Support Vector Machines

Algorithm

- Trade off

Example

Mass

Lets starts by imagining we measured mass of people of employees who are obese

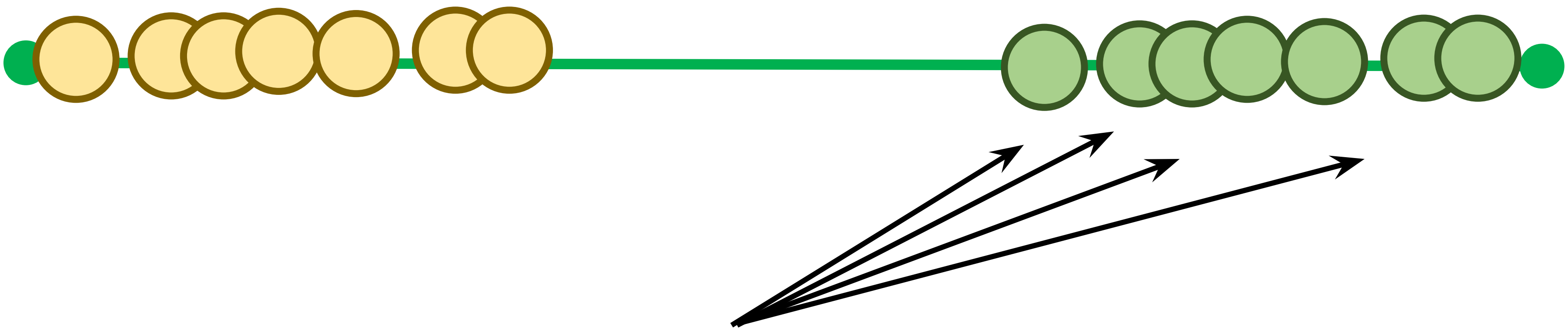
Example

Age

Yellow represent people who are not obese

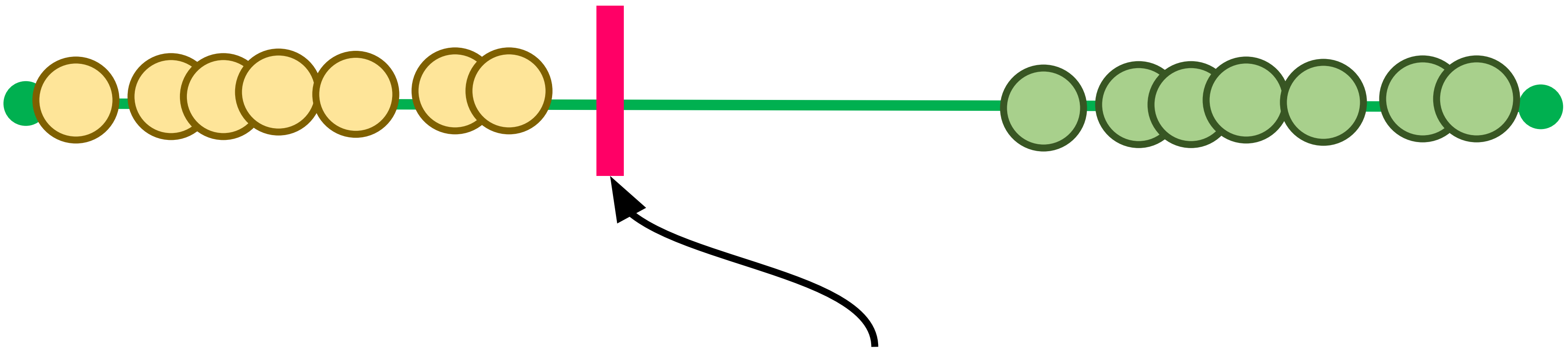
Example

Age



Green represent people who are obese

Example

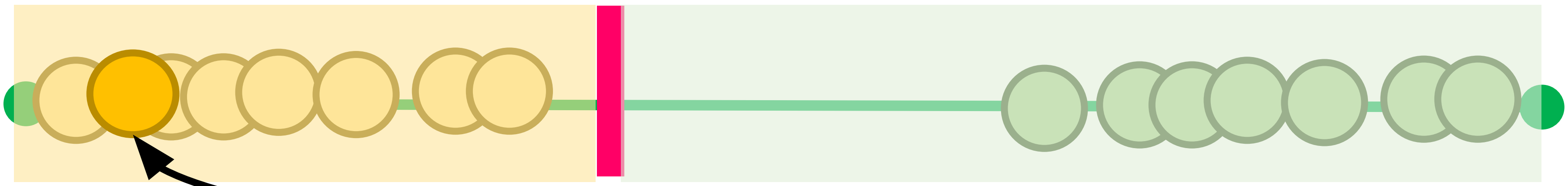


- Based on the observation we can pick a threshold

Case-1

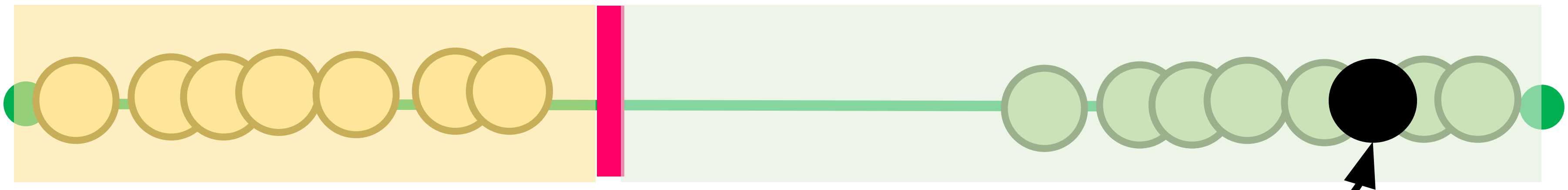
- For a new observation on left side of the threshold point ?

Case-1



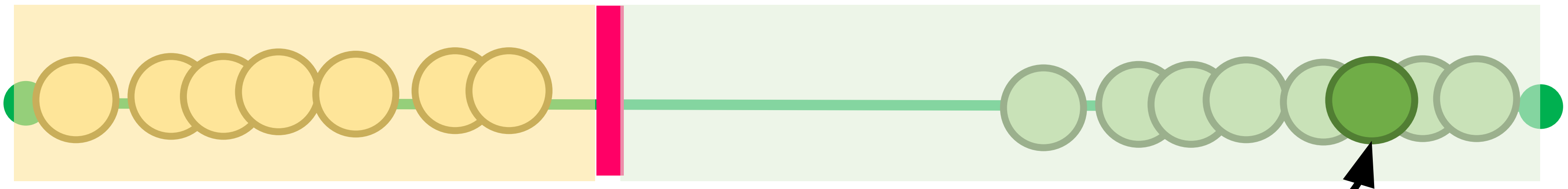
- We can classify it as not obese

Case-2



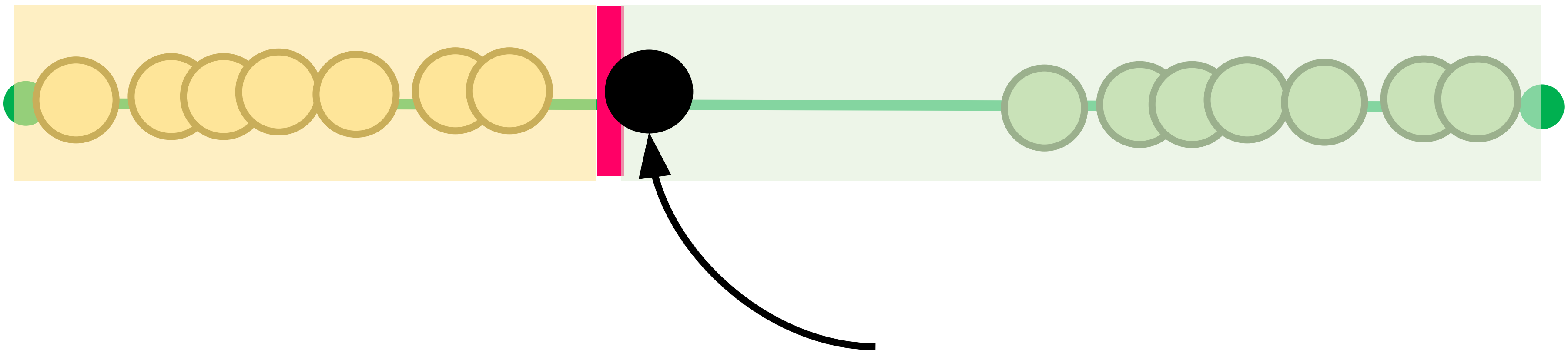
- For a new observation on right side or more mass than the threshold point ?

Case-2



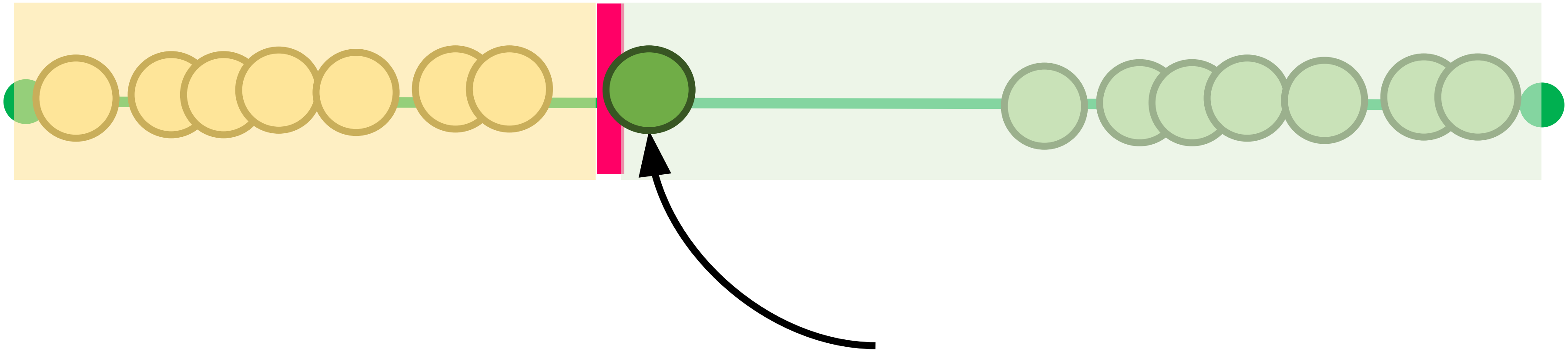
- We can classify as “obese”

Case-3



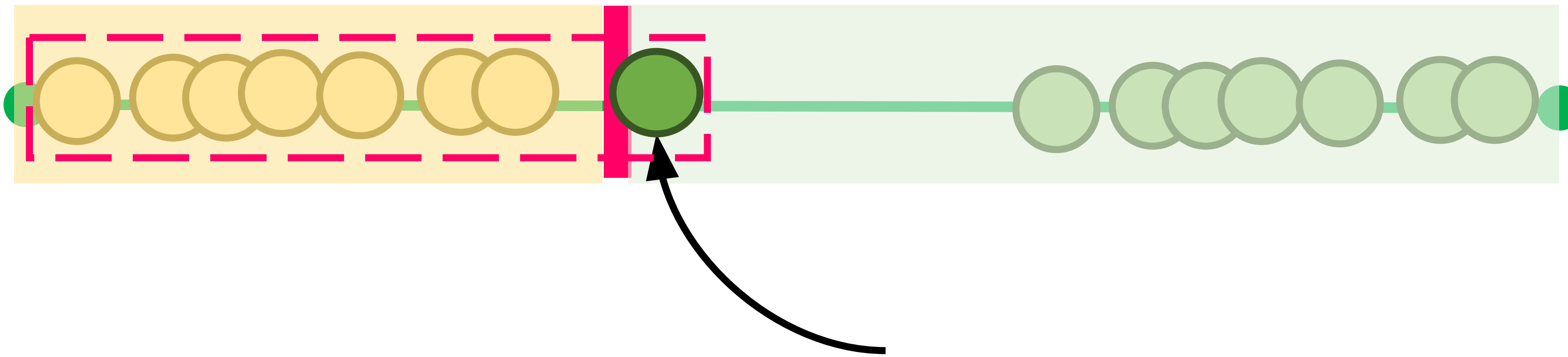
- What if let's say new mass came just above the threshold mass ??

Case-3



• It is classify as “obese”

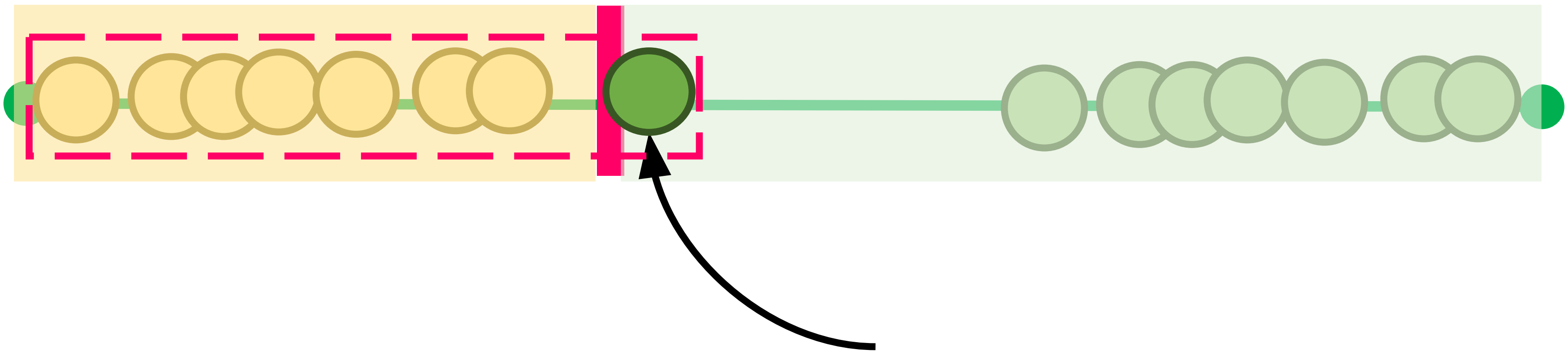
Case-3



- It is classify as “obese”

- It doesn’t make sense, because it is much closer to the observation that are **not obese**

