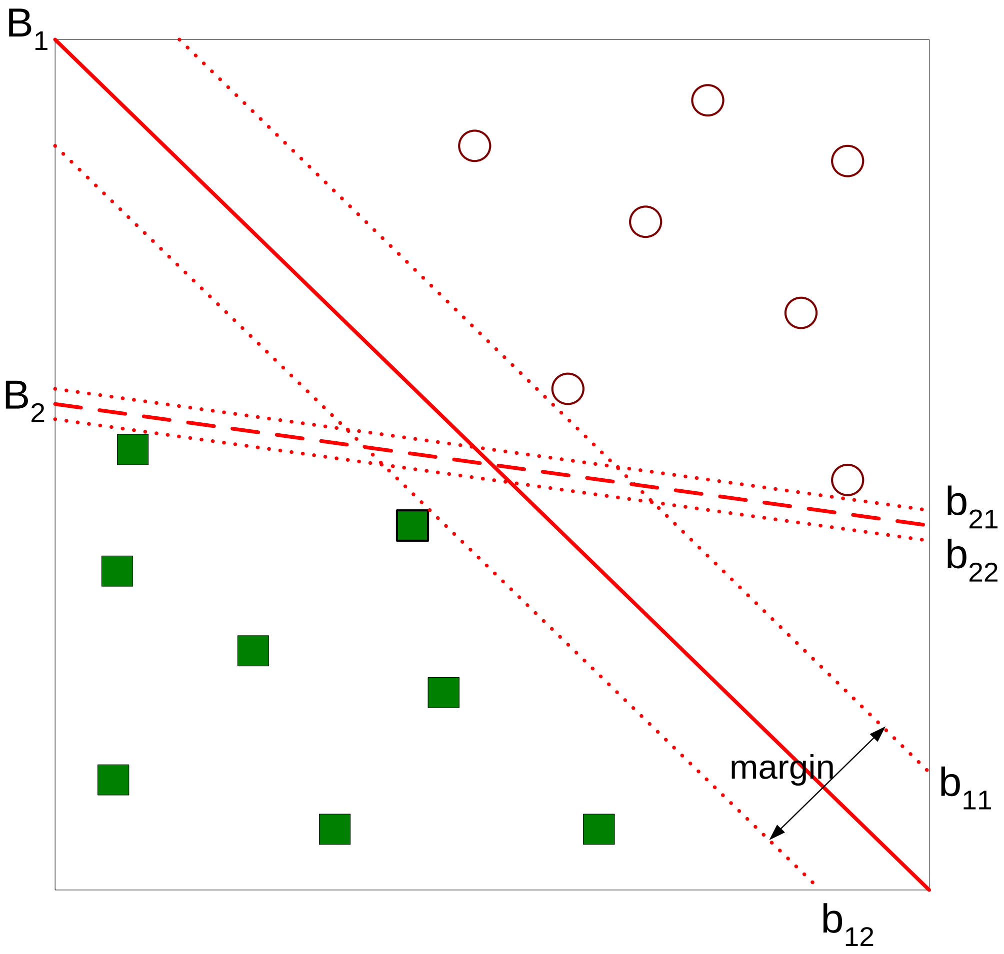
- Other possible solutions

Support Vector Machines



- Which one is better? B1 or B2?
- How do you define better?

