Introduction to ML and Support Vector Machines

- Vikas Thammanna Gowda

Agenda

- What is ML? Why ML?
- Categories of ML
- Support Vector Machines
  - Basic idea
  - Special Cases
  - History and Development
  - Math

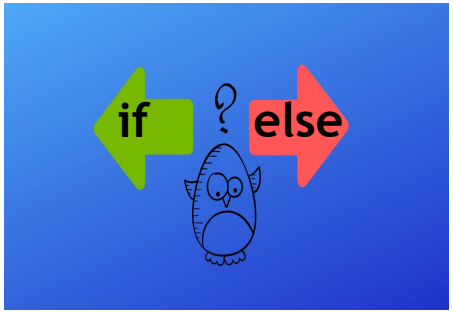
What is Machine Learning?

Arthur Samuel: Machine Learning is the field of study that gives computers the ability to learn without being explicitly programmed.

ML is a collection of **algorithms** and techniques used to create computational systems that **learn from data** in order to make **predictions** and inferences.

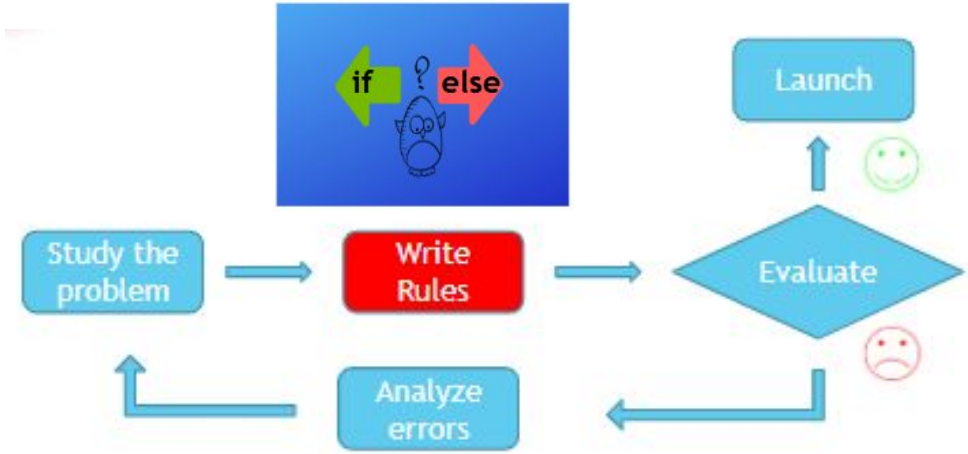
Why Machine Learning?

4U
credit card
free
amazing

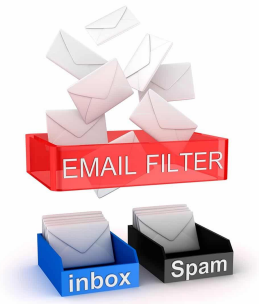


Why Machine Learning?

4U
credit card
free
amazing

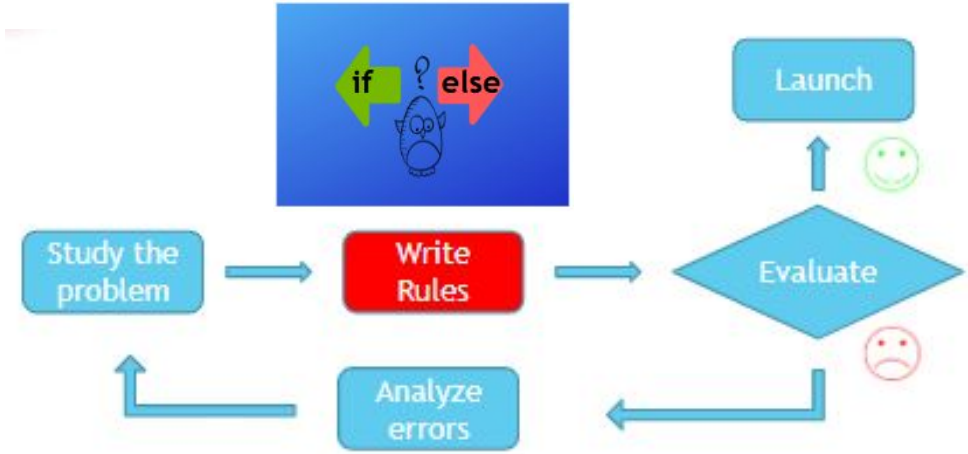


The traditional approach



Why Machine Learning?

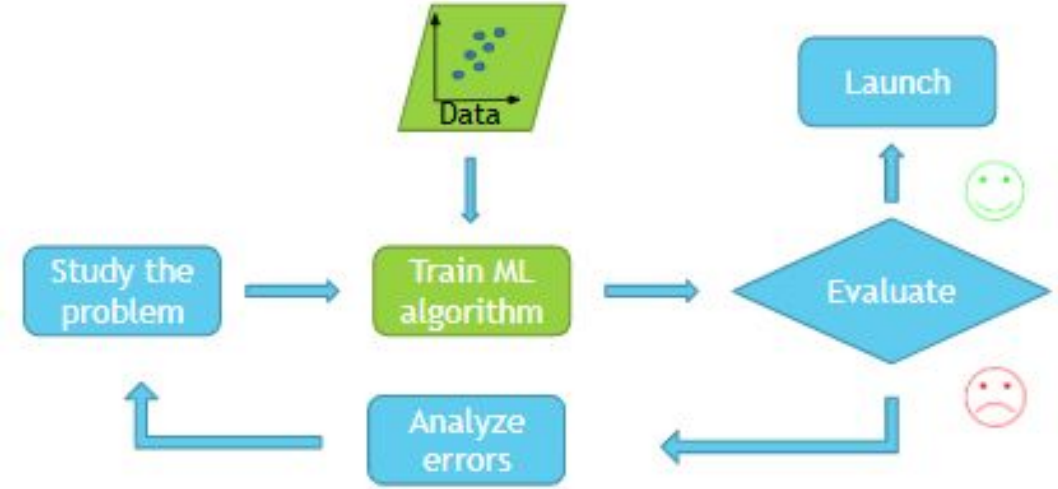
For U
credit card
free
amazing



The traditional approach

Why Machine Learning?