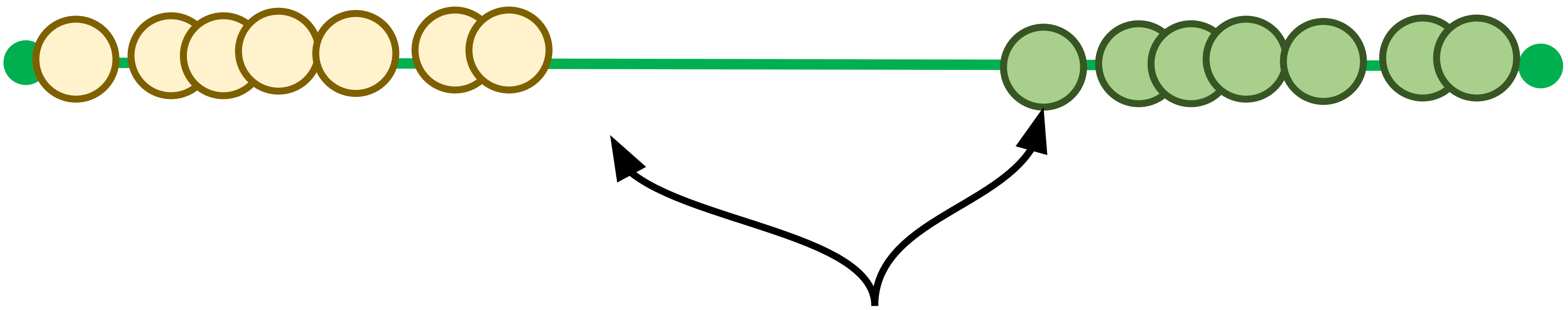
Case-3



- So, this threshold is lame

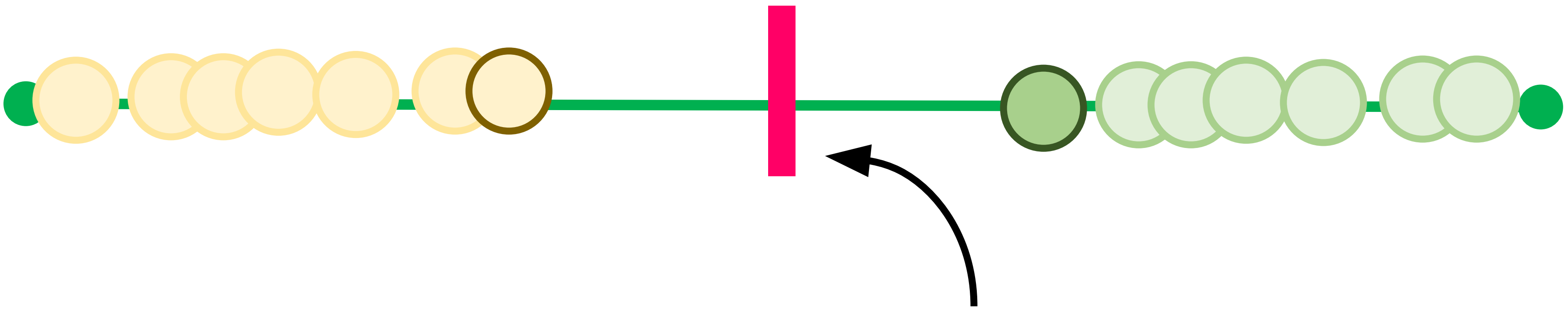
Can we do better ??

Remedy



We will focus on the observations on the edges of each cluster

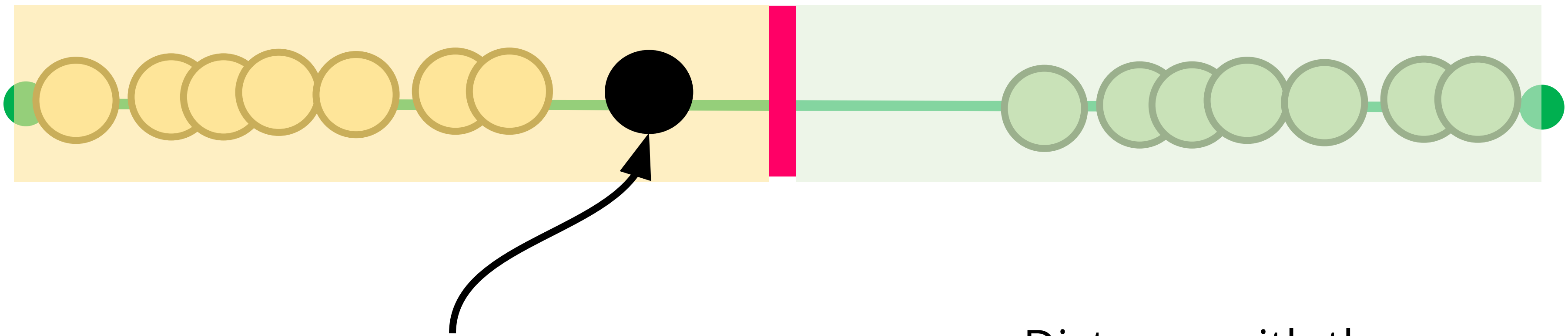
Remedy



Use mid point between them as threshold

✔ select a threshold such a way that it should have equidistant from the vector to allow maximum margin

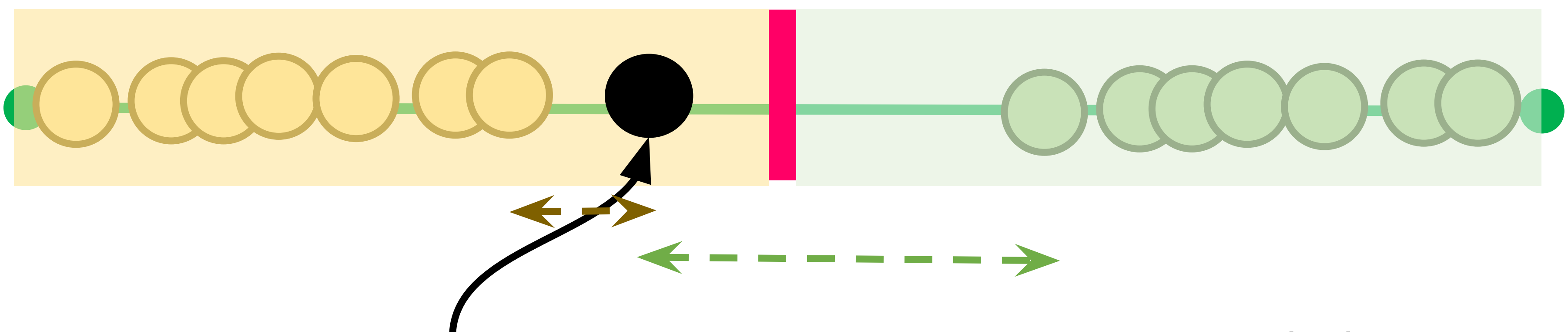
Case-3



- What if let's say new mass came just above the threshold mass??

- Distance with the edge of not obese is less

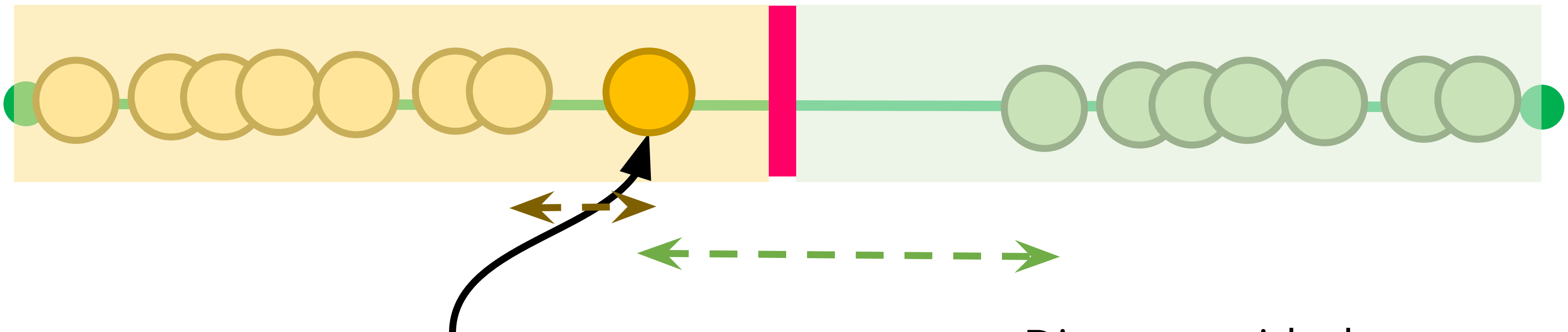
Case-3



- What if let's say new mass came just above the threshold mass ??

- Distance with the edge of not obese is less than the distance for having a obese hence it will classify as "not obese"

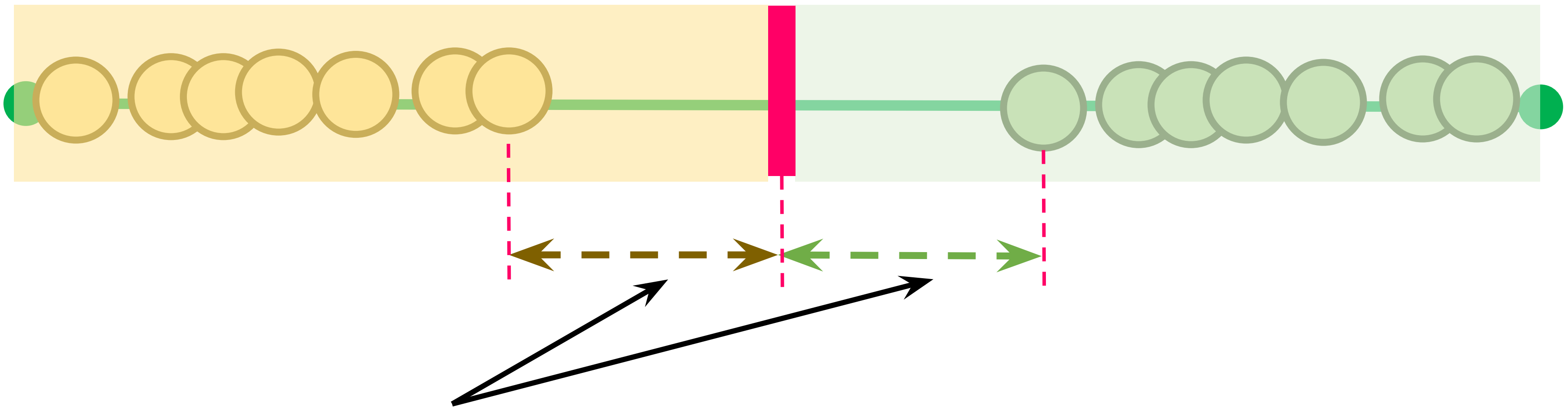
Case-3



- Classify as “not obese”

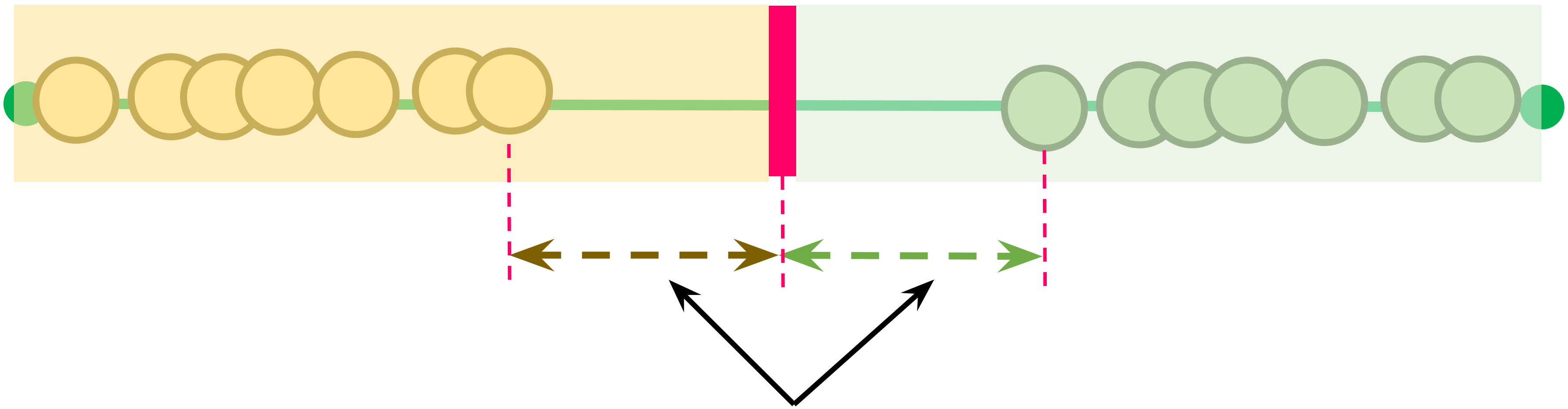
- Distance with the edge of not obese is less than the distance for having obese hence it will classify as “not obese”

Terminology



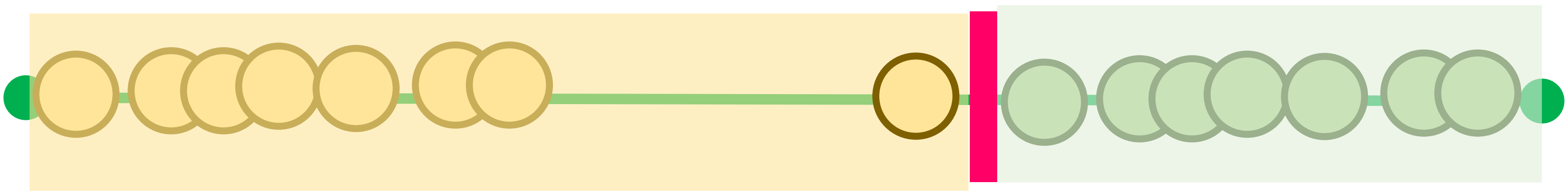
- The shortest distance between the observations and threshold is called **margin**

Terminology



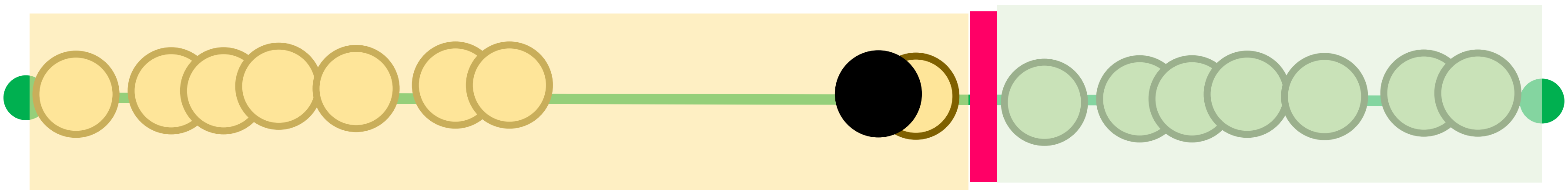
- When we are using the threshold that gives us the largest margin to make classification is called **Maximum Margin Classifier**

Case-4



- **Maximum Margin Classifier**, create a margin between two closest points

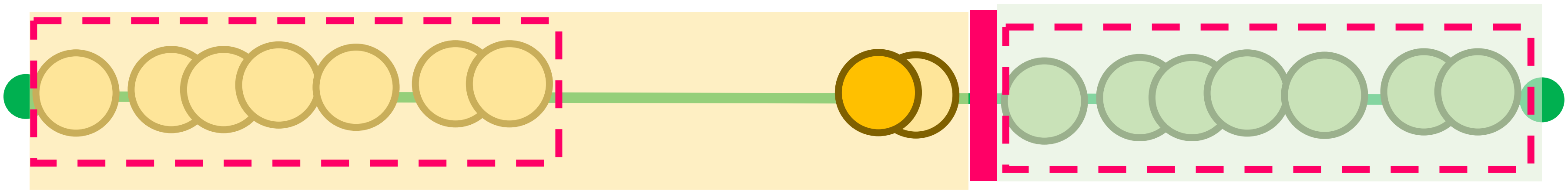
Case-4



- Maximum Margin Classifier classifies this points a “not obese”

❖ Note: most of the green points are closes to the predictor.