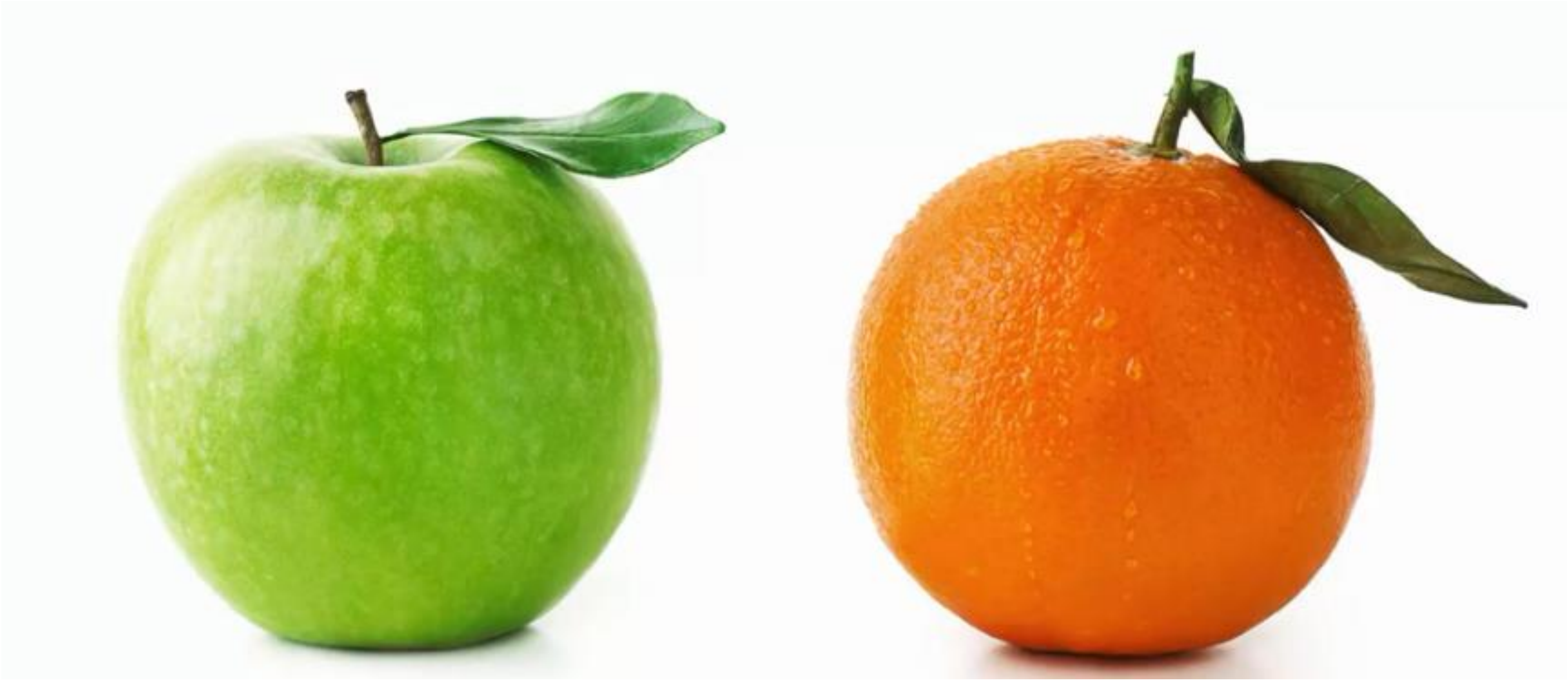
Support Vector Machines



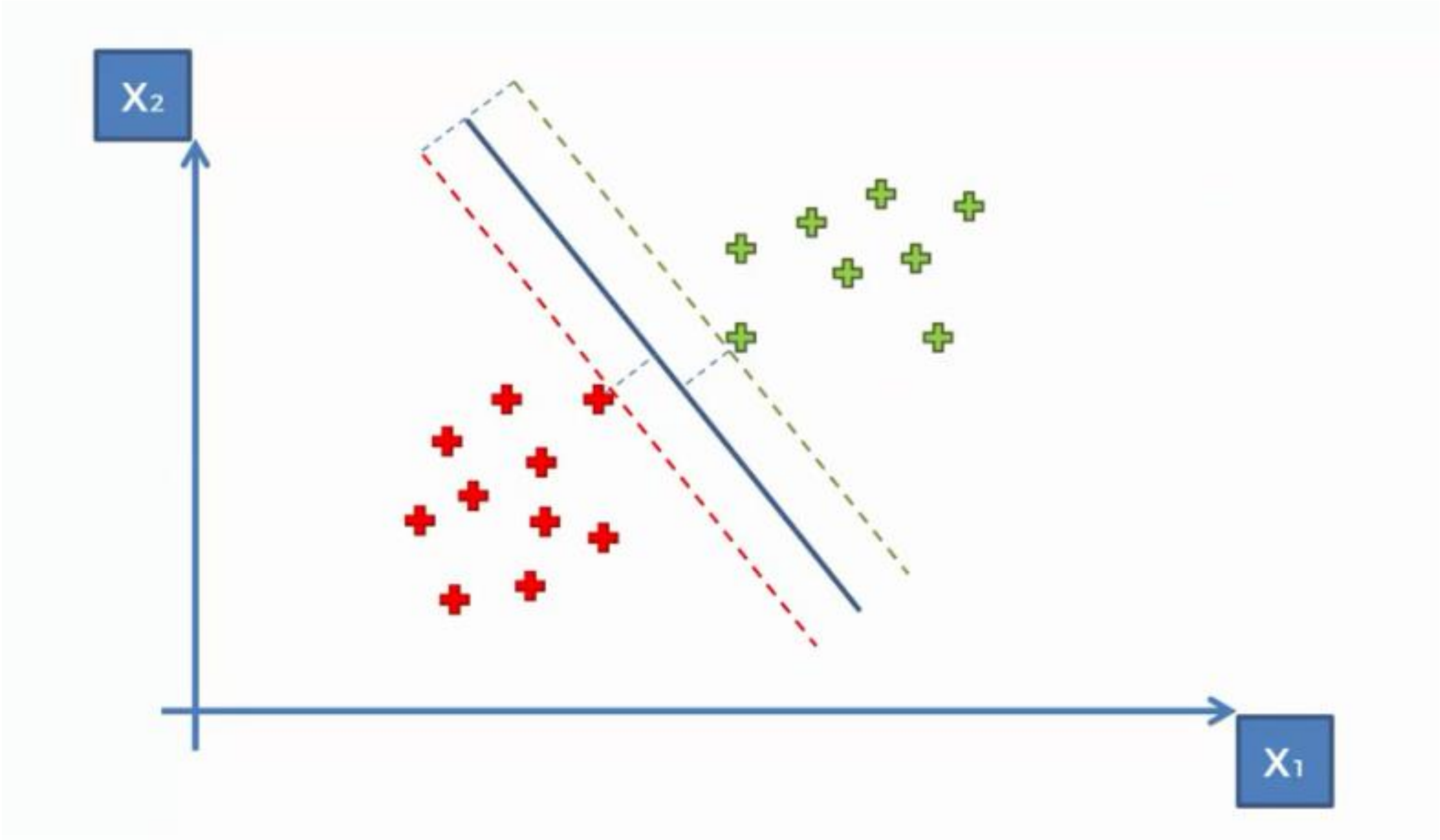
- Find hyperplane maximizes the margin => B1 is better than B2

What's So Special About SVMs?

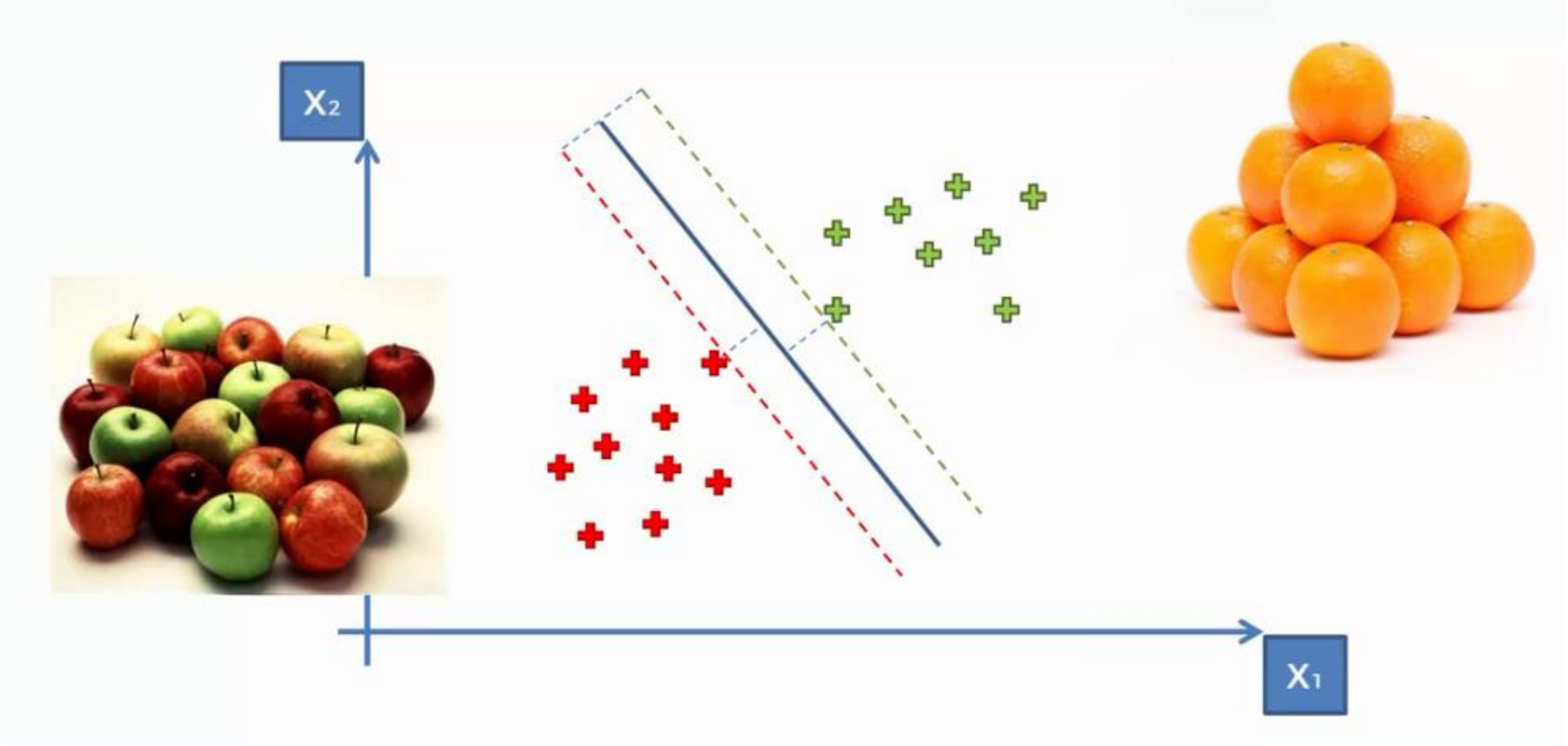
What's so special about SVMs?