For U
credit card
free
amazing

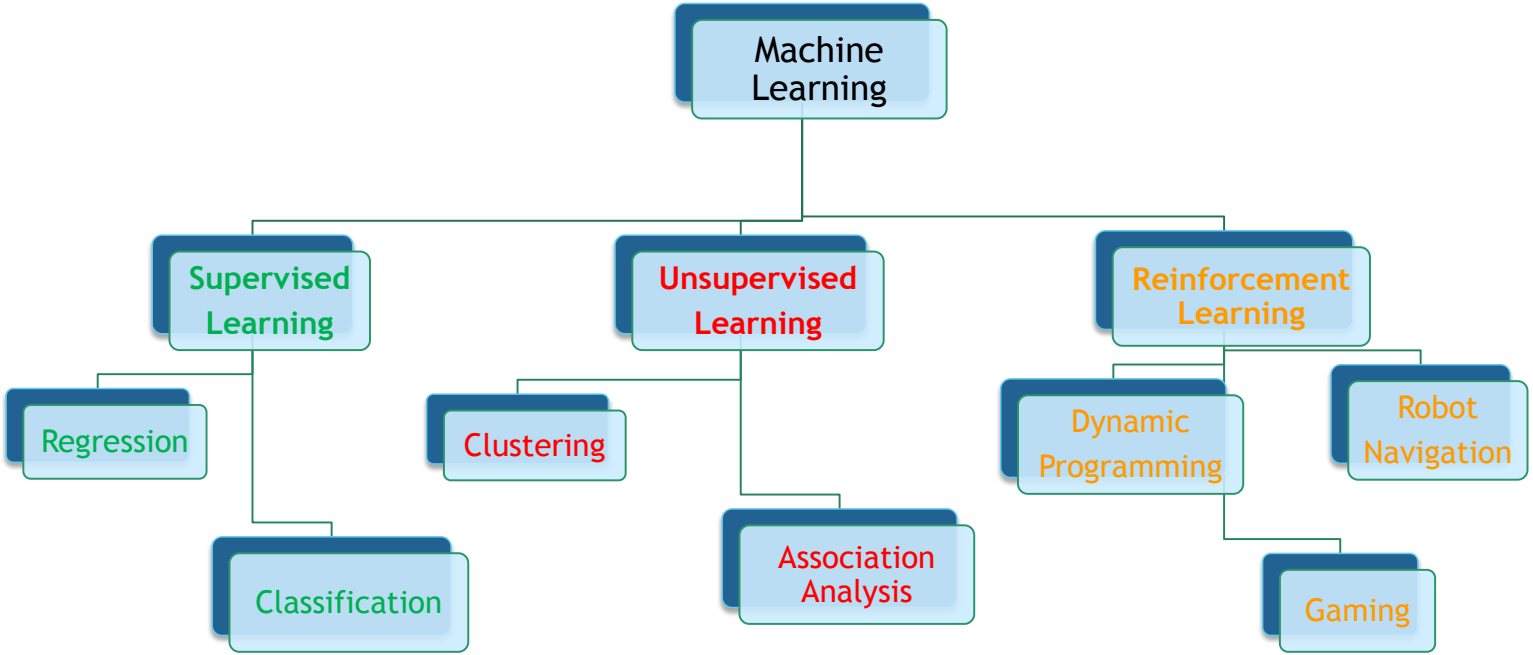


Machine Learning approach

Agenda

- What is ML? Why ML?
- **Categories of ML**
- Support Vector Machines
  - Basic idea
  - Special Cases
  - History and Development
  - Math

Categories of Machine Learning



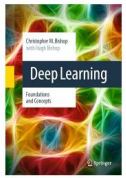
Categories of Machine Learning

- Machine Learning
  - Supervised Learning
    - Regression
    - Classification
  - Unsupervised Learning
    - Clustering
    - Association Analysis
  - Reinforcement Learning
    - Dynamic Programming
    - Gaming
    - Robot Navigation

Categories of Machine Learning

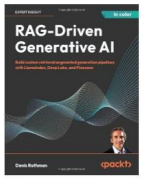
Products related to this item

Sponsored



Deep Learning: Foundations and Concepts
Christopher M. Bishop
★★★★½ 115
Hardcover
$81.10
prime

Just released



RAG-Driven Generative AI: Build custom retrieval augmented generation pipelines wit...
Denis Rothman
Minimize AI hallucinations and build accurate, custom generative AI pipelines with RAG using embedded vector databases and integrated

Just released

LLM Engineer's Handbook: Master the art of engineering large language models from...
Paul Iusztin
Discover practical insights into designing, training, and deploying LLMs in real-world scenarios by leveraging MLOps best

Machine Learning

Regression

Classification

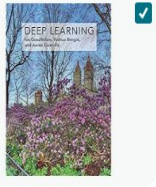
Clustering

Unsupervised Learning

Association Analysis

Gaming

Frequently bought together



+

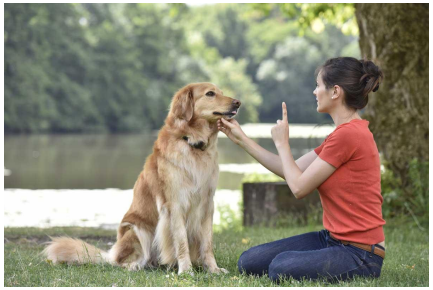
+

This item: Deep Learning (Adaptive Computation and Machine Learning series)
$60.00

Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow: Concepts, Tool...
$49.99

The Elements of Statistical Learning: Data Mining, Inference, and Prediction,...
$84.99

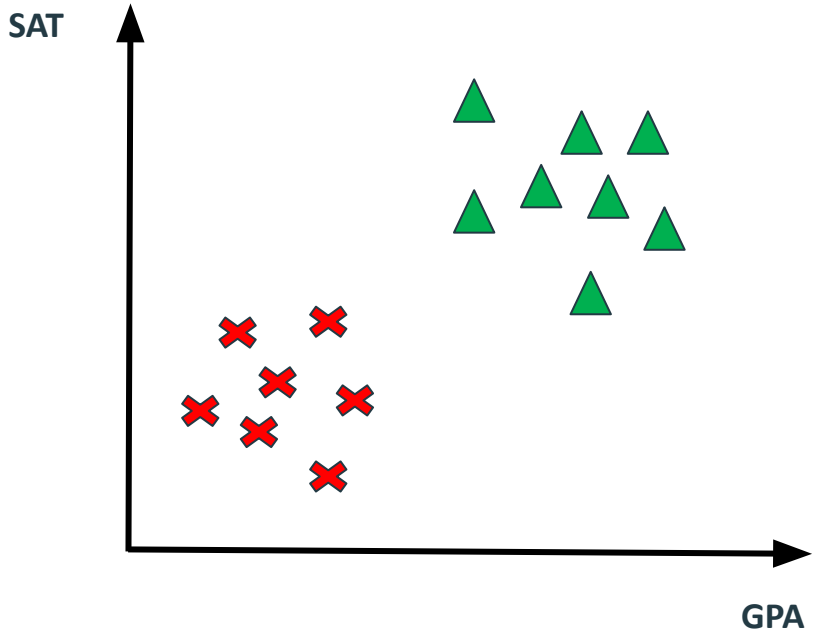
Categories of Machine Learning



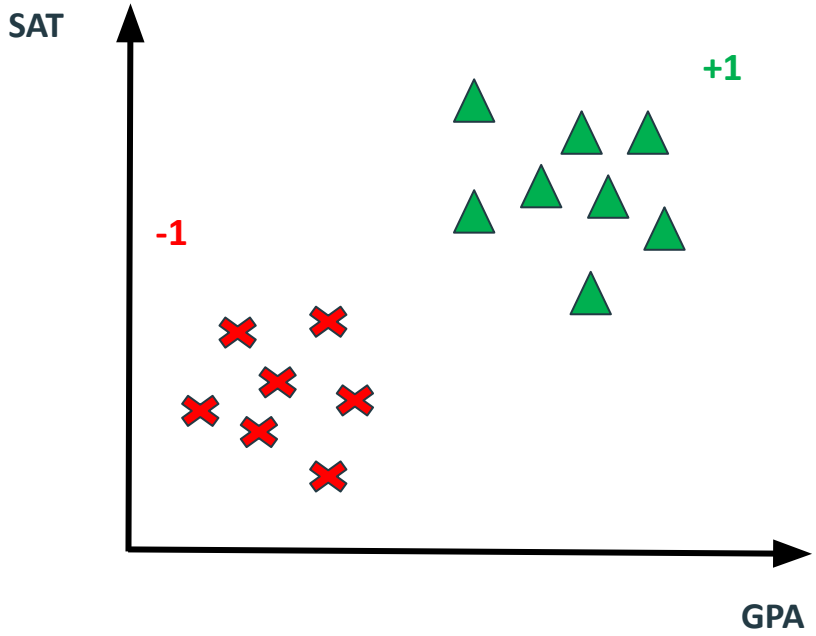
Agenda

- What is ML? Why ML?
- Categories of ML
- **Support Vector Machines**
  - **Basic idea**
  - Special Cases
  - History and Development
  - Math