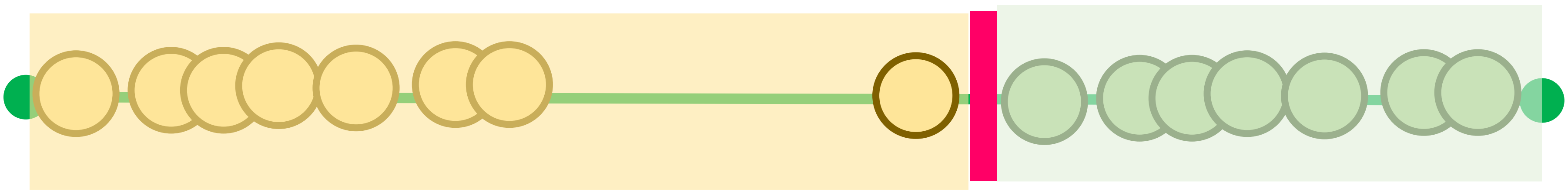
Case-4



- Maximum Margin Classifier classifies this points a “not obese”

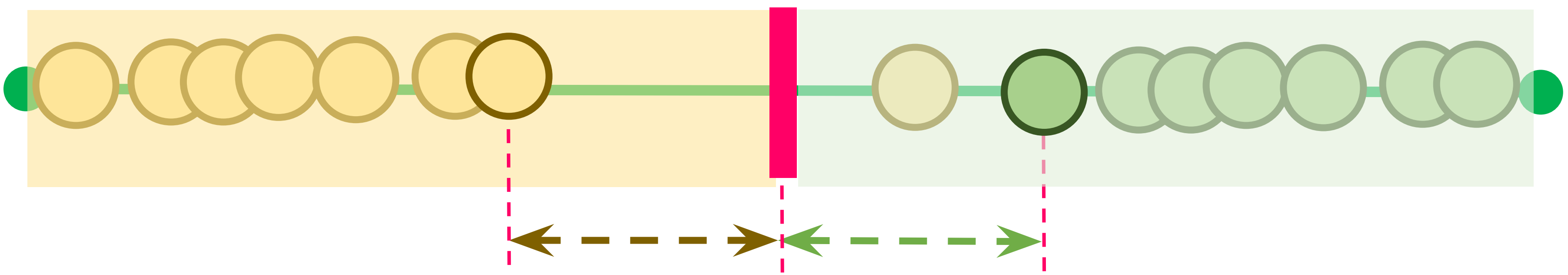
❖ Note: most of the green points are closes to the predictor.

Case-4 - Remedy



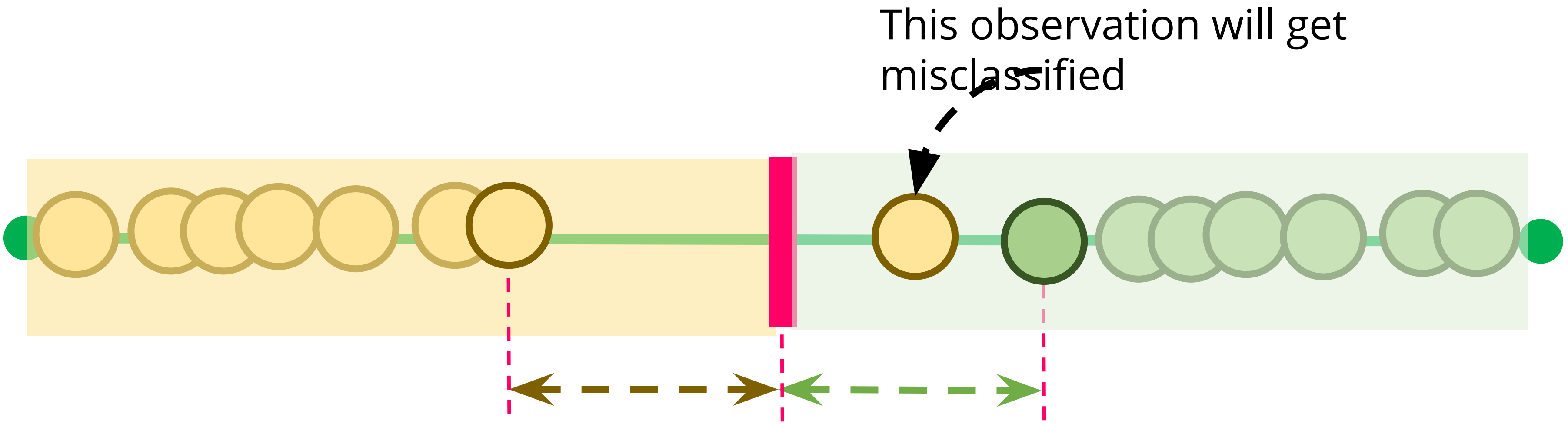
- **Threshold should not so sensitive to outlier**
  - ✔ **Allow misclassifications.**

Case-4 - Solution



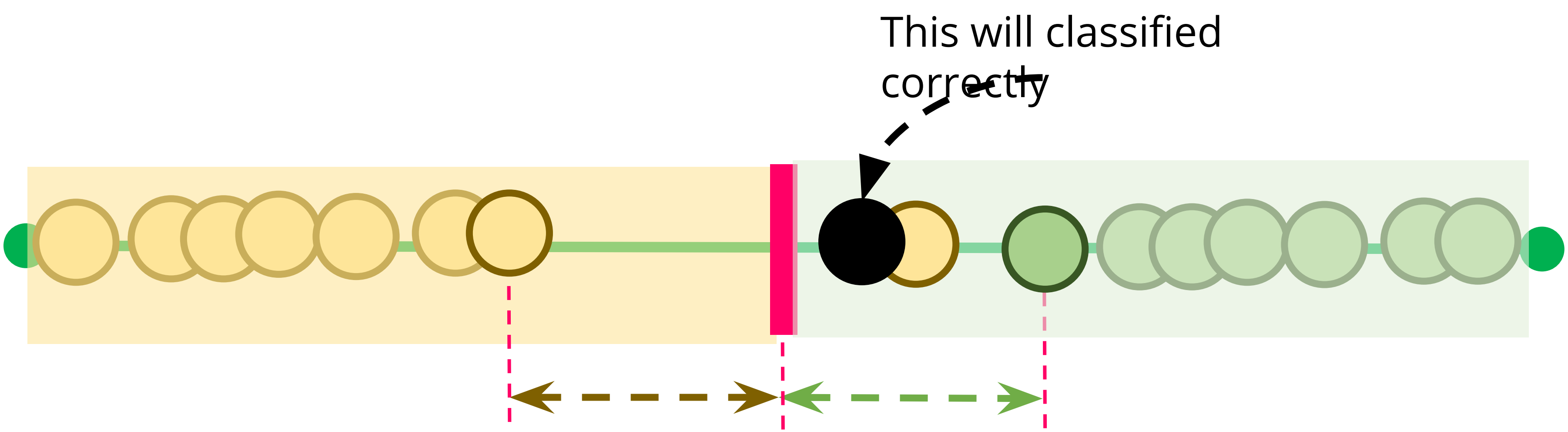
- If you put threshold halfway between these two observations

Case-4 - Solution



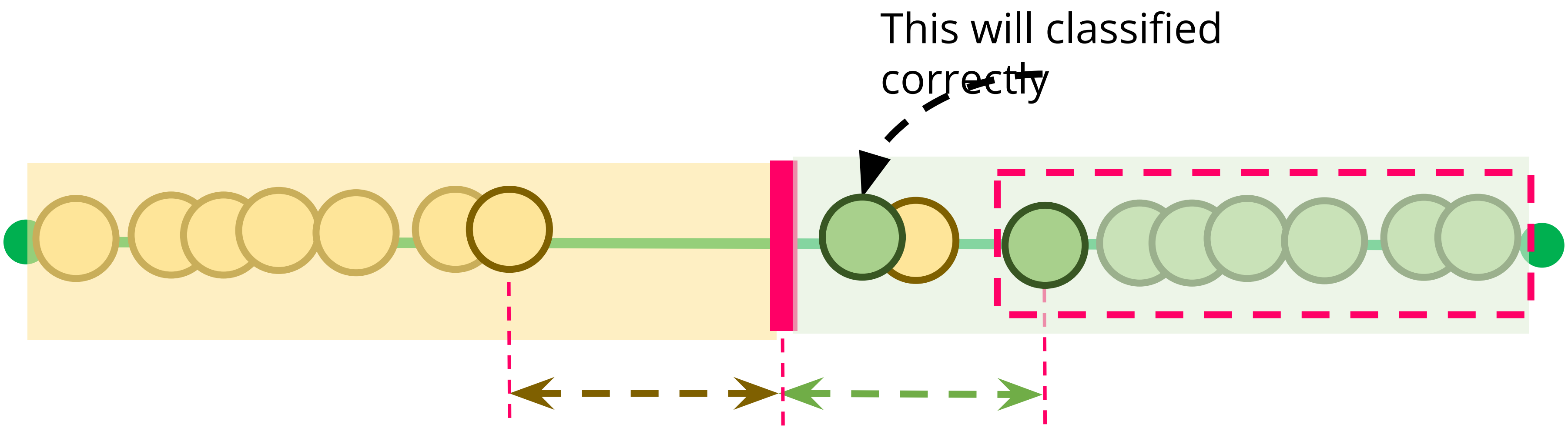
- Few observation might get misclassified

Case-4 - Solution



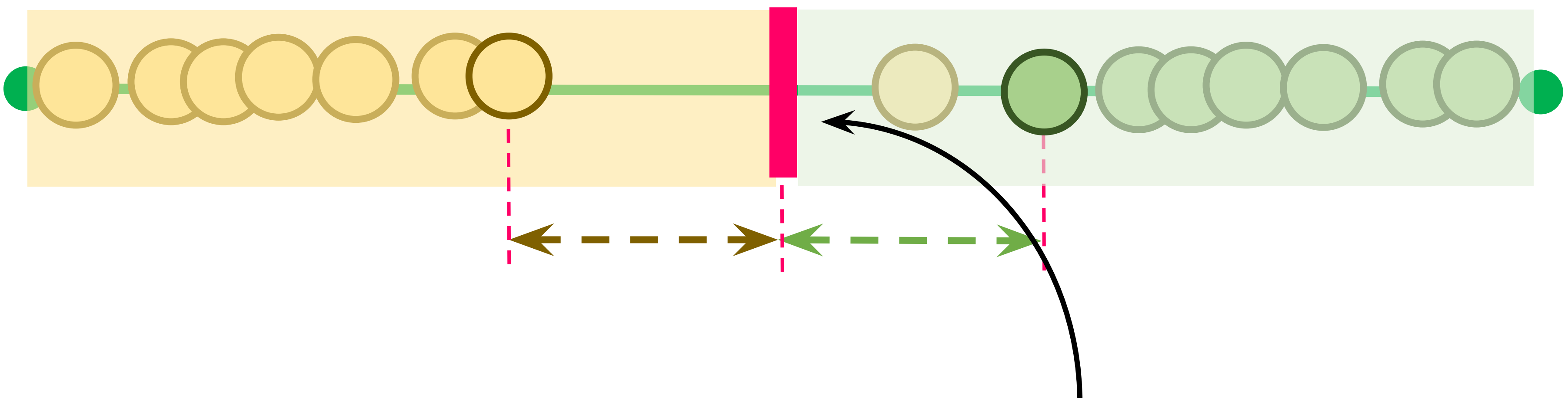
- For new observation it will classify correctly

Case-4 - Solution



- For new observation it will classify correctly

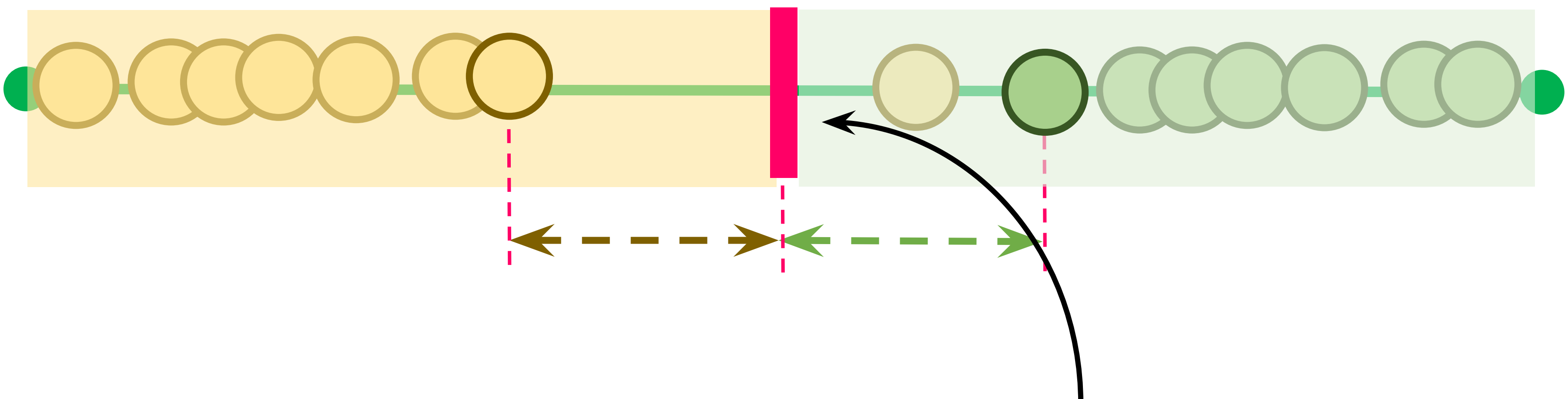
Case-4 - Solution



- Choosing a threshold that allows misclassification is an example of Bias/Variance Tradeoff