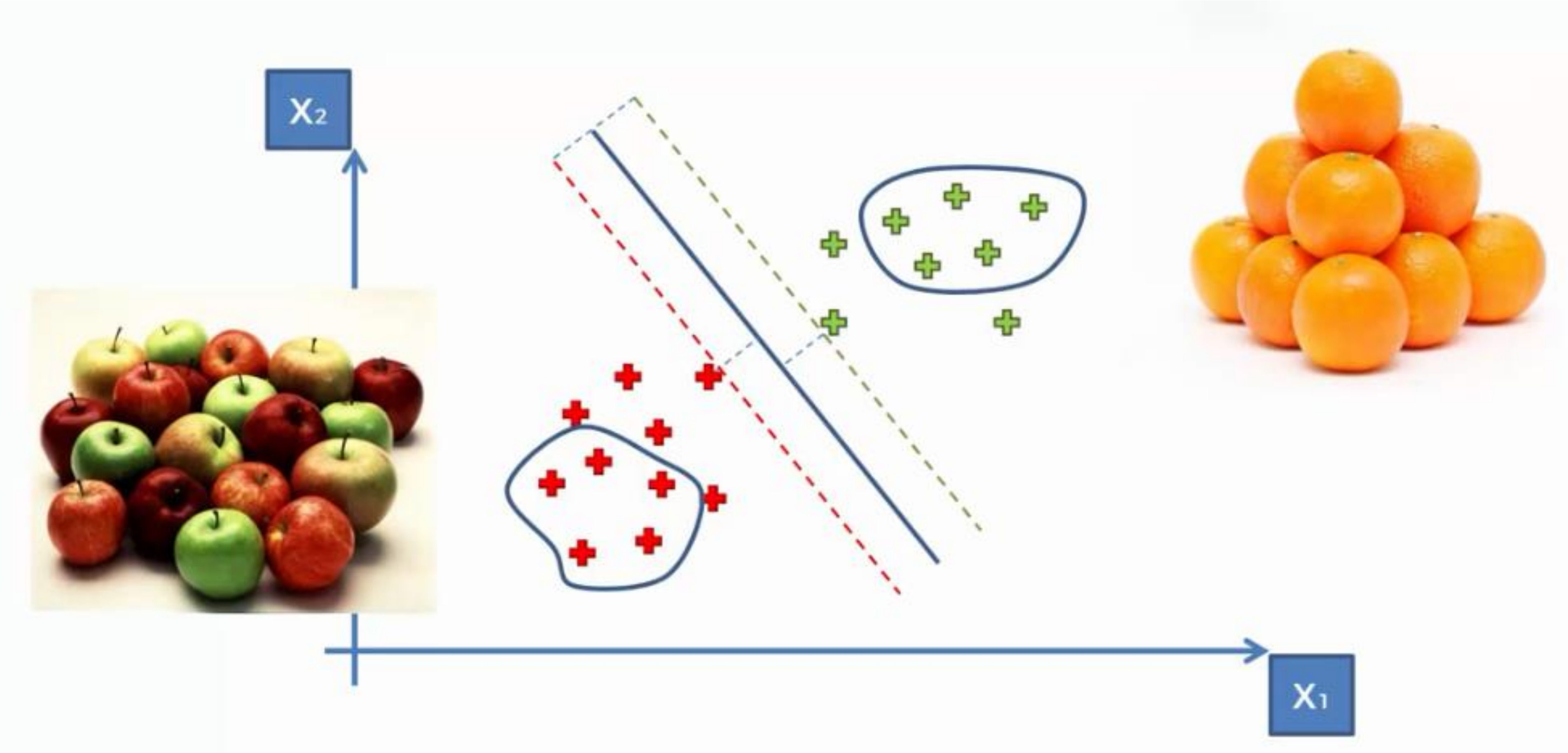
What’s so special about SVMs?



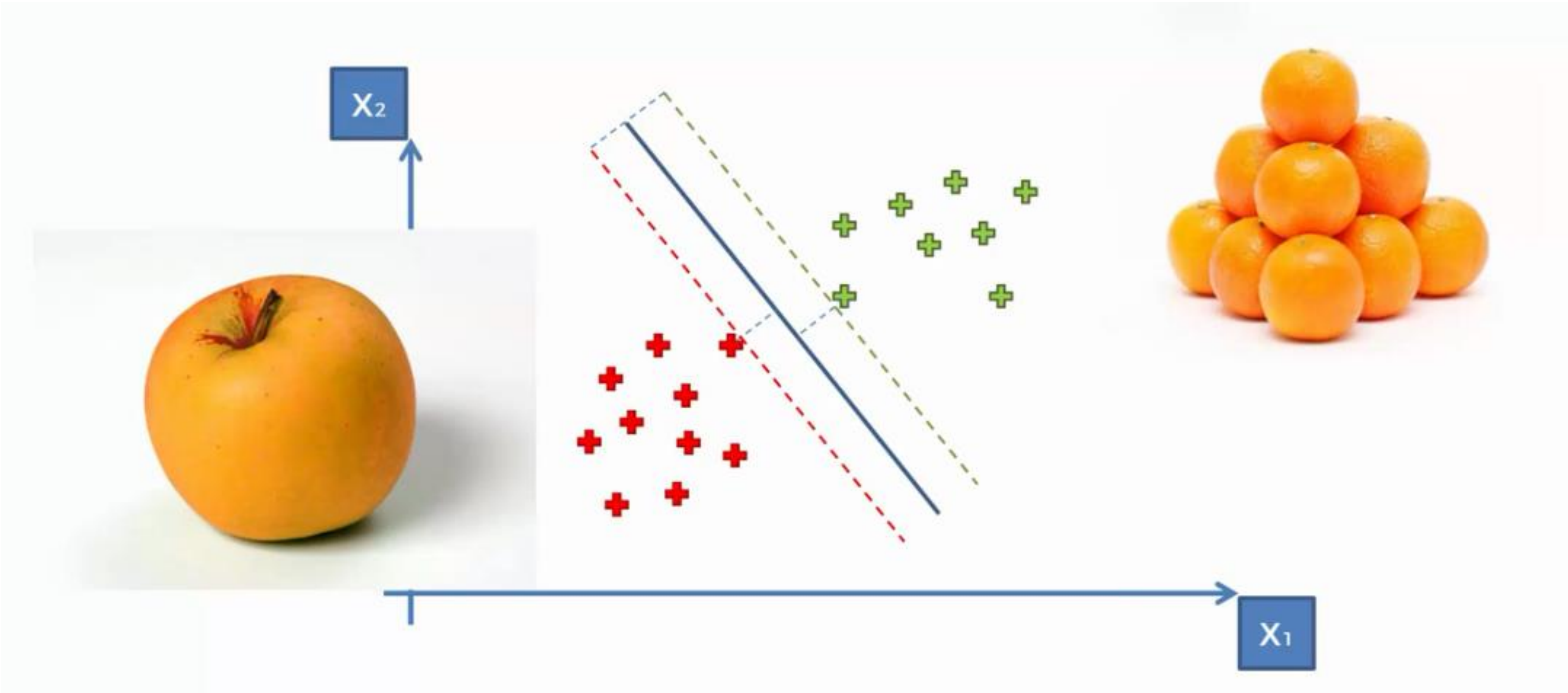
What's so special about SVMs?