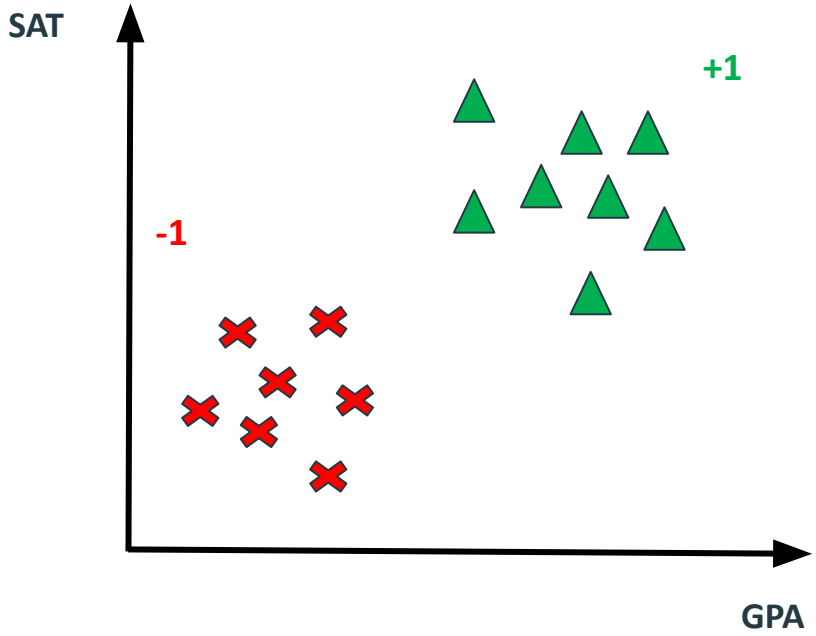
Support Vector Machines
The basic idea



Support Vector Machines
The basic idea

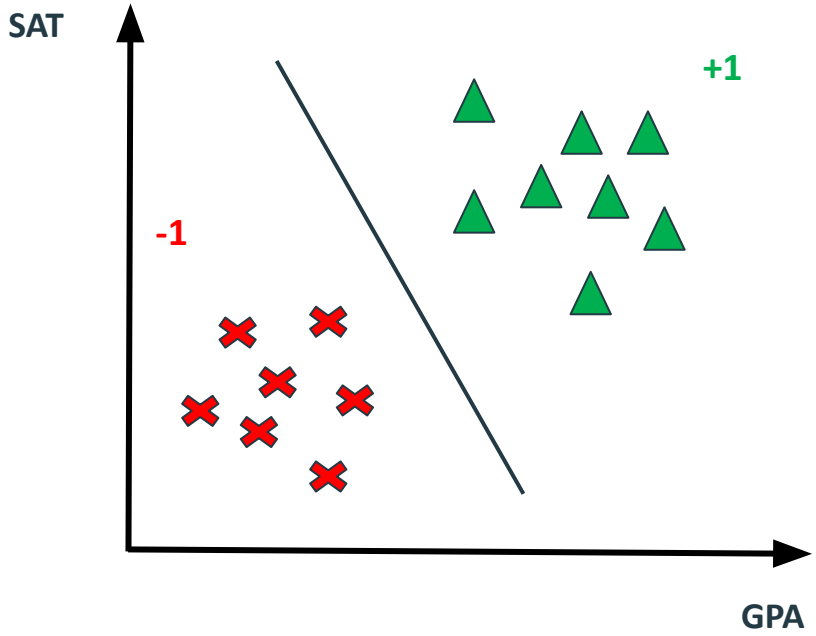


Support Vector Machines
The basic idea



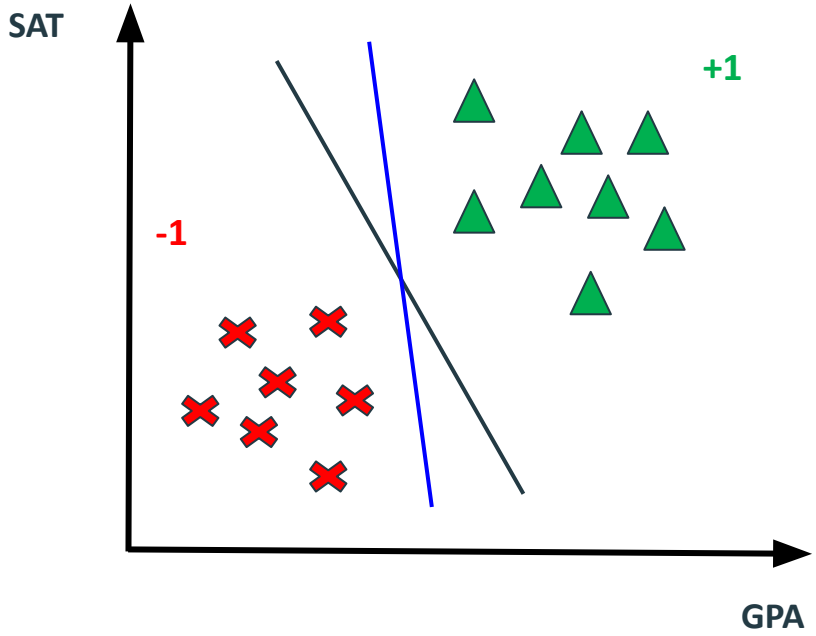
Can we find a straight line that separates the two classes?

Support Vector Machines
The basic idea



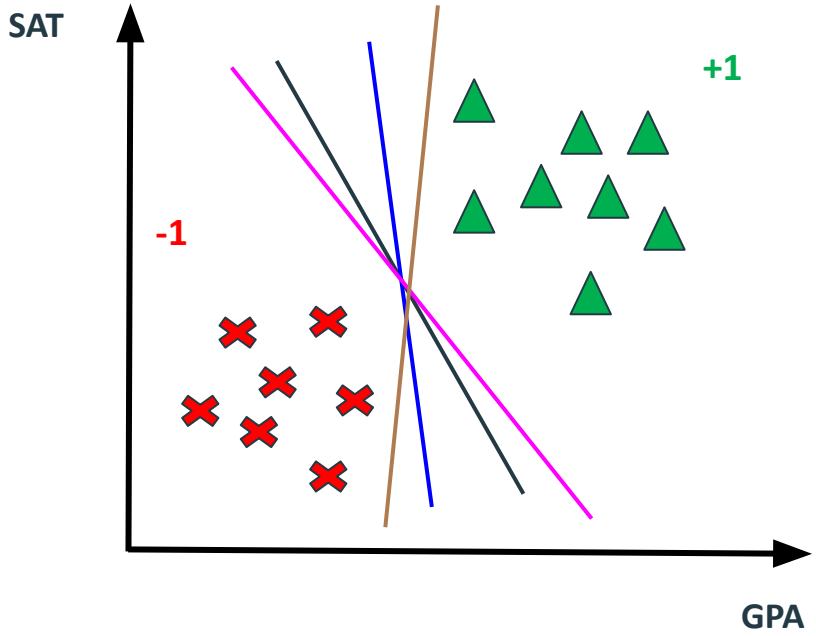
Can we find a straight line that separates the two classes?

Support Vector Machines
The basic idea



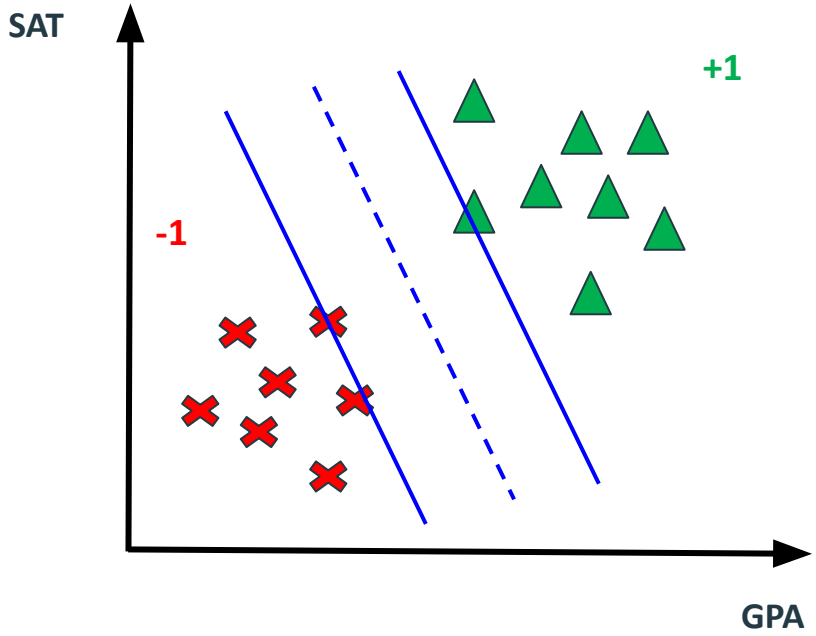
Can we find a straight line that separates the two classes?