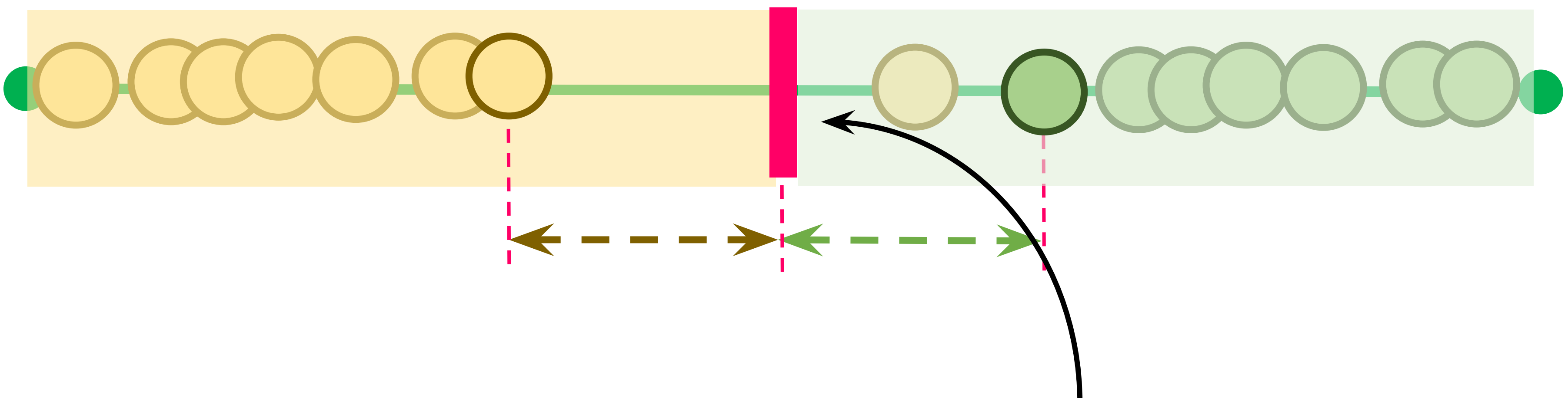
Soft Margin

Soft Margin



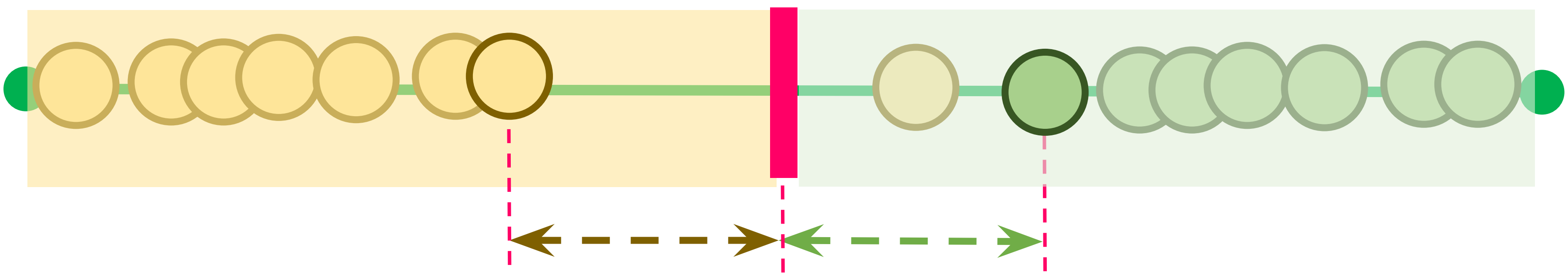
- When we allow misclassification, the distance between the observations and the threshold is called **Soft Margin**

Soft Margin

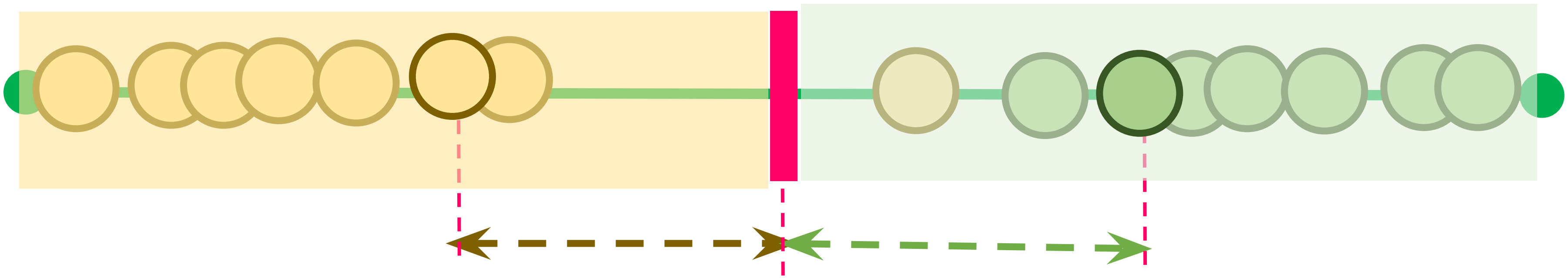


- When we allow misclassification, the distance between the observations and the threshold is called **Soft Margin**

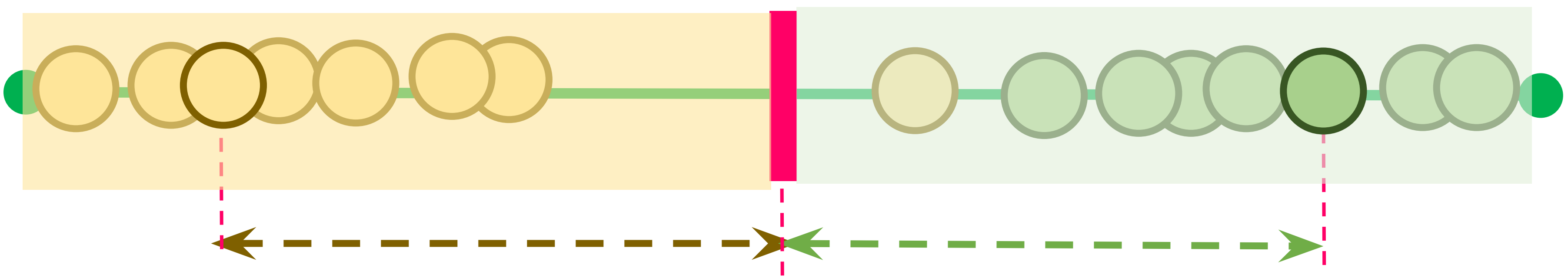
How to determine Soft Margin



How to determine Soft Margin



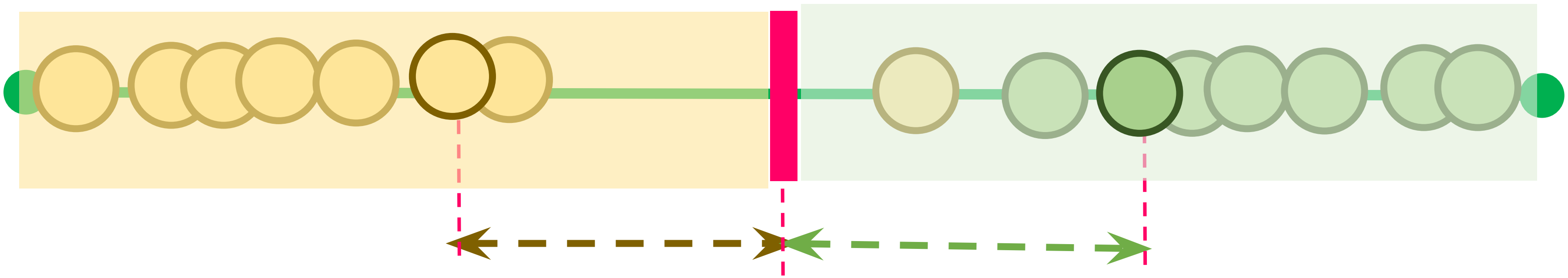
How to determine Soft Margin



How to determine Soft Margin

- We use Cross Validation to determine how many misclassifications and observations to allow inside of the Soft Margin to get the best classification.

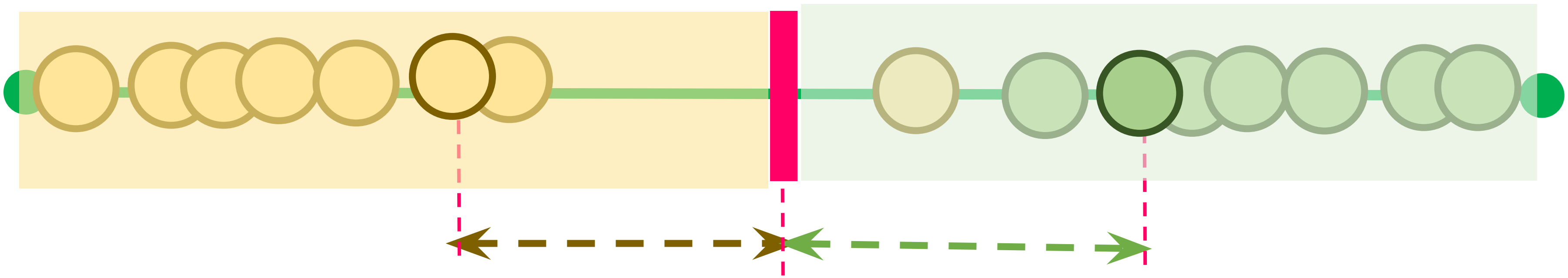
Cross Validation for Soft Margin



- If Cross Validation determined that it was the best Soft Margin.
  - There will be one misclassified
  - Two observation will be correctly classified to be within the **soft margin.**

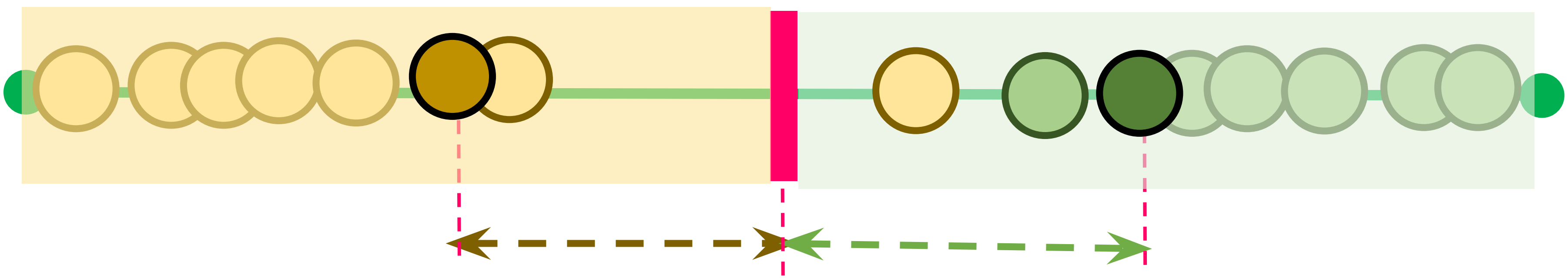
Soft Margin Classifier
or
Support Vector Classifier

Soft Margin Classifier



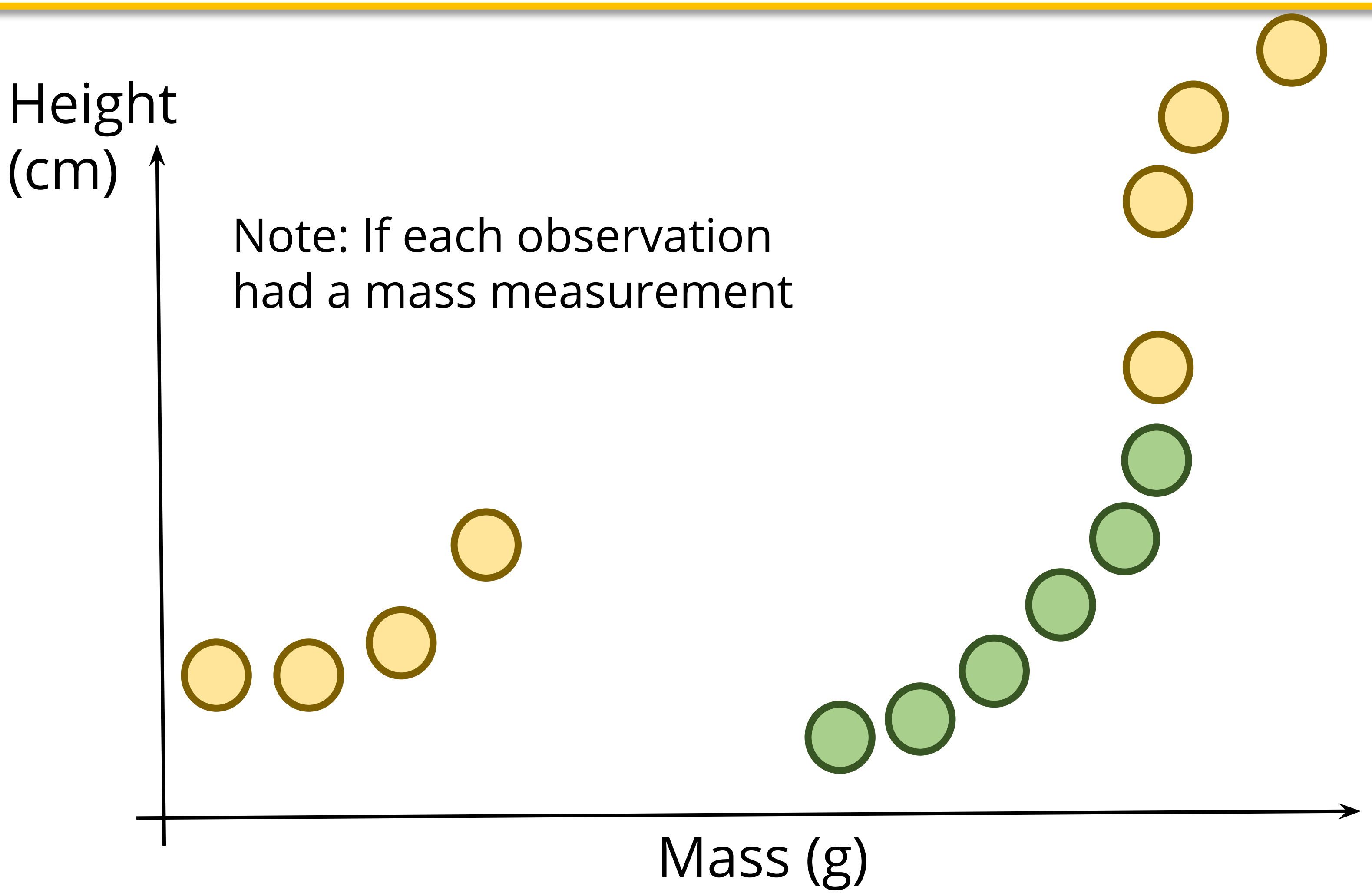
- When we use a Soft Margin to determine the location of a threshold which is called Soft Margin Classifier also known as Support Vector Classifier to classify observation.