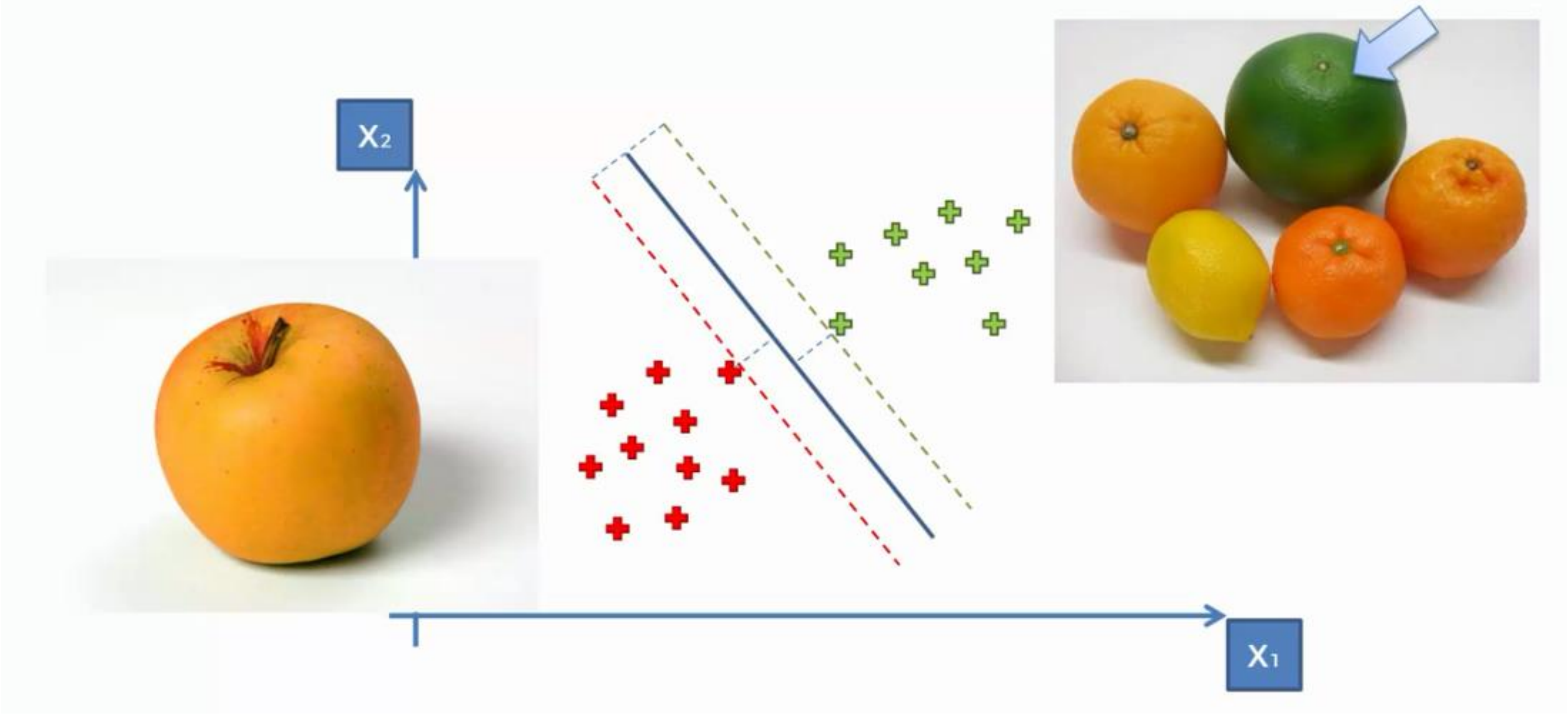
What's so special about SVMs?



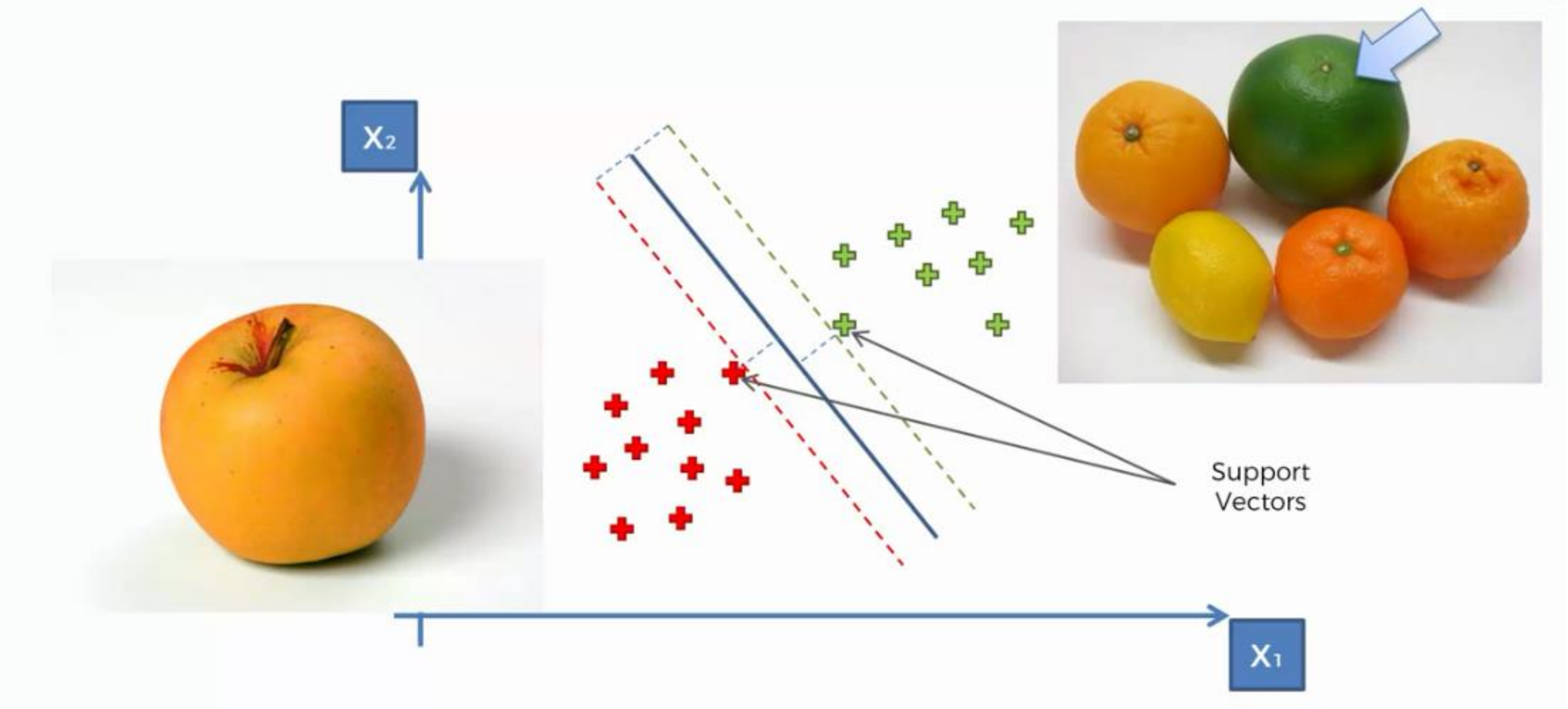
What’s so special about SVMs?