Support Vector Machines
The basic idea



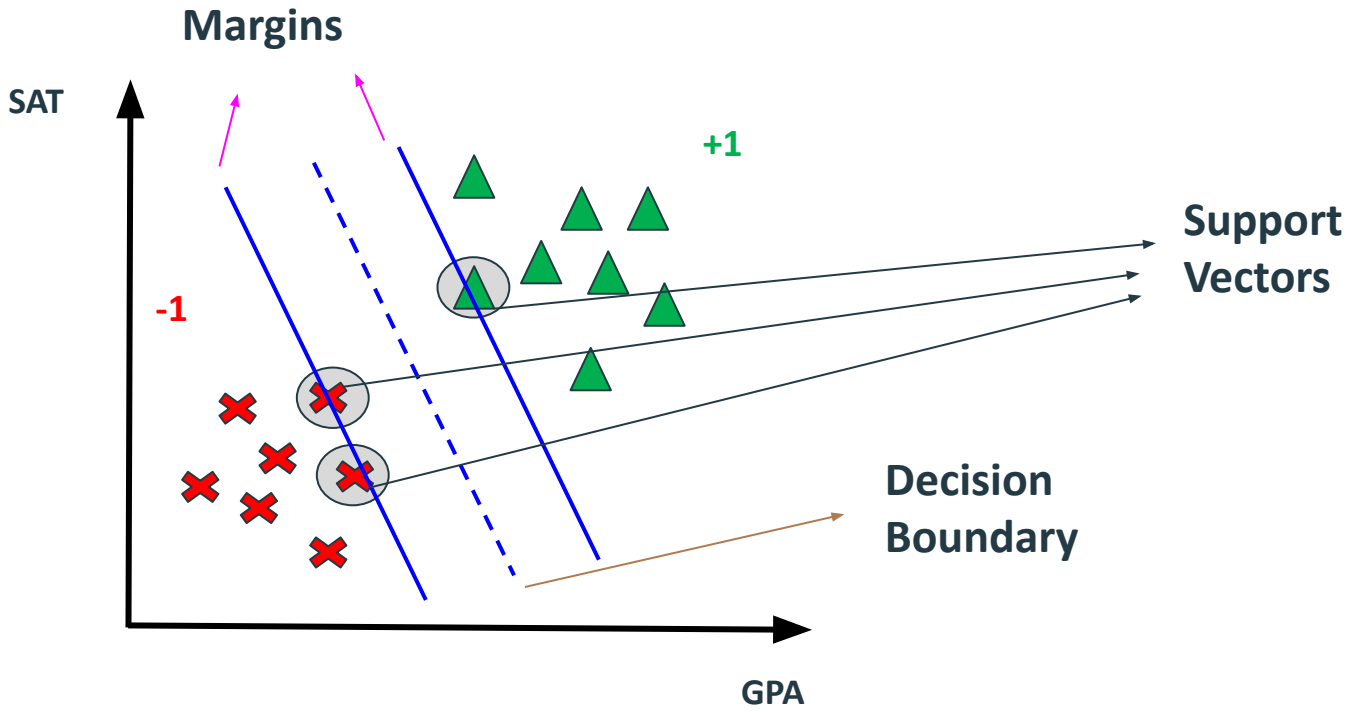
Can we find a straight line that separates the two classes?