Soft Margin Classifier

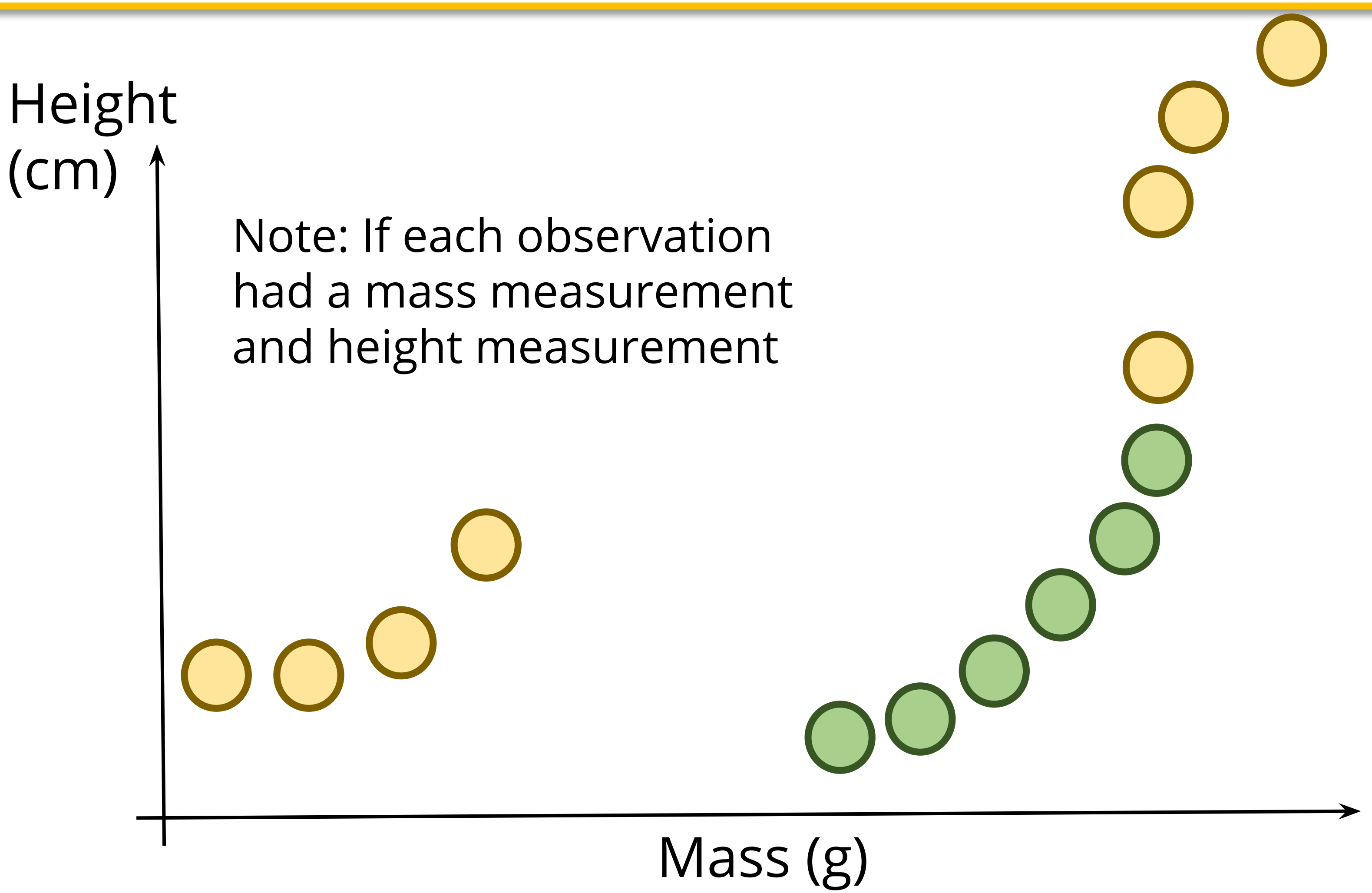


- The name Support Vector Classifier comes from the fact that the observations on the edge and within the **Soft Margin** are called **Support Vectors**

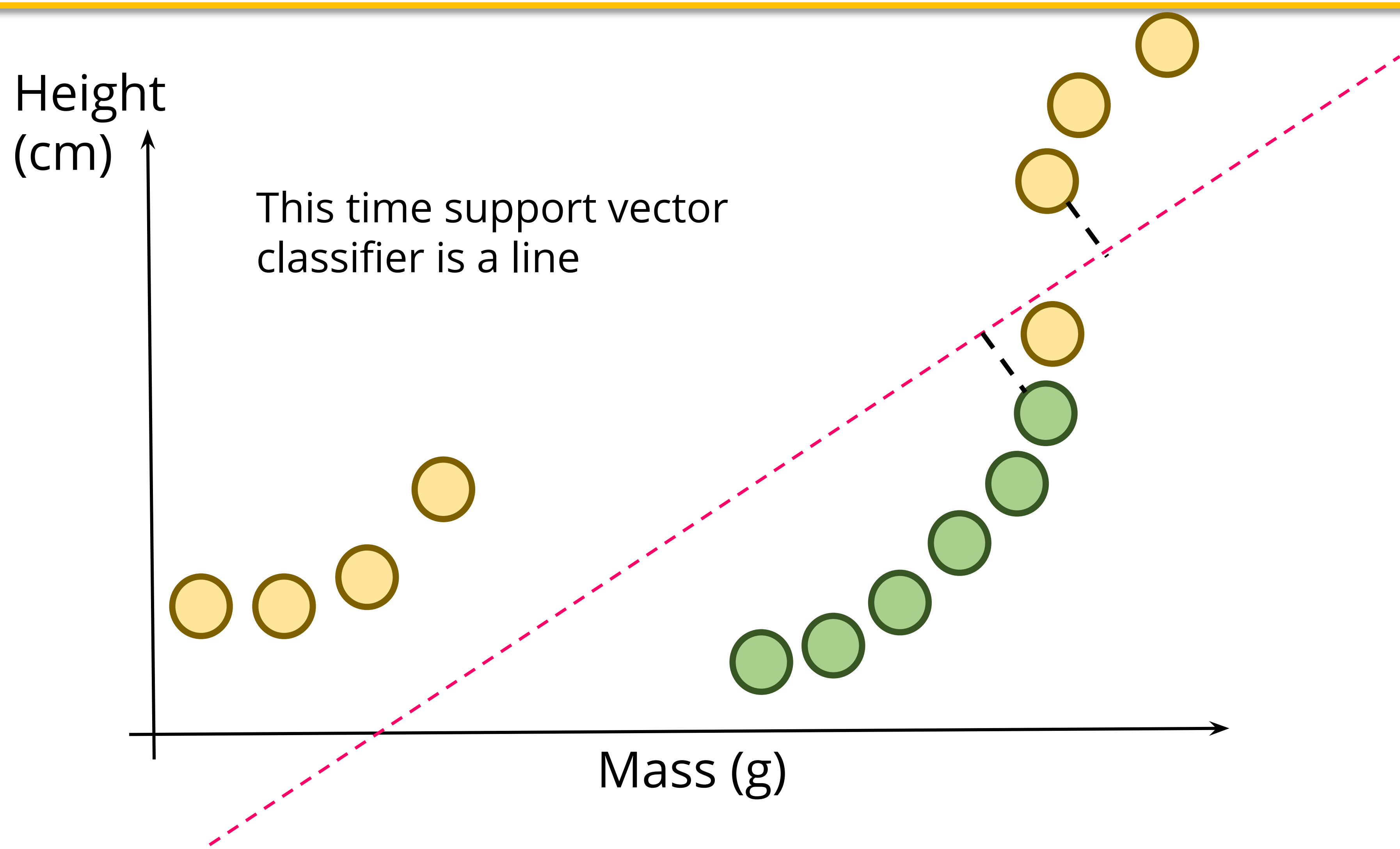
2D


Height
(cm)
Note: If each observation
had a mass measurement
Mass (g)