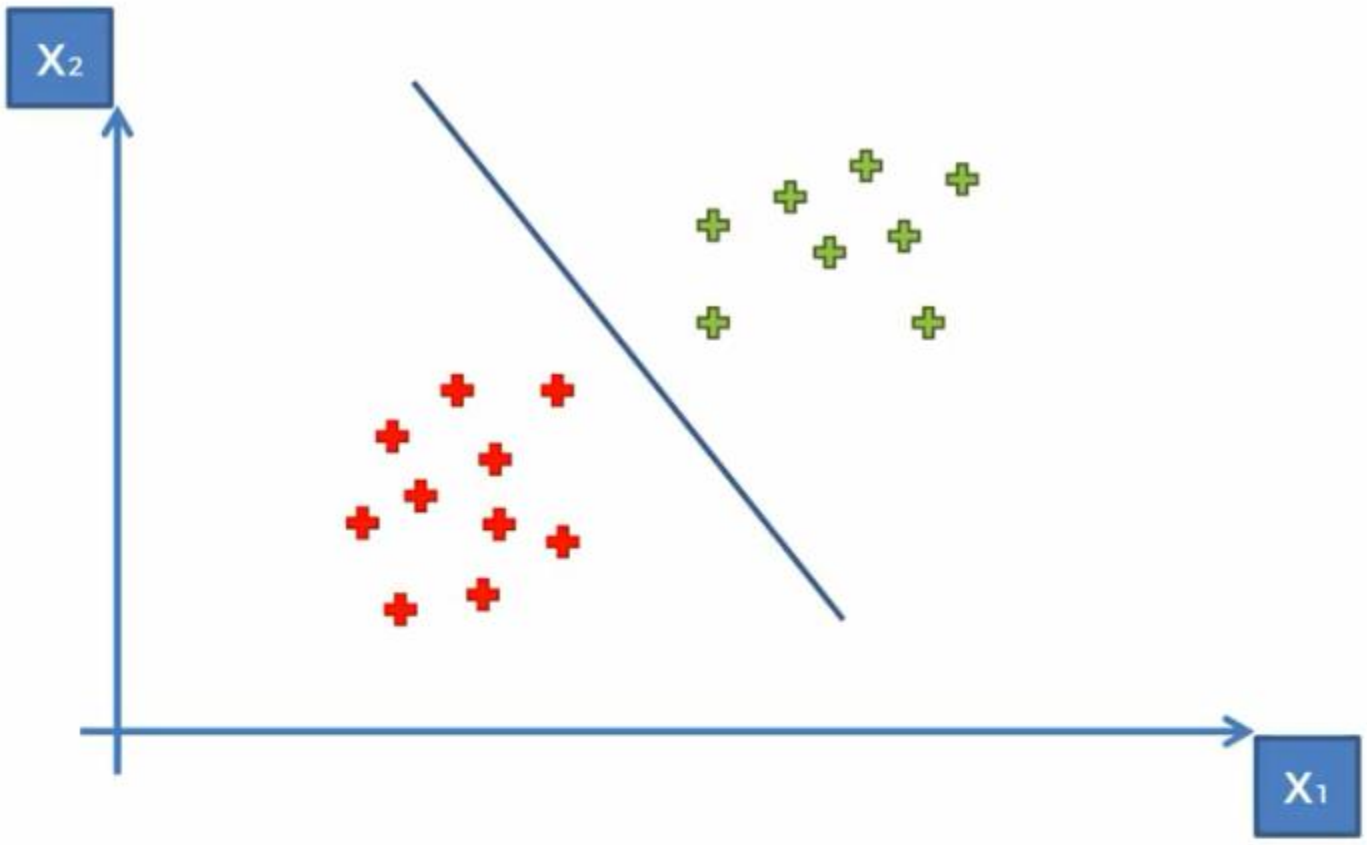
What’s so special about SVMs?



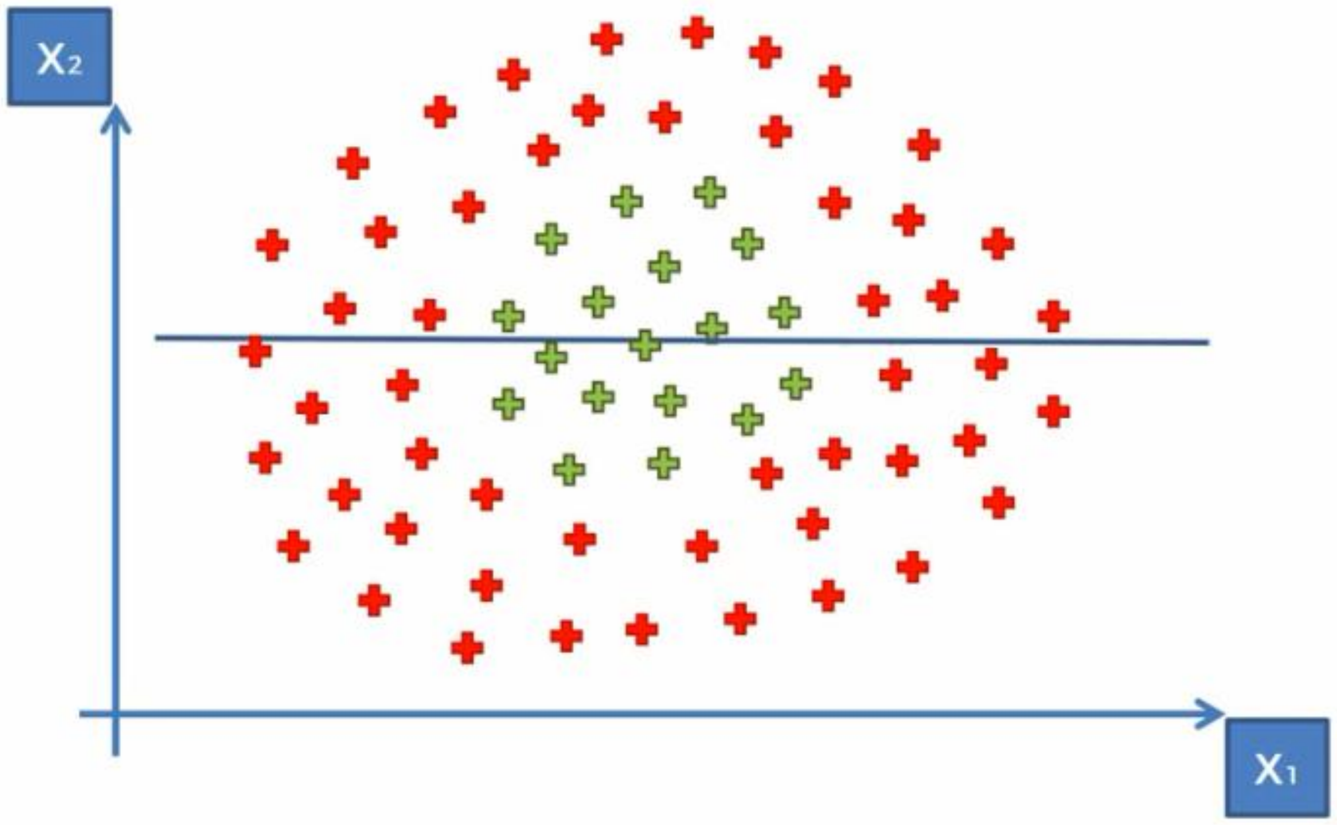
What’s so special about SVMs?