Support Vector Machines
The basic idea



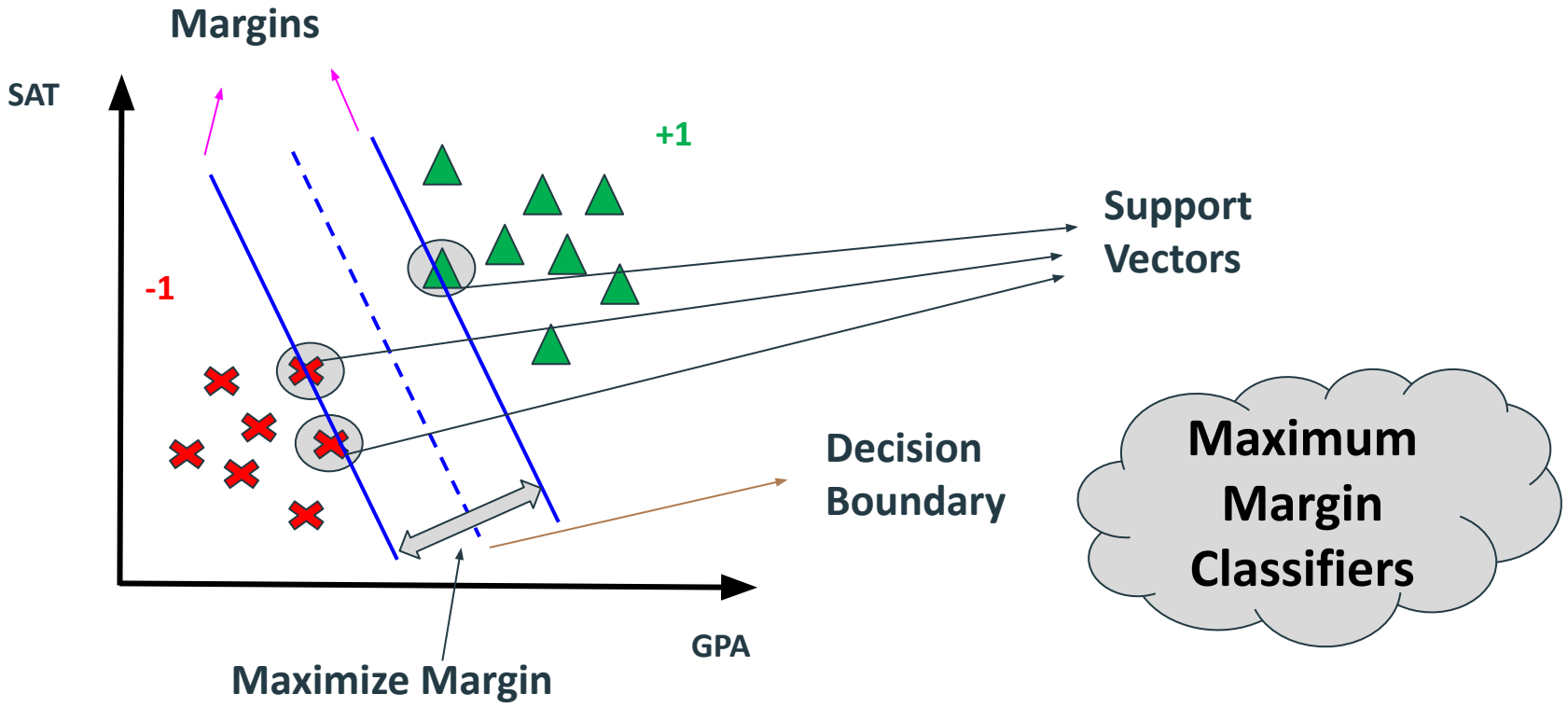
Support Vector Machines

The basic idea



Support Vector Machines

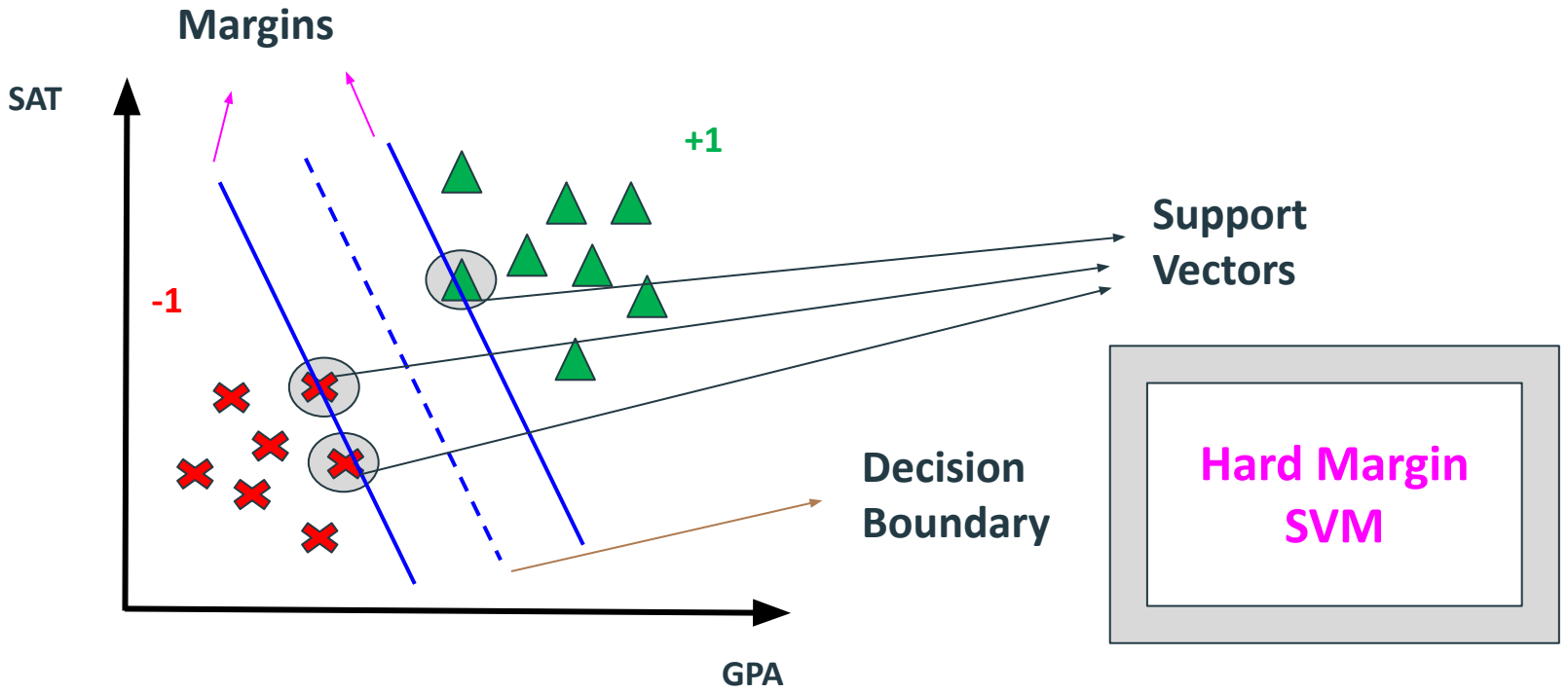
The basic idea



Agenda

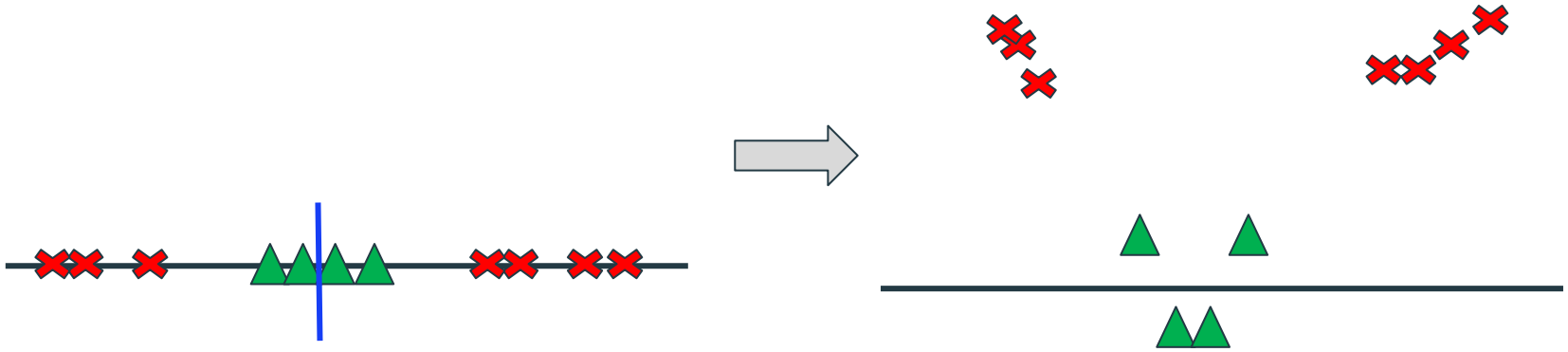
- What is ML? Why ML?
- Categories of ML
- **Support Vector Machines**
  - Basic idea
  - **Special Cases**
  - History and Development
  - Math