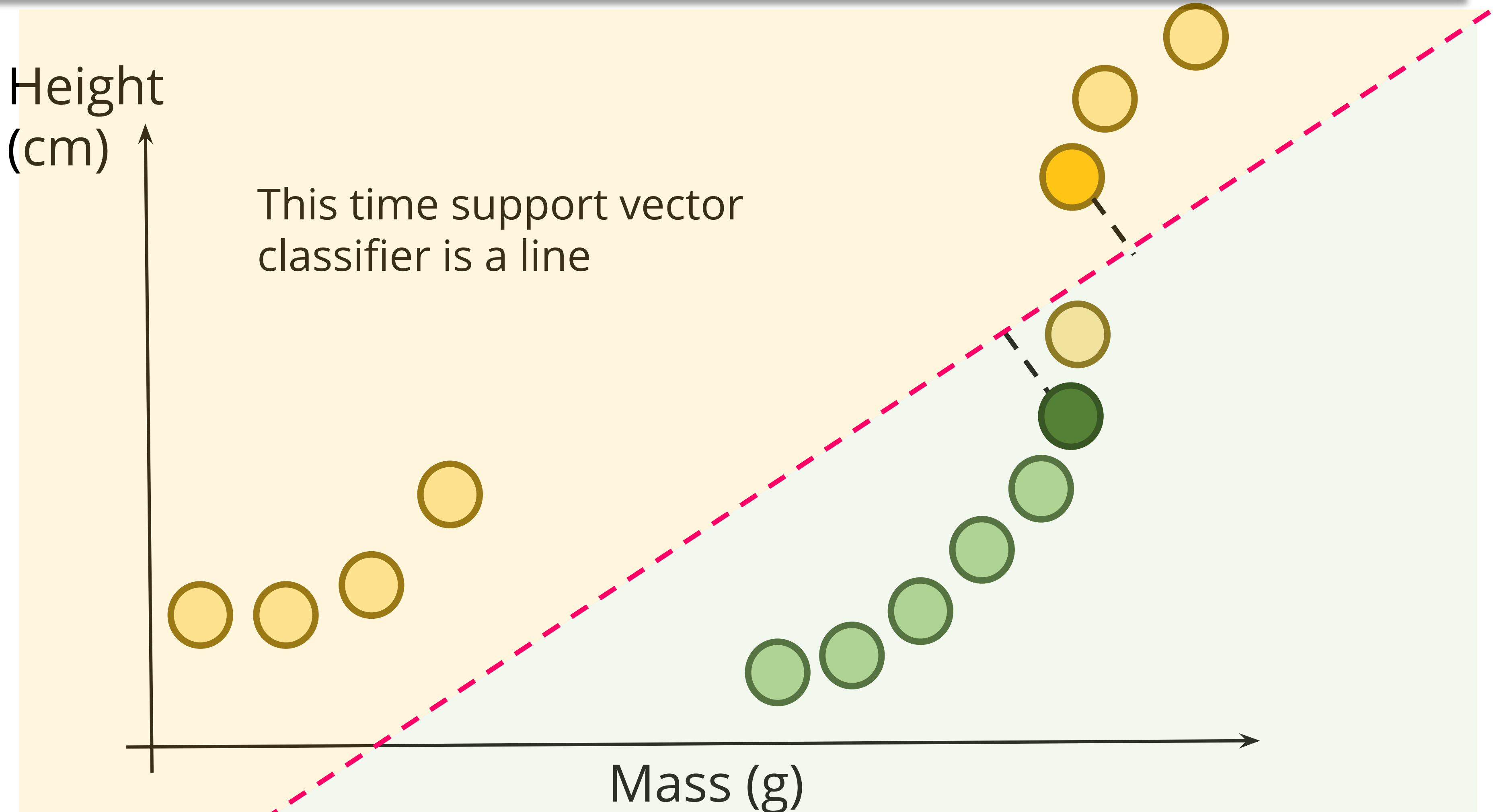
2-Dimension Data



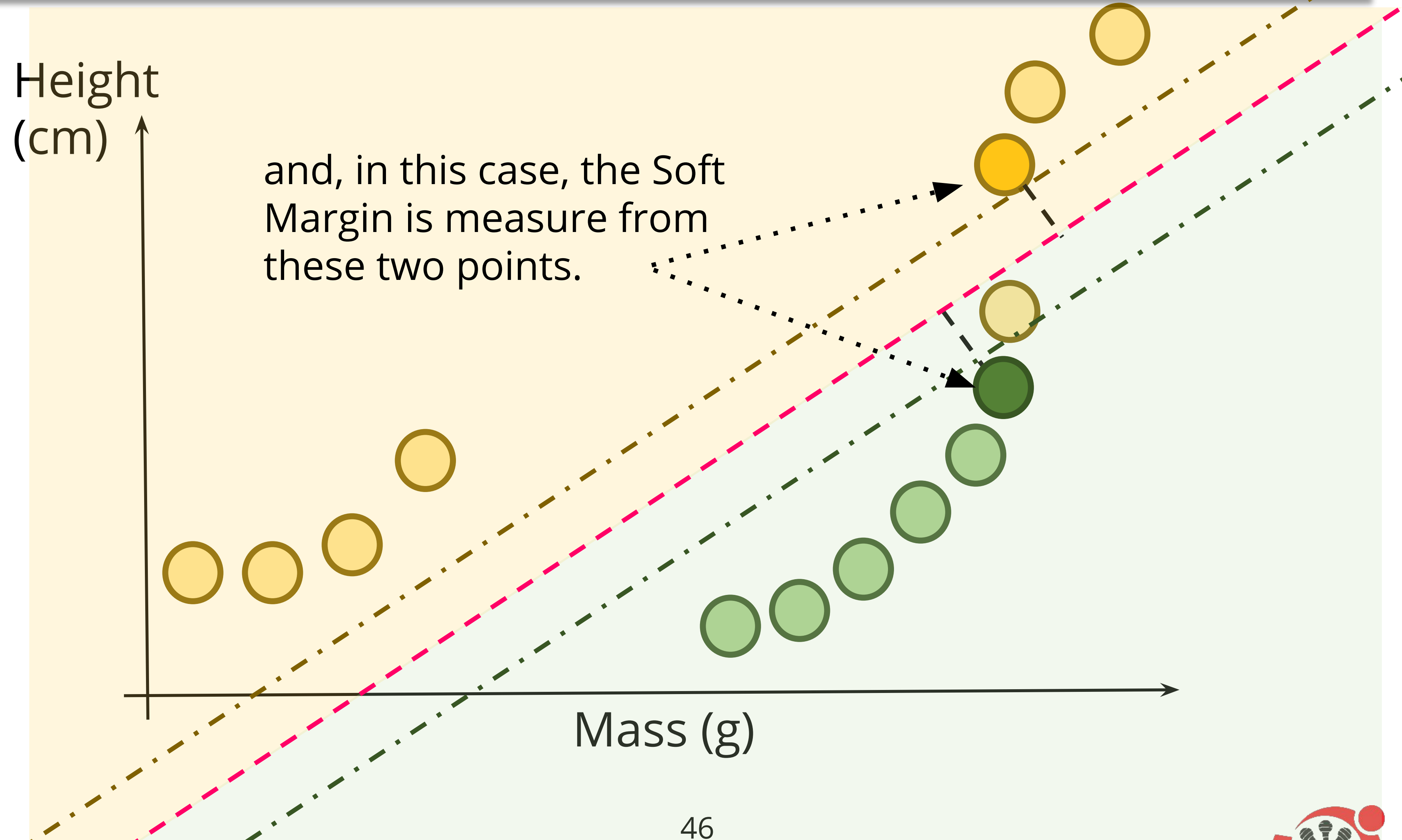
2-Dimension Data



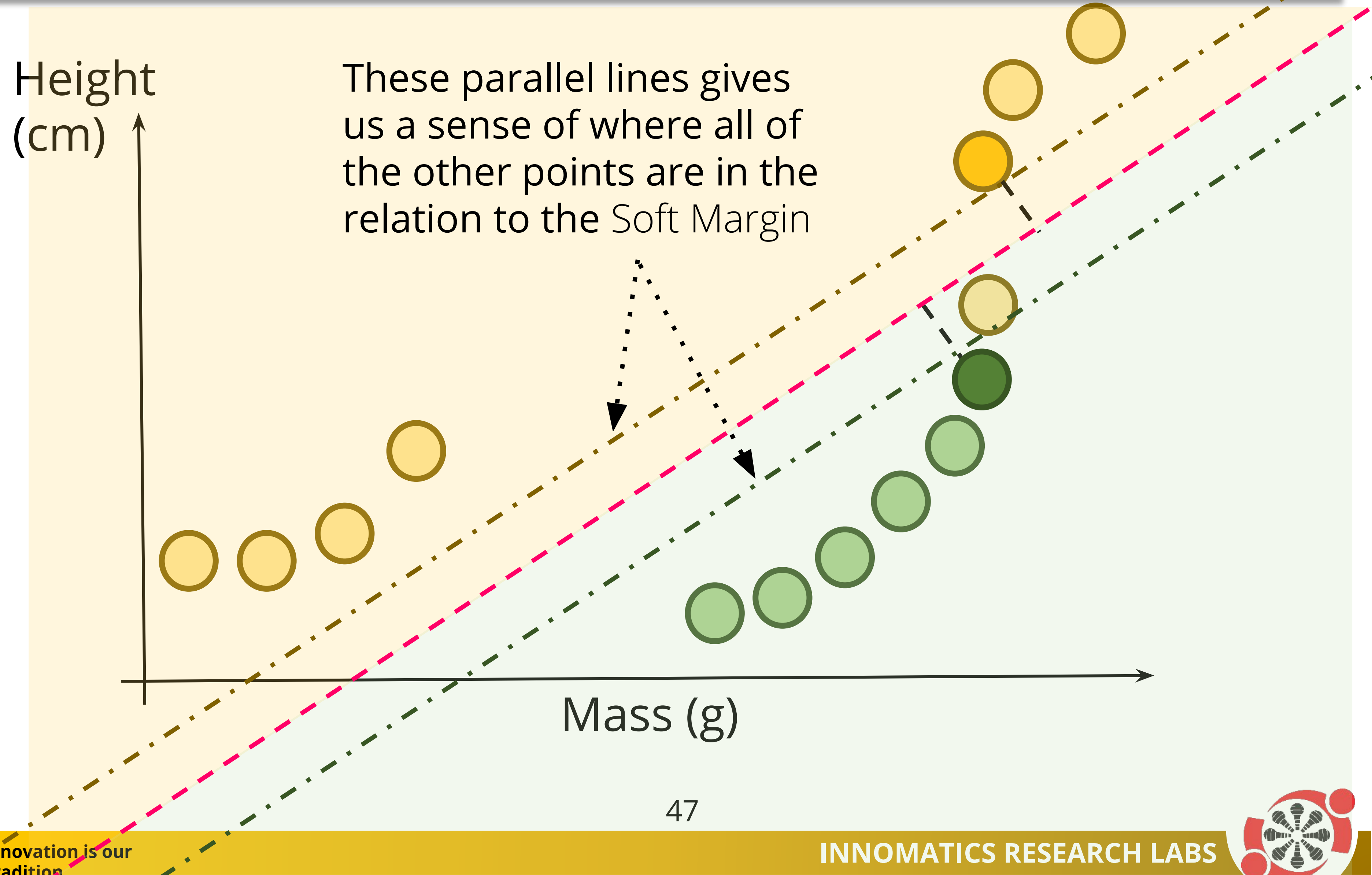
2-Dimension Data



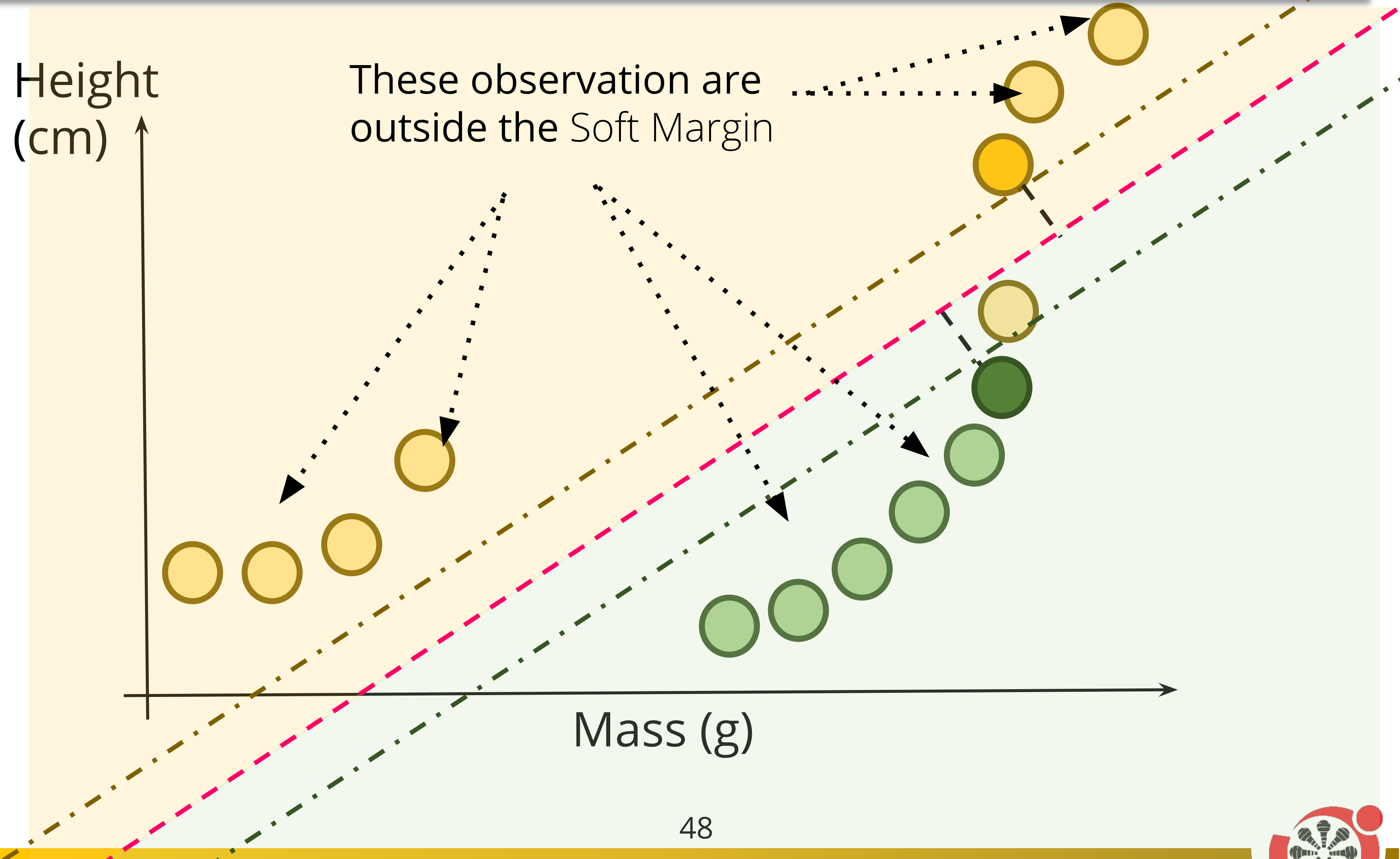
2-Dimension Data



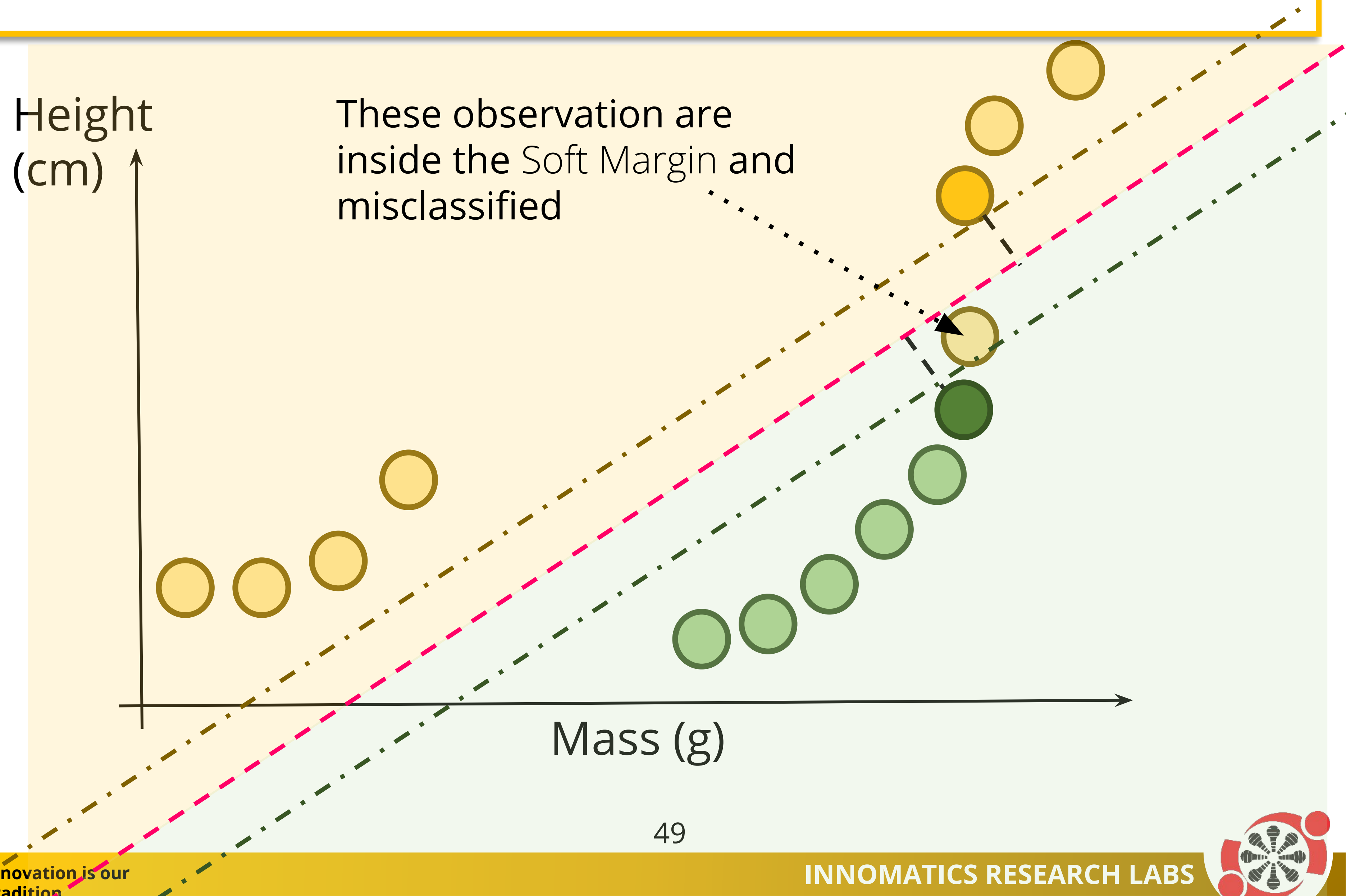
2-Dimension Data



2-Dimension Data



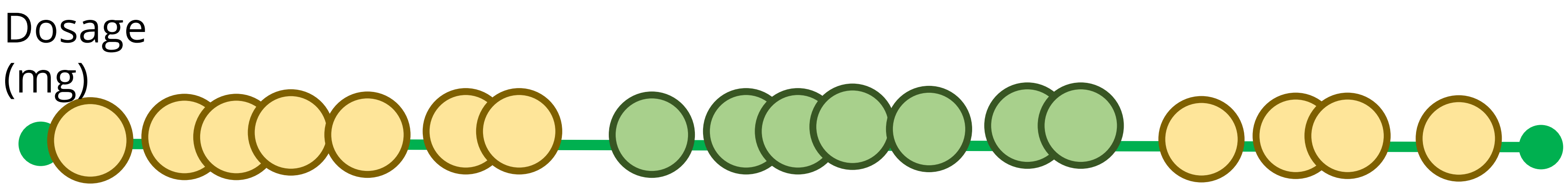
2-Dimension Data



Support Vector Classifiers

- Support Vector Classifiers seem pretty cool because they can handle outliers.
- They allow misclassifications, they can handle overlapping classifications.