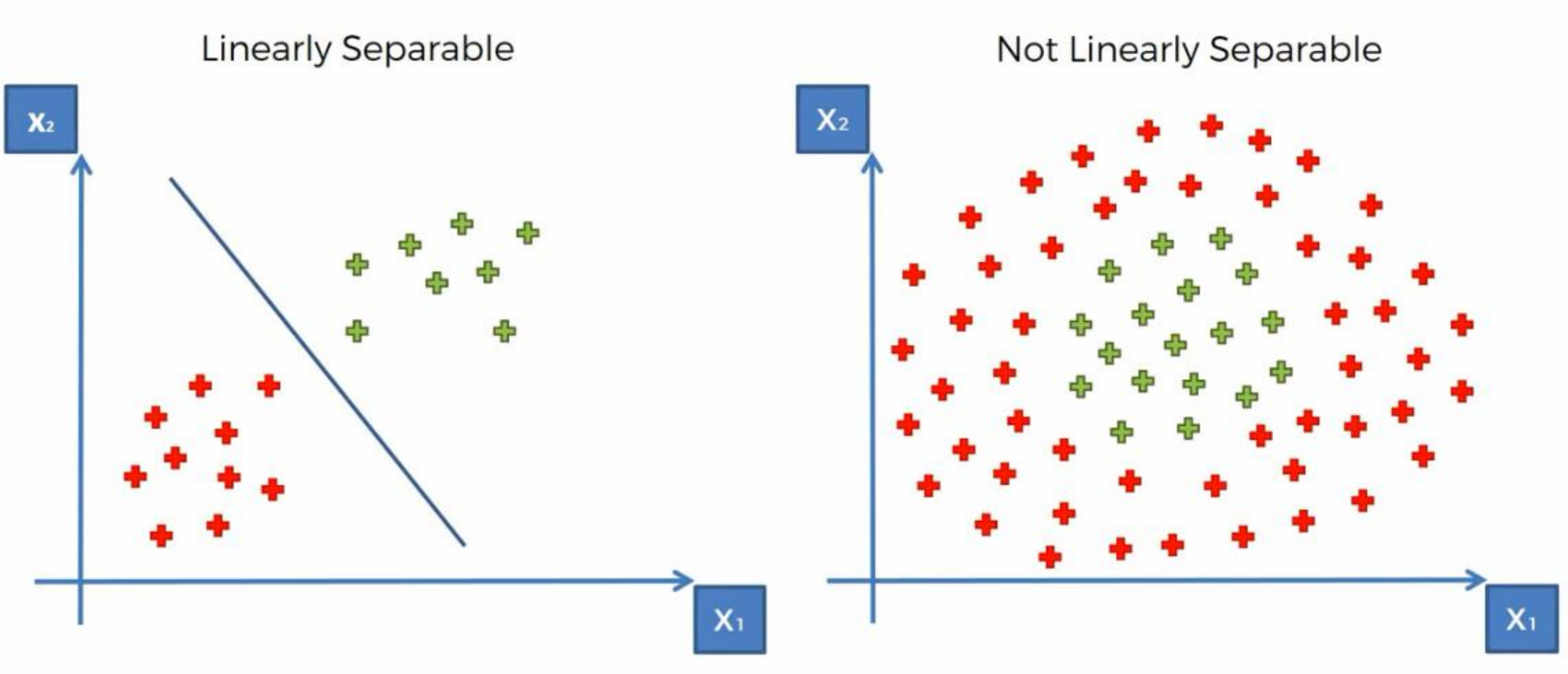
SVM separates these points



What about these points?