Support Vector Machines
Case 1: Linearly Separable Data

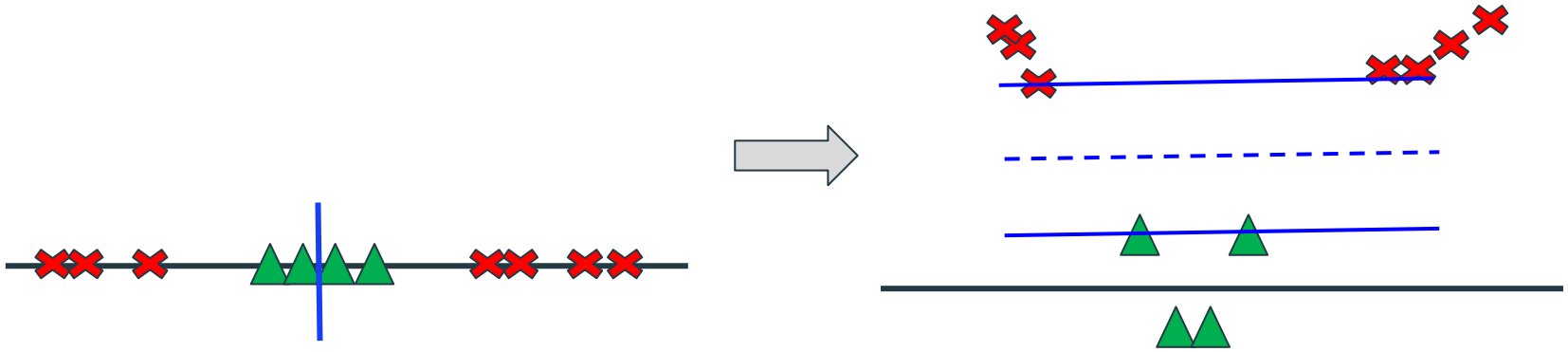


Support Vector Machines
Case 2a: Linearly Inseparable

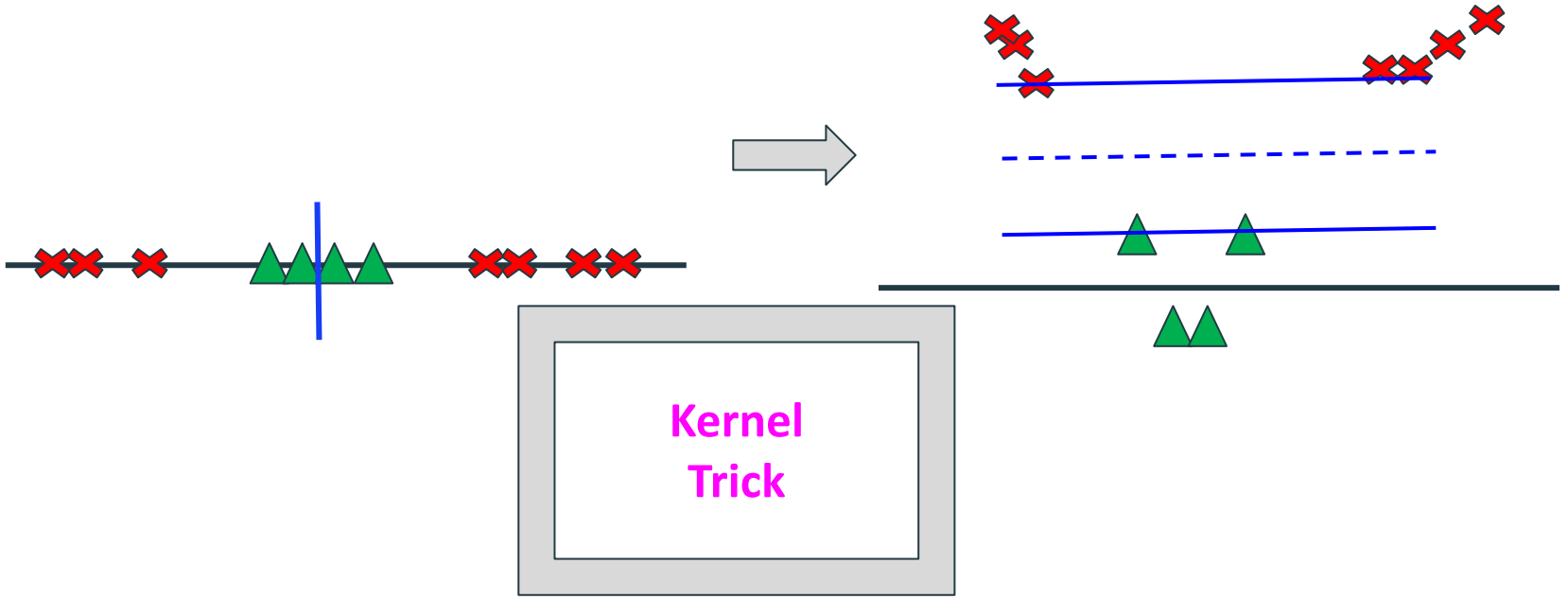
Support Vector Machines
Case 2a: Linearly Inseparable