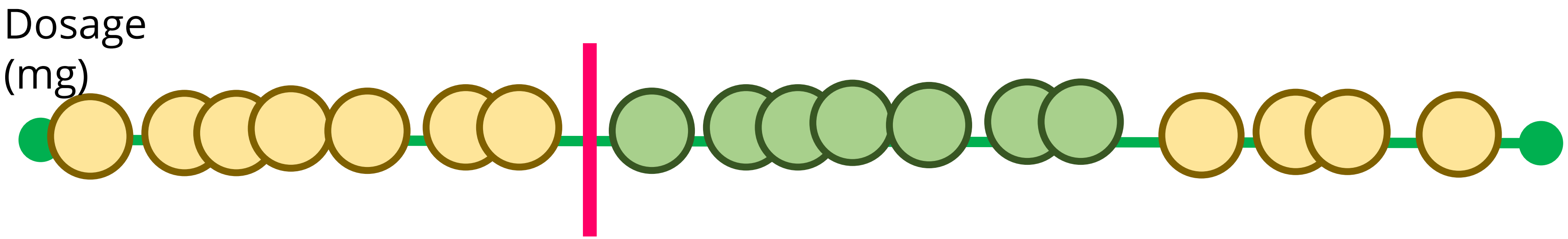
Case-5



Yellow : Drug Not Cured
Green: Drug Cured

- **What if this was our training data and we has tons of overlap ?**

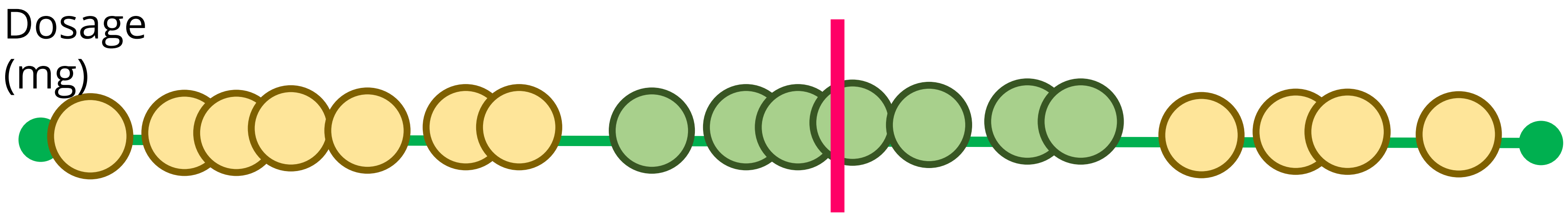
Case-5



Yellow : Drug Not Cured
Green: Drug Cured

- **Support Vector Classifiers are only semi-cool, since they don’t perform will with type of data**

Case-5



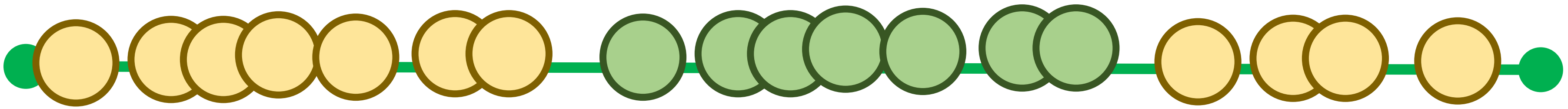
Yellow : Drug Not Cured
Green: Drug Cured

- **Support Vector Classifiers are only semi-cool, since they don’t perform will with type of data**

Support Vector Machines

Case -5

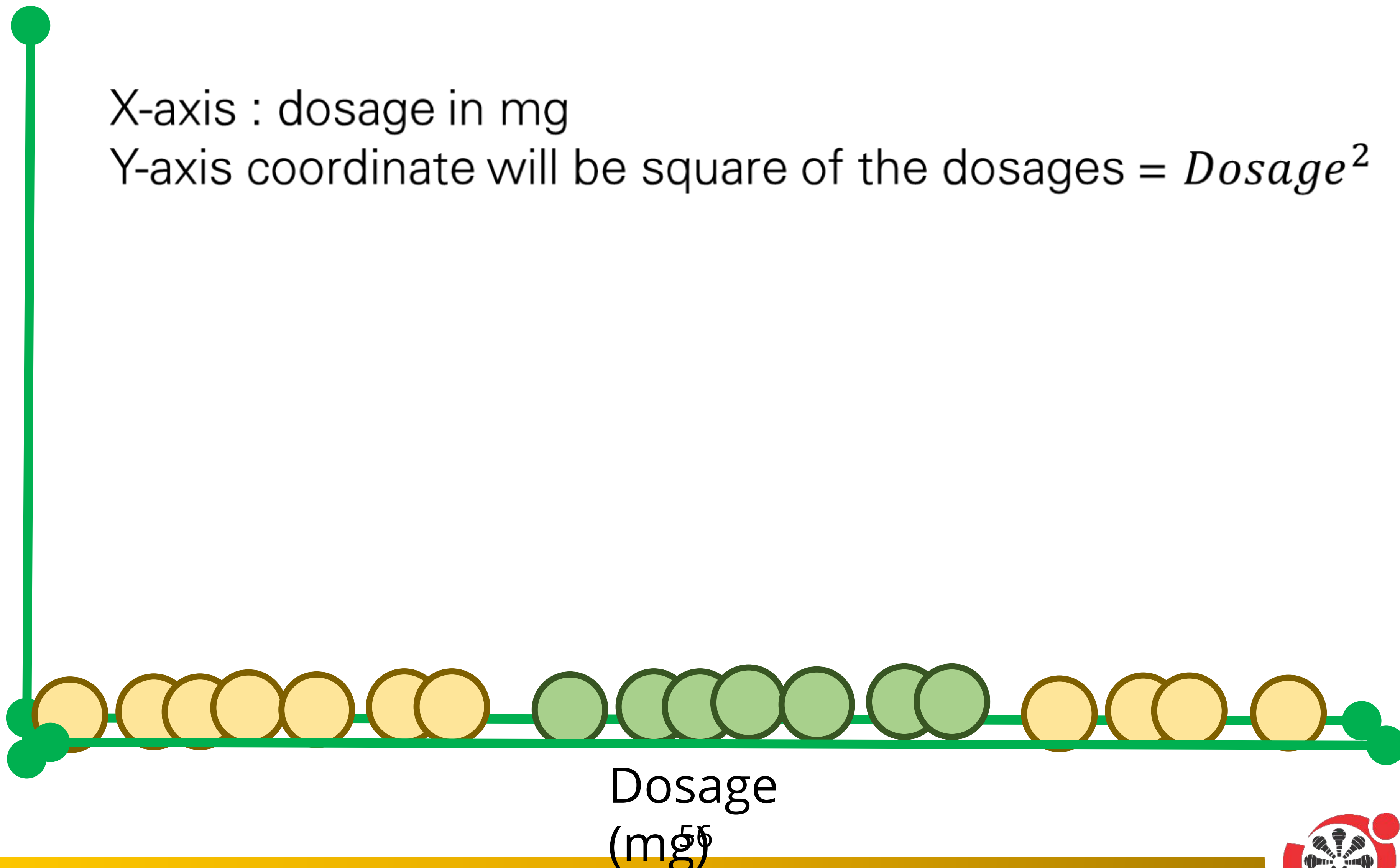
- Since Maximal Margin Classifier and Support Vector Classifiers can’t handle this data



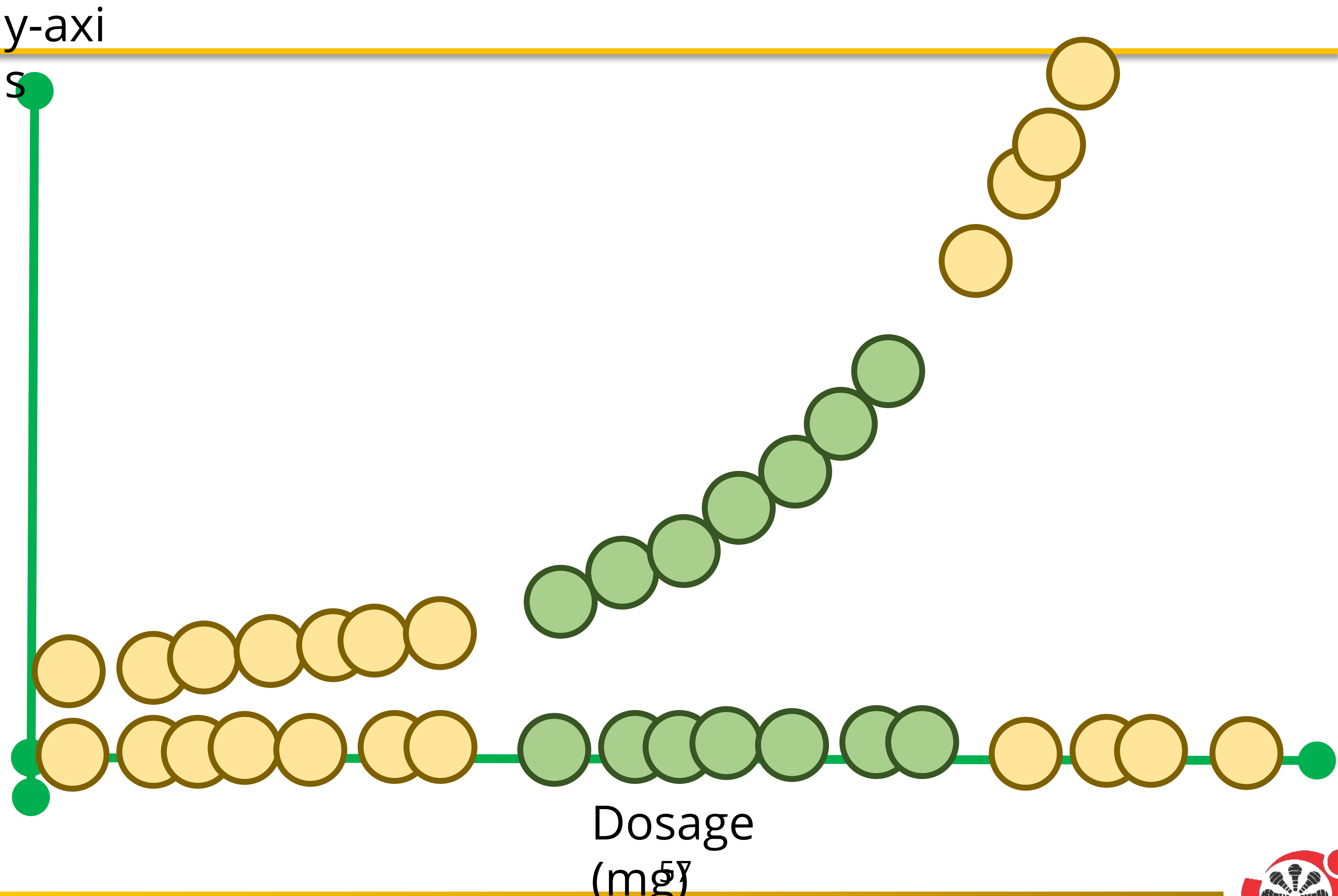
- Let’s start by getting get an intuitive sense of the main ideas behind Support Vector Machines

Case-5

X-axis : dosage in mg
Y-axis coordinate will be square of the dosages = $Dosage^2$