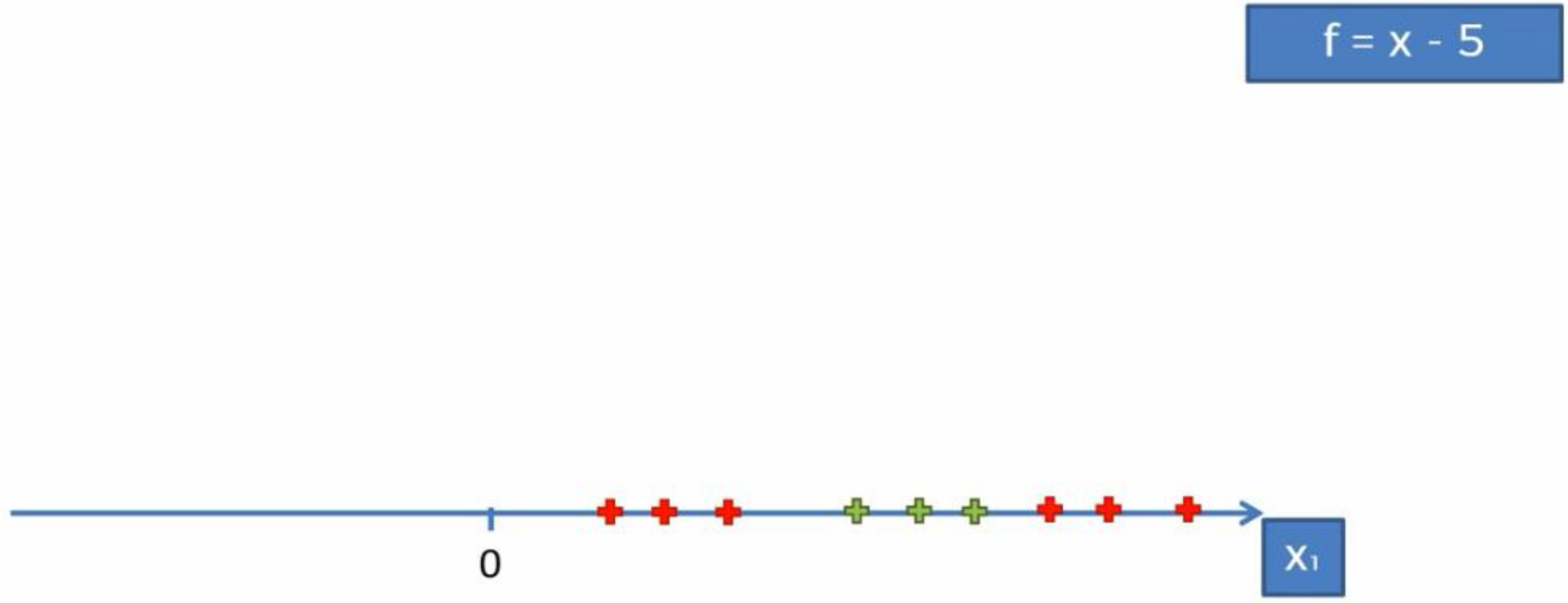
Linearly Separability



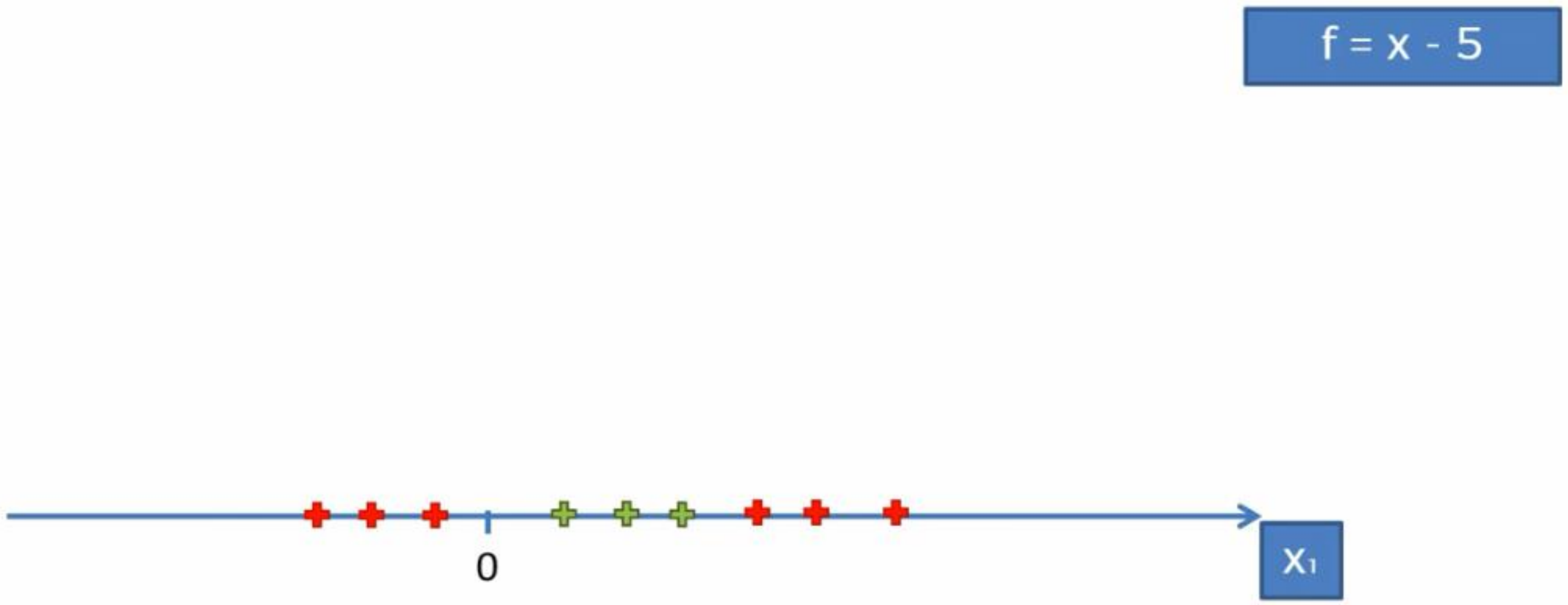
A Higher-Dimensional Space

Mapping to a higher dimension

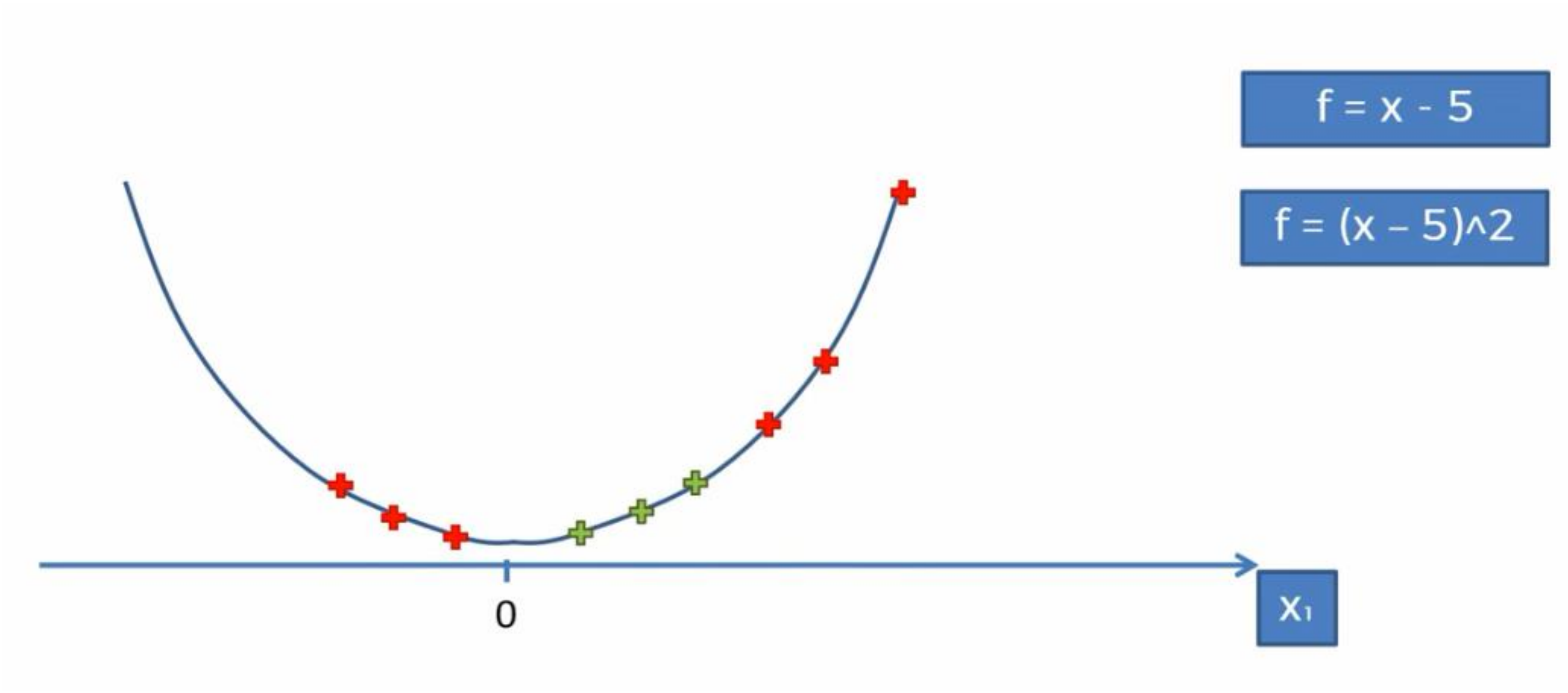
Mapping to a higher dimension



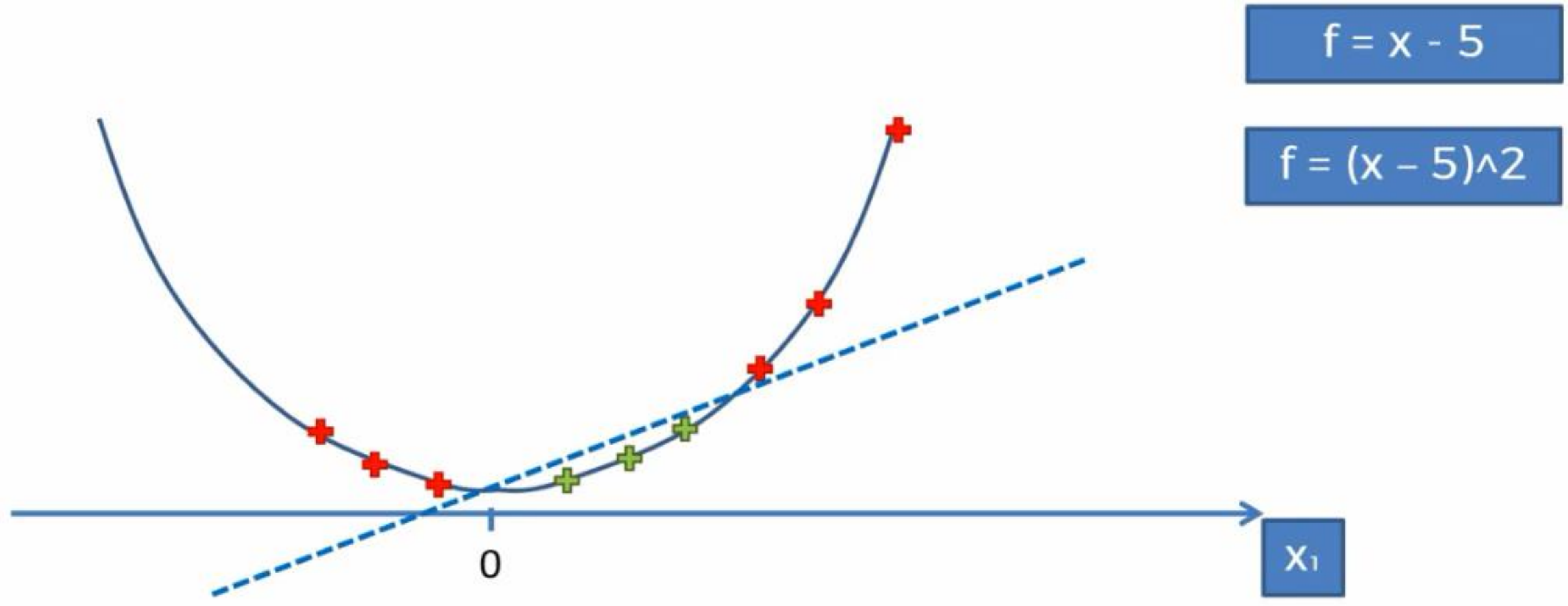
Mapping to a higher dimension



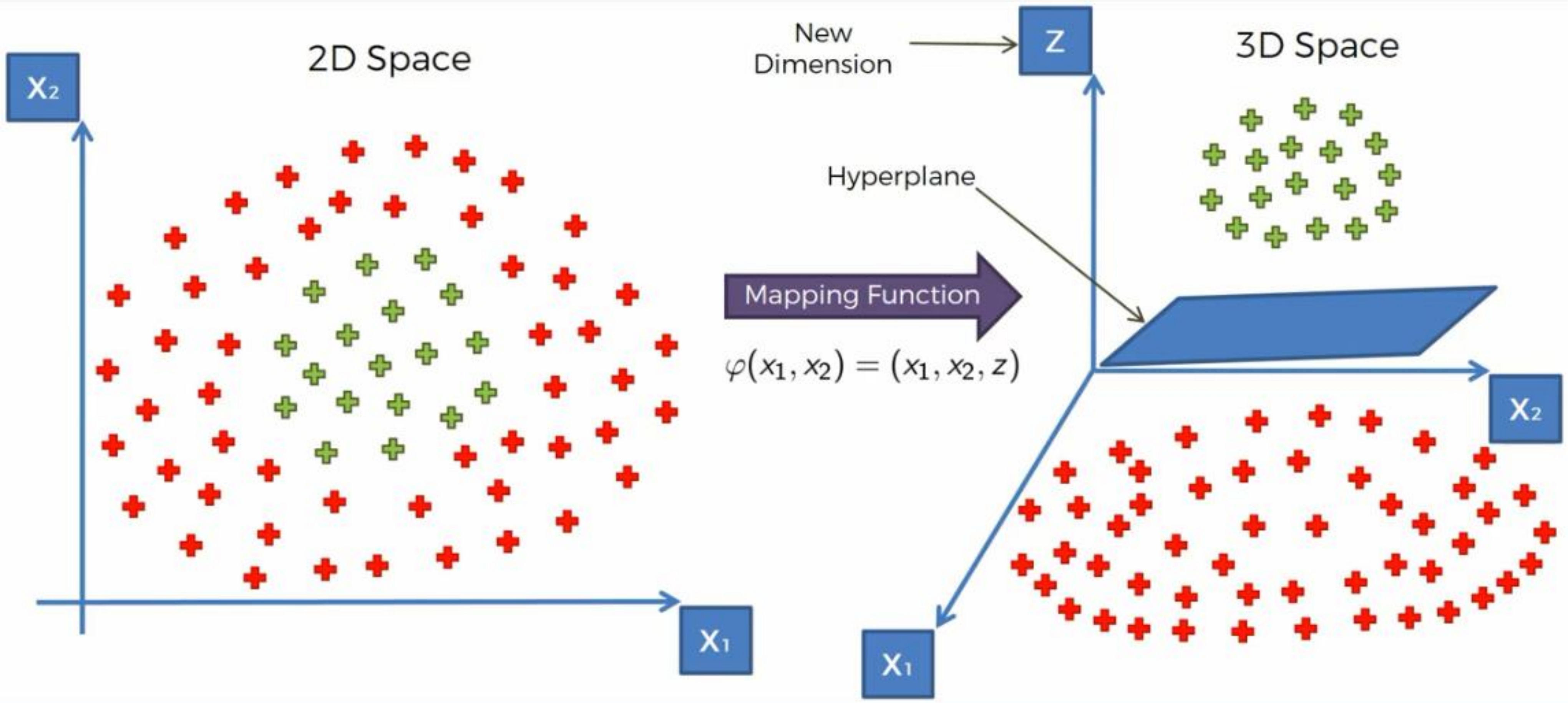
Mapping to a higher dimension



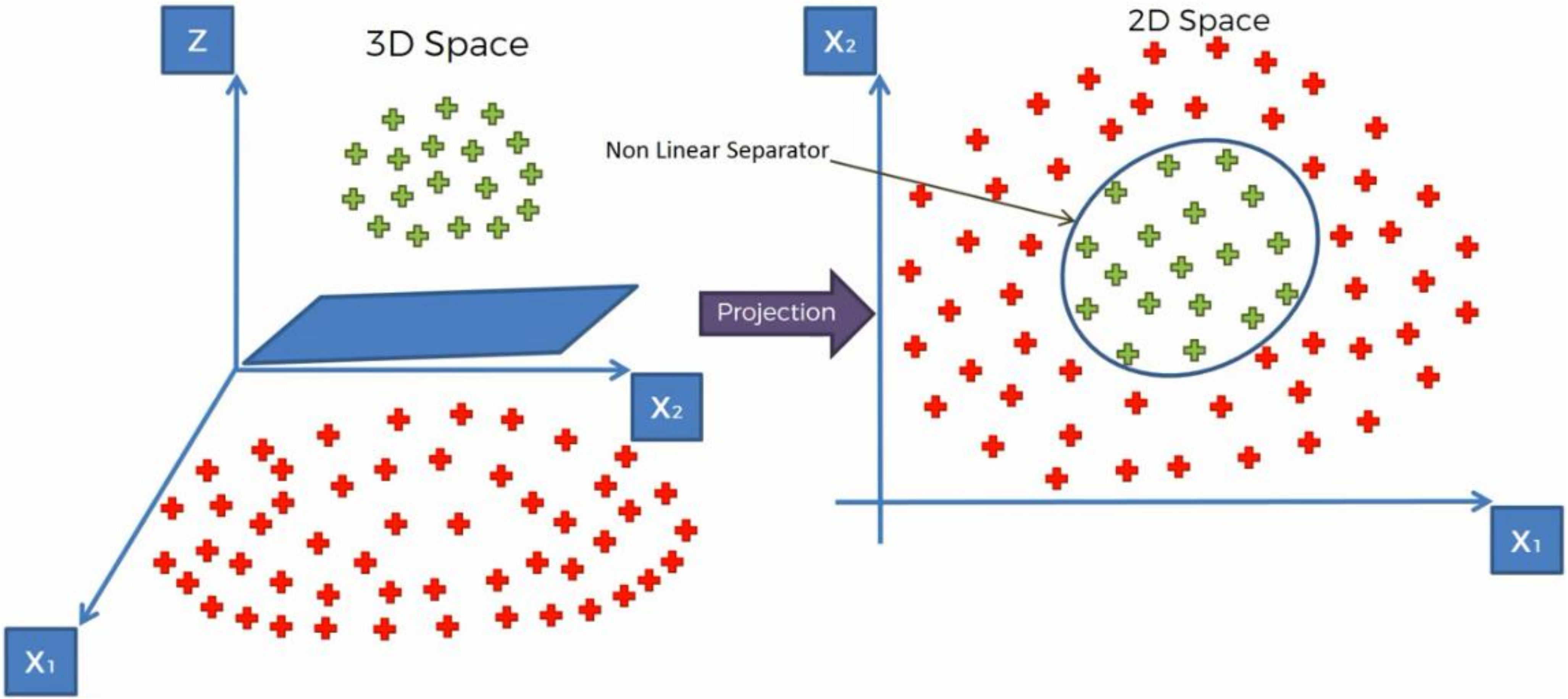
Mapping to a higher dimension



Mapping to a higher dimension



Projecting back to 2D space



Drawback

Mapping to a Higher Dimensional Space can be highly compute-intensive

Characteristics of SVM

- Since the learning problem is formulated as a convex optimization problem, efficient algorithms are available to find the global minima of the objective function (many of the other methods use greedy approaches and find locally optimal solutions)
- Overfitting is addressed by maximizing the margin of the decision boundary, but the user still needs to provide the type of kernel function and cost function
- Difficult to handle missing values
- Robust to noise
- High computational complexity for building the model