Support Vector Machines
Case 2a: Linearly Inseparable

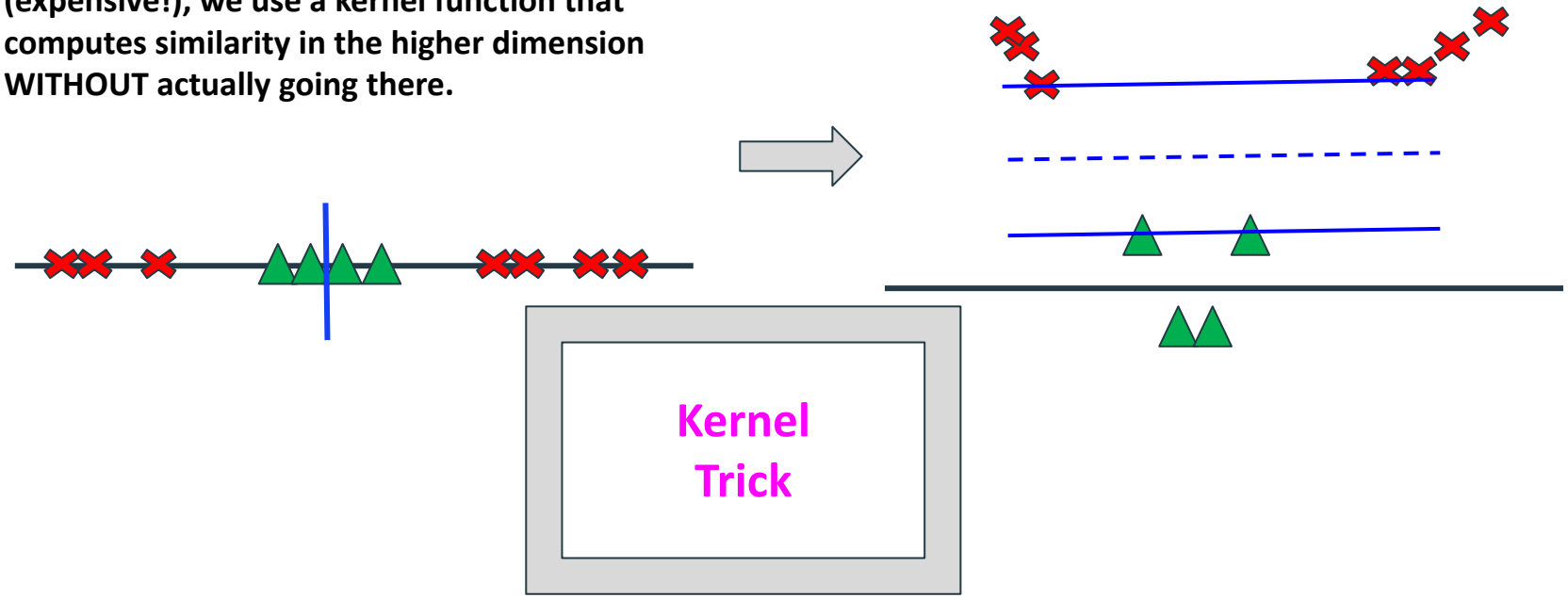


Support Vector Machines
Case 2a: Linearly Inseparable

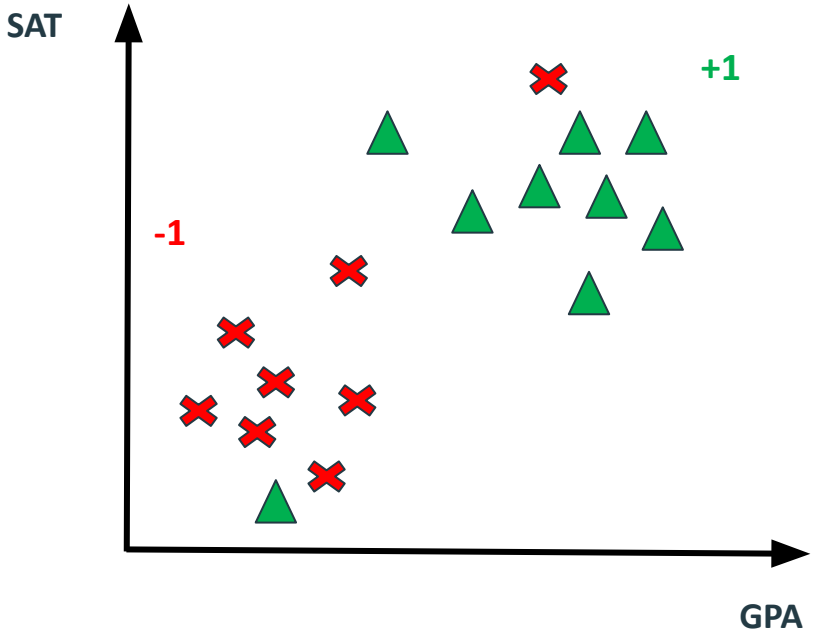


Support Vector Machines
Case 2a: Linearly Inseparable

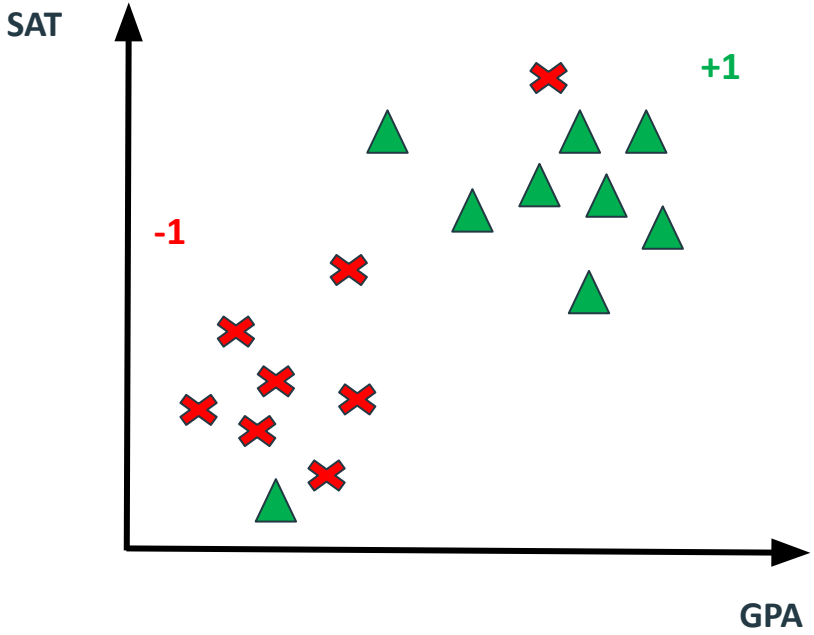
Instead of explicitly transforming data (expensive!), we use a kernel function that computes similarity in the higher dimension WITHOUT actually going there.

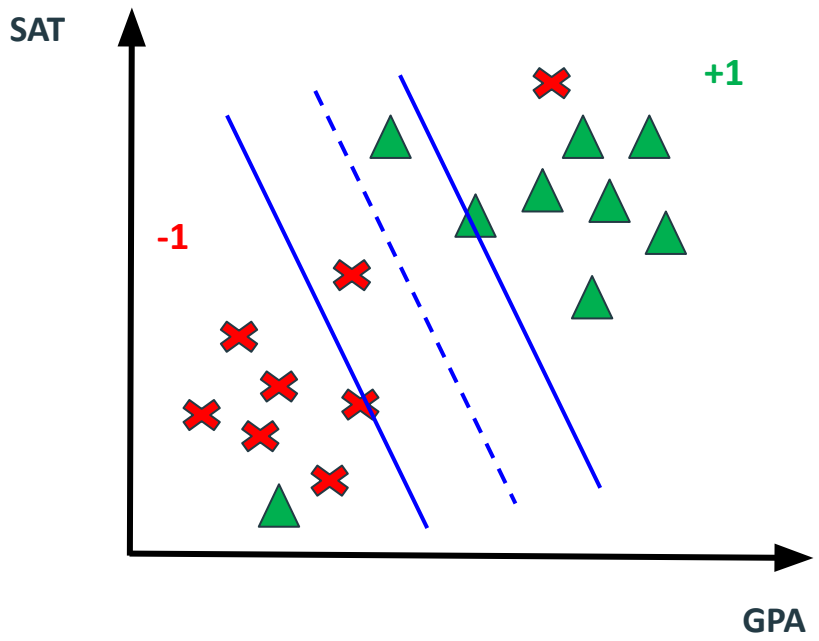


Support Vector Machines
Case 2b: Linearly Inseparable

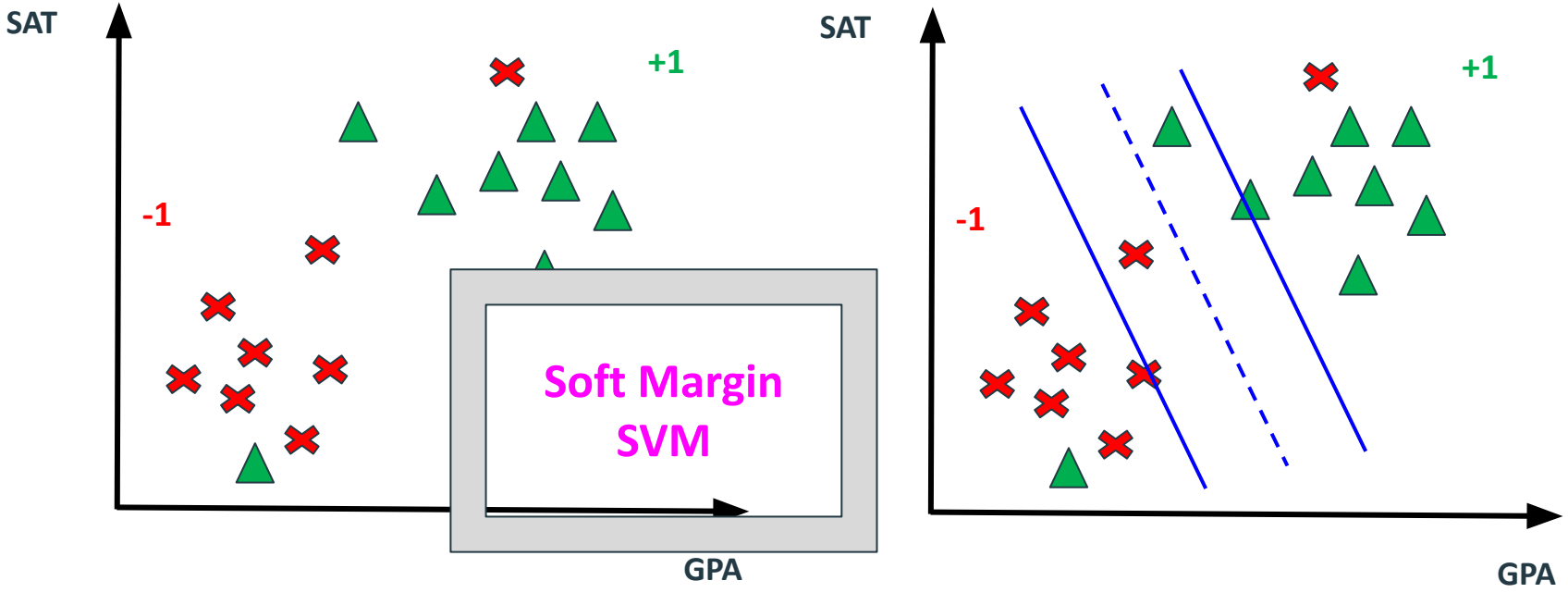


Support Vector Machines
Case 2b: Linearly Inseparable





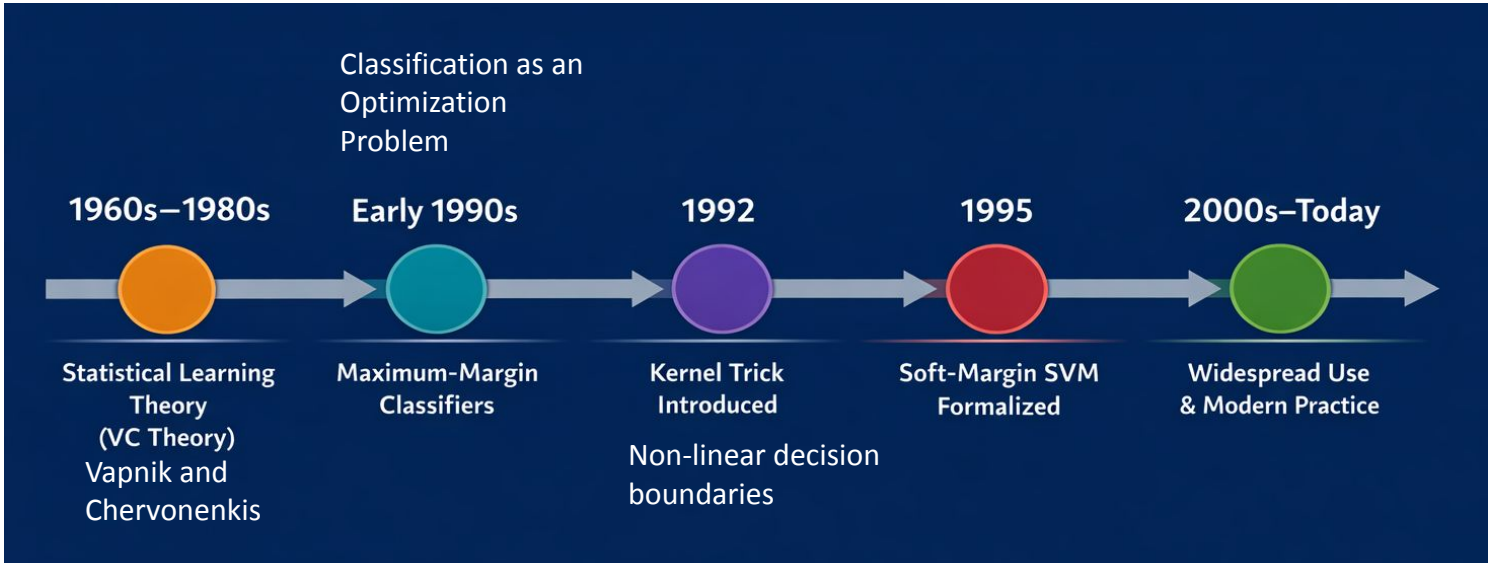
Support Vector Machines
Case 2b: Linearly Inseparable