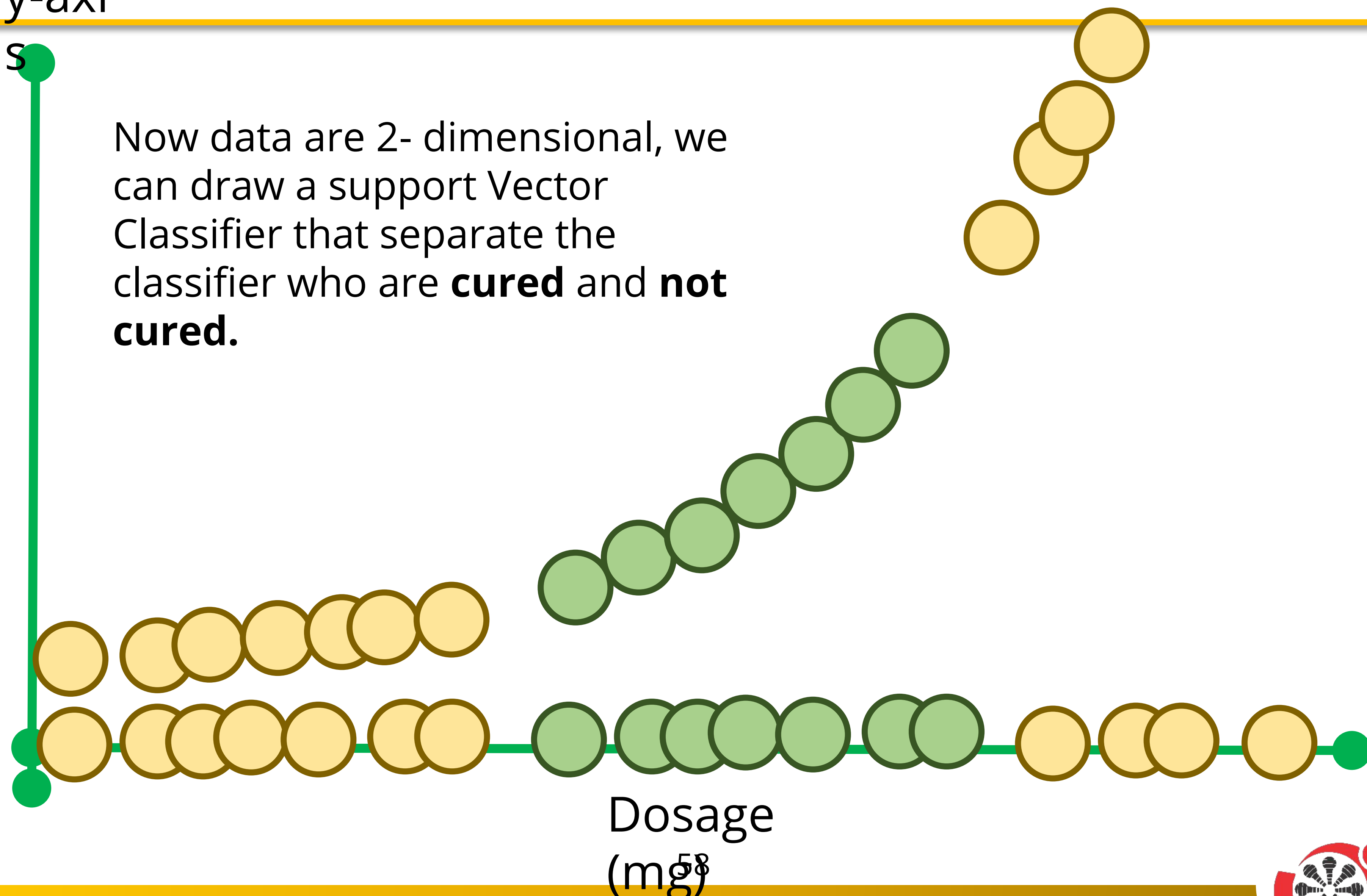


Case-5



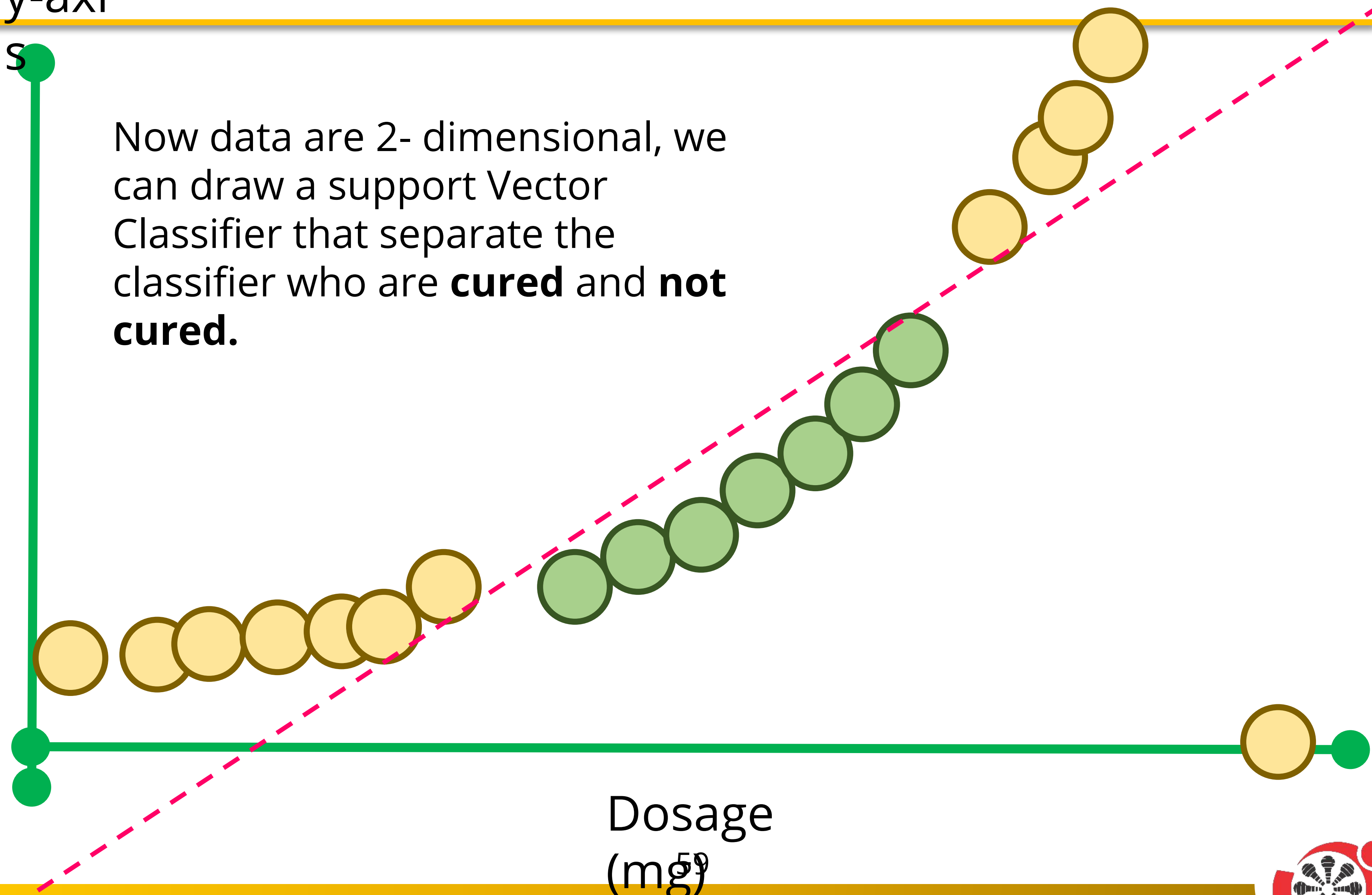
Case-5

Now data are 2- dimensional, we can draw a support Vector Classifier that separate the classifier who are **cured** and **not cured.**

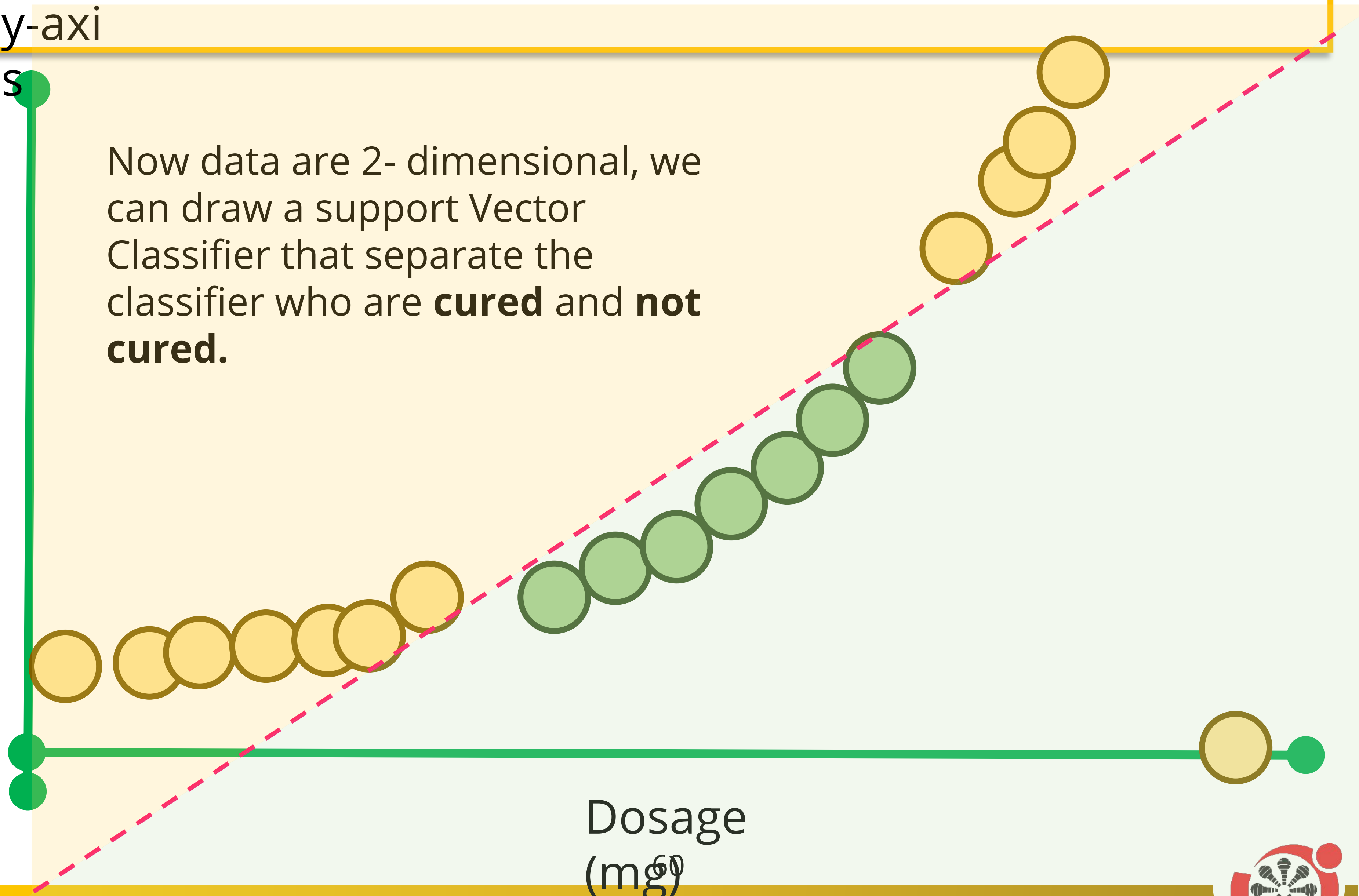


Case-5

Now data are 2- dimensional, we can draw a support Vector Classifier that separate the classifier who are **cured** and **not cured.**

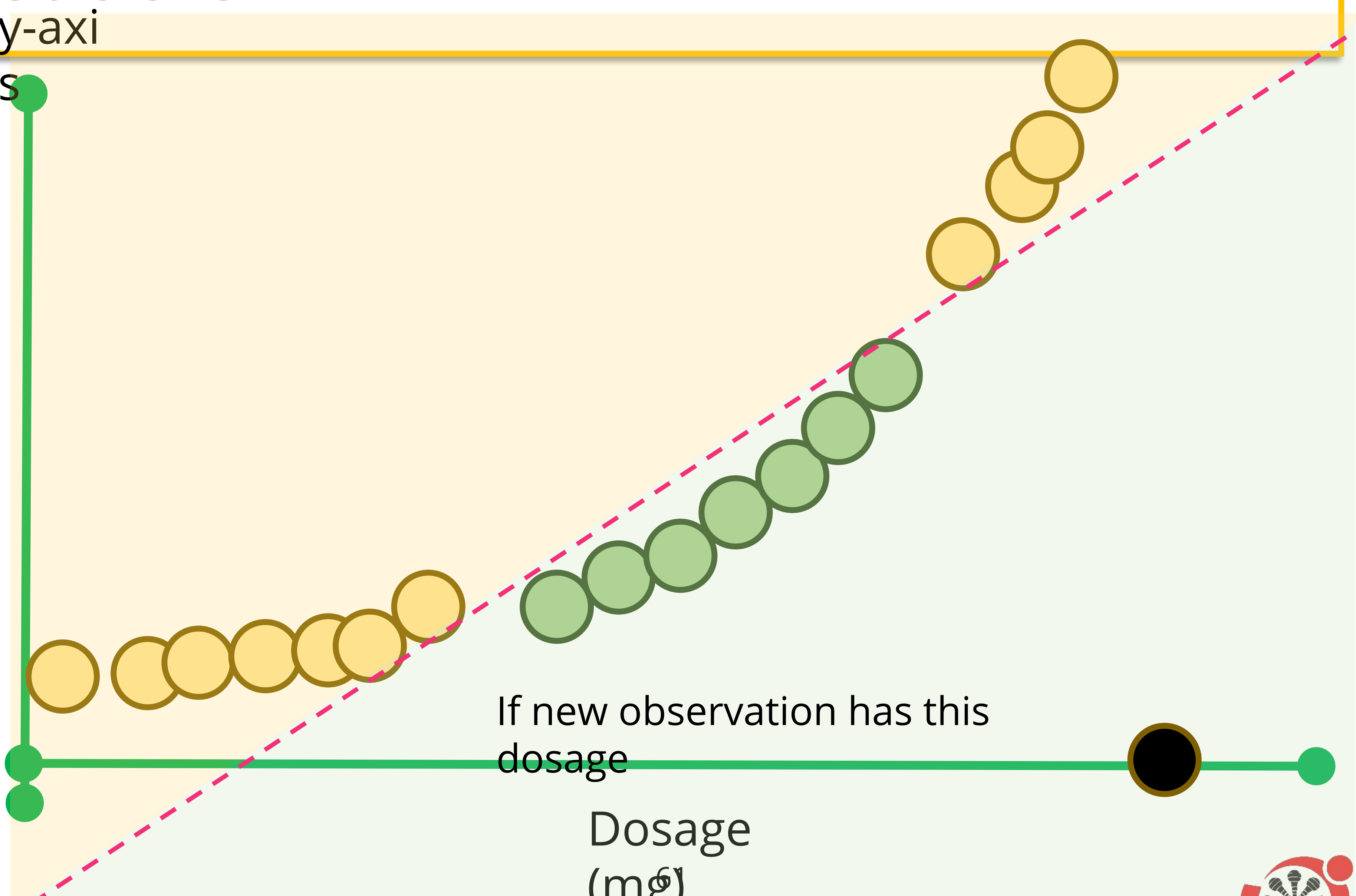


Case-5

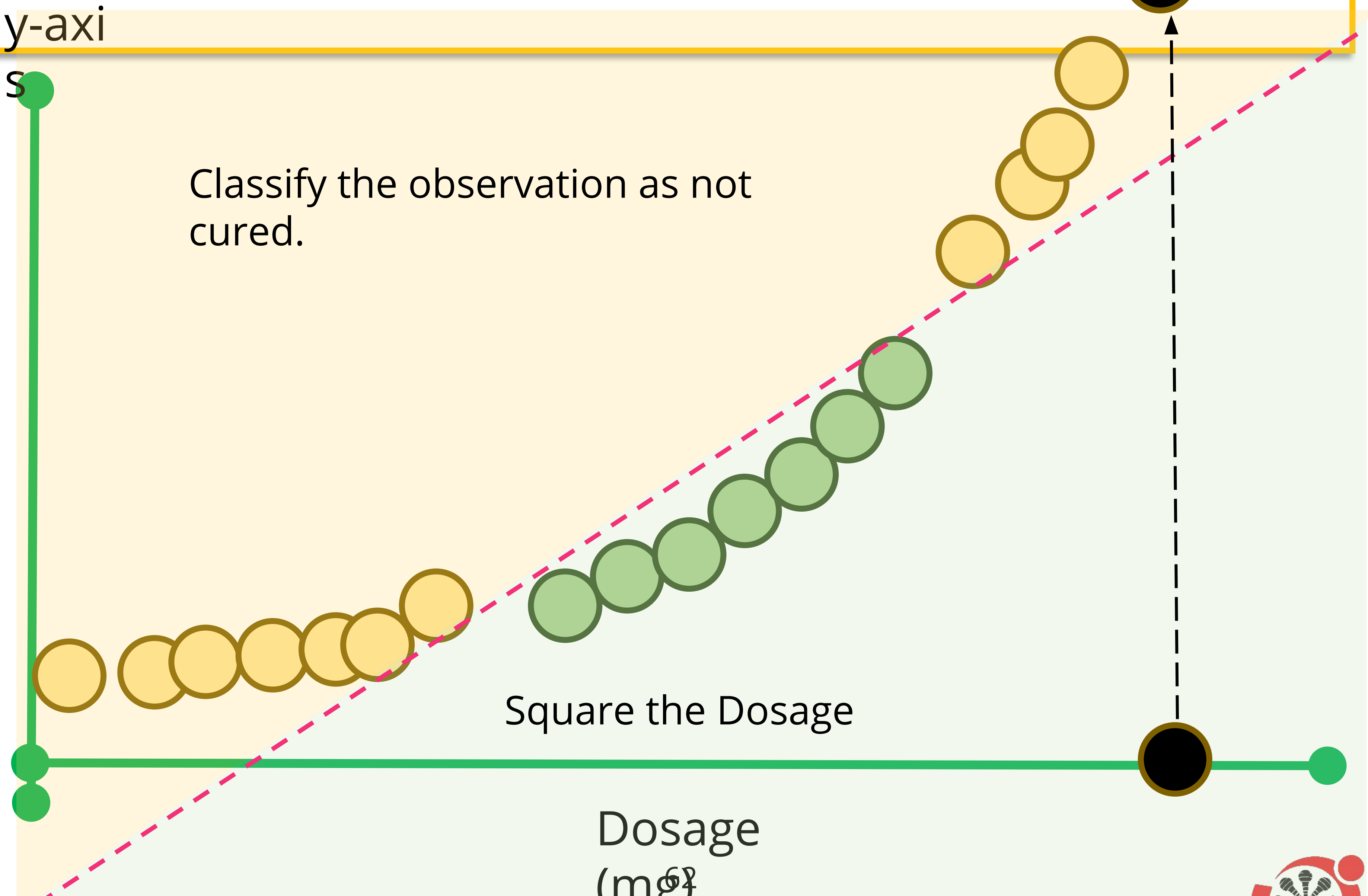


Now data are 2- dimensional, we can draw a support Vector Classifier that separate the classifier who are **cured** and **not cured.**

Case-5



Case-5


y-axi
s
Classify the observation as not cured.
Square the Dosage
Dosage
(mg)

Support Vector Machines

The main idea of support vector machine is

1. Start with data in a relatively low dimension
2. Move the data into a higher dimension
3. Find a support vector classifier that separate the higher dimension data into two groups

Transform the data ?

- How do we decide how to transform the data ?

To make mathematics possible, Support Vector Machines use something called Kernel Functions to systematically find Support Vector Classifier in higher dimensions.