Agenda

- What is ML? Why ML?
- Categories of ML
- **Support Vector Machines**
  - Basic idea
  - Special Cases
  - **History and Development**
  - Math

Support Vector Machines
The History and Development



Agenda

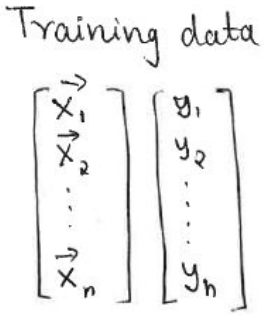
- What is ML? Why ML?
- Categories of ML
- **Support Vector Machines**
  - Basic idea
  - Special Cases
  - History and Development
  - **Math**

Hard Margin SVM
The Math

Problem Setup: Binary Classification
We want to classify data into two categories. For example, predicting whether a student gets into tech school (y = +1) or does not (y = -1).

The Data:

- n total students
- For each student i, we have:
  - A feature vector x_i (e.g., GPA, SAT scores, …)
  - A label $y_i \in \{-1, +1\}$

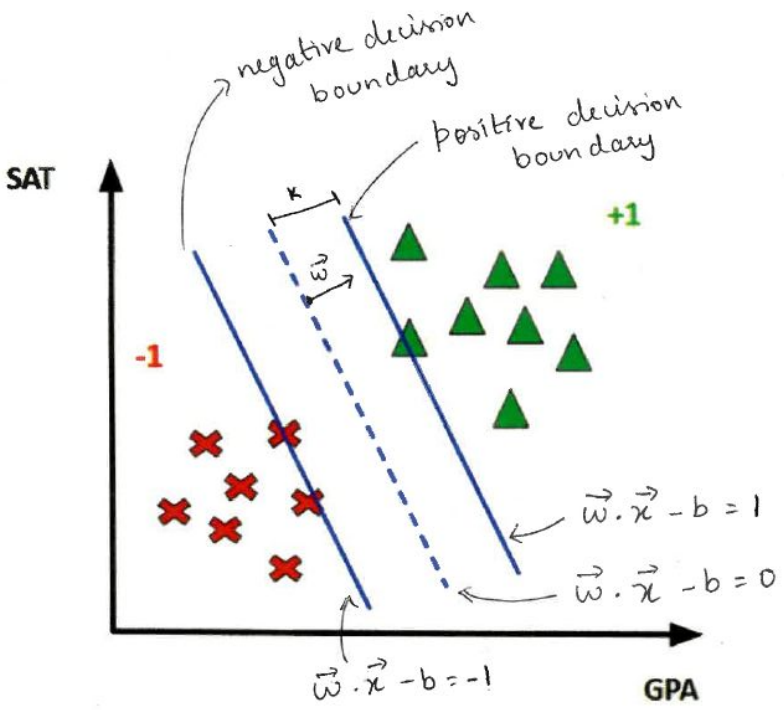


The Goal: Build an SVM model that creates a maximum margin classifier—a decision boundary with the largest possible "safety zone" between the two classes. This breathing room helps the model generalize better to new, unseen data.

Hard Margin SVM
The Math

To find the best separating boundary, we define three parallel hyperplanes

What is the distance between the two margin boundaries?



Hard Margin SVM
The Math

Let $\vec{x}_0$ be any point on the decision boundary

$$\Rightarrow \quad \vec{w} \cdot \vec{x}_0 - b = 0 \quad - (1)$$

We want to walk from $\vec{x}_0$ in the direction of w (the perpendicular direction) until we reach the positive margin boundary.

Lets walk a distance of k units in the direction of the unit vector, $\frac{w}{\|w\|}$

$\therefore$ the new position after walking k units:

$$\vec{x}_1 = \vec{x}_0 + k \frac{w}{\|w\|} \quad - (2)$$

Since x_1 is on the positive margin boundary,

$$\vec{w} \cdot \vec{x}_1 - b = 1 \quad - (3)$$

substitute (2) in (3)

$$\vec{w} \cdot \left(\vec{x}_0 + k \frac{w}{\|w\|} \right) - b = 1$$

$$\vec{w} \cdot \vec{x}_0 + k \frac{w \cdot w}{\|w\|} - b = 1$$

$$\left(\vec{w} \cdot \vec{x}_0 - b \right) + k \frac{w \cdot w}{\|w\|} = 1 \quad - (4)$$

substitute (1) in (4), $w \cdot w = \|w\|^2$

$$0 + k \frac{\|w\|^2}{\|w\|} = 1$$

$$\Rightarrow \boxed{k = \frac{1}{\|\vec{w}\|}}$$

$\therefore$ The margin size is $\frac{2}{\|\vec{w}\|}$

So, we have to maximize the margin

$\Rightarrow$ minimizing $\|\vec{w}\|$

Hard Margin SVM
The Math

Constraints:

1. We need every data point be correctly classified
2. Every data point be outside or on the margin

ie., for positive classes,

$$y_i = +1 \quad , \quad \vec{w} \cdot \vec{x}_i - b \geq +1$$

for negative classes

$$y_i = -1 \quad , \quad \vec{w} \cdot \vec{x}_i - b \leq -1$$

we combine both constraints by multiplying y_i

$$y_i (\vec{w} \cdot \vec{x}_i - b) \geq 1$$

$\therefore$ minimize $||w||$

subject to $y_i (\vec{w} \cdot \vec{x}_i - b) \geq 1 \quad \forall \; i = 1, 2 \ldots n$

GitHub

LinkedIn

Exercise 1: Match the Following

Match each scenario to the appropriate approach:

Scenario	Approach
Calculate employee salary based on hours worked and fixed hourly rate	
Predict customer churn based on past behavior patterns	

Options: Traditional Programming, Machine Learning

Exercise 2: Match the Following

Match each scenario to the correct ML type:

Scenario	ML Type
Determining if a tumor is benign or malignant based on medical scans	
Finding groups of similar customers for targeted marketing campaigns	
Predicting the sale price of a used car based on mileage, age, and condition	
Discovering that customers who buy diapers often buy baby formula	
Teaching a robot to navigate a maze by rewarding it for reaching the exit	

Options: Regression, Classification, Clustering, Association Analysis, Reinforcement Learning

Exercise 3: True or False

Statement	T/F
A. Support vectors are all the data points in the dataset	
B. A larger margin generally leads to better generalization	
C. The kernel trick explicitly computes coordinates in higher dimensions	
D. Soft margin SVM is more practical for real-world data than hard margin	