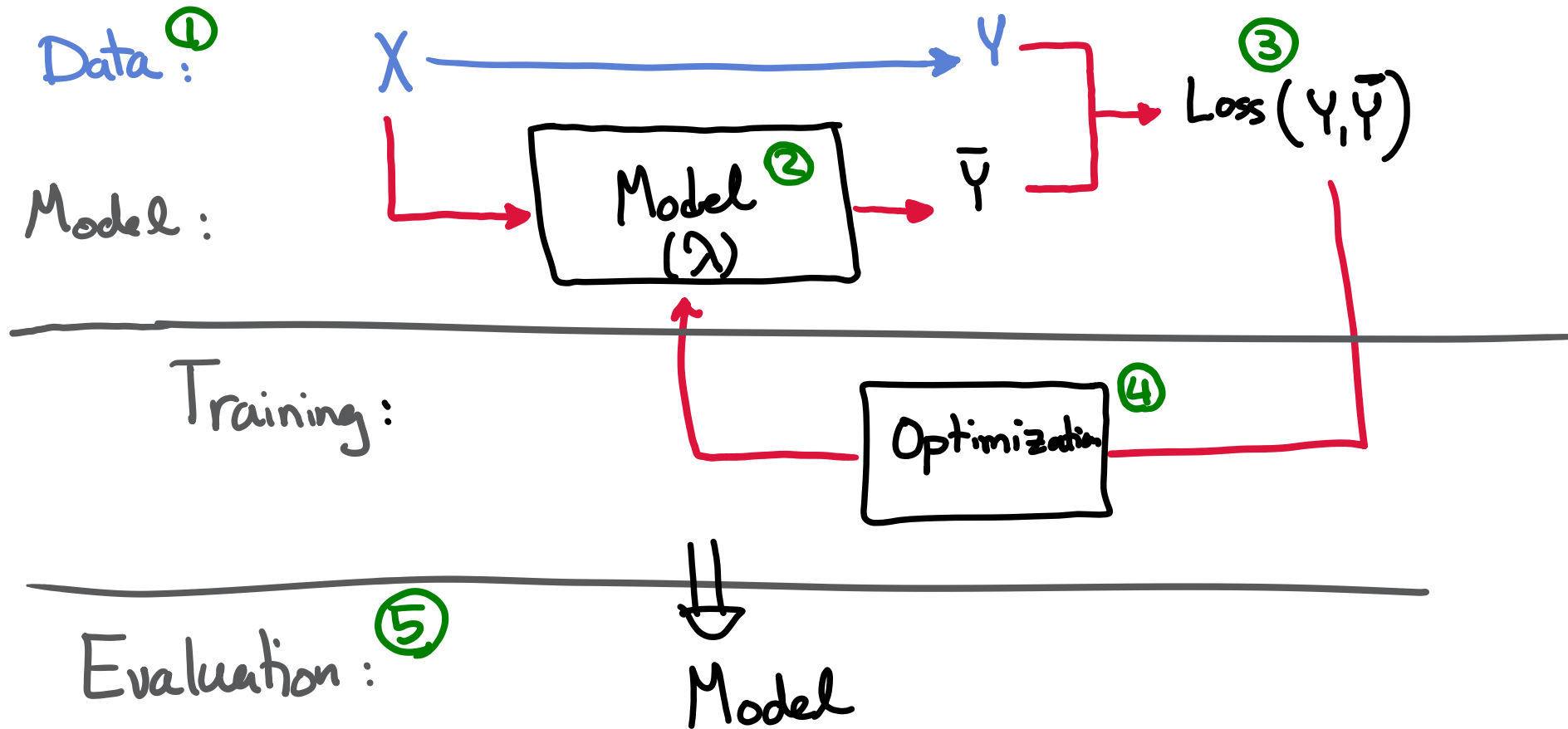


The background of the slide features a complex, abstract pattern of blue and white wavy lines, resembling a topographical map or a fluid simulation, set against a solid black background. The lines are dense and intricate, creating a sense of depth and movement.

Machine Learning in Physics: **Models**

Sadegh Raeisi

Supervised: Ingredients



Outline



A Simple model



Linear vs non-linear



Inherently non-linear models

Notation

The diagram illustrates a machine learning model function f_w . The function is represented by the symbol f with a subscript w . A red circle highlights the w , with a red arrow pointing down to the text "Parameters of the model". The function takes as input the features of a sample, denoted by \vec{X}^i , which is enclosed in a red circle with a red arrow pointing up to the text "Features of the sample". The output of the function is the label of the sample, denoted by \bar{Y} , which is also enclosed in a red circle with a red arrow pointing up to the text "Label of the sample". The entire expression is set within large parentheses, followed by an equals sign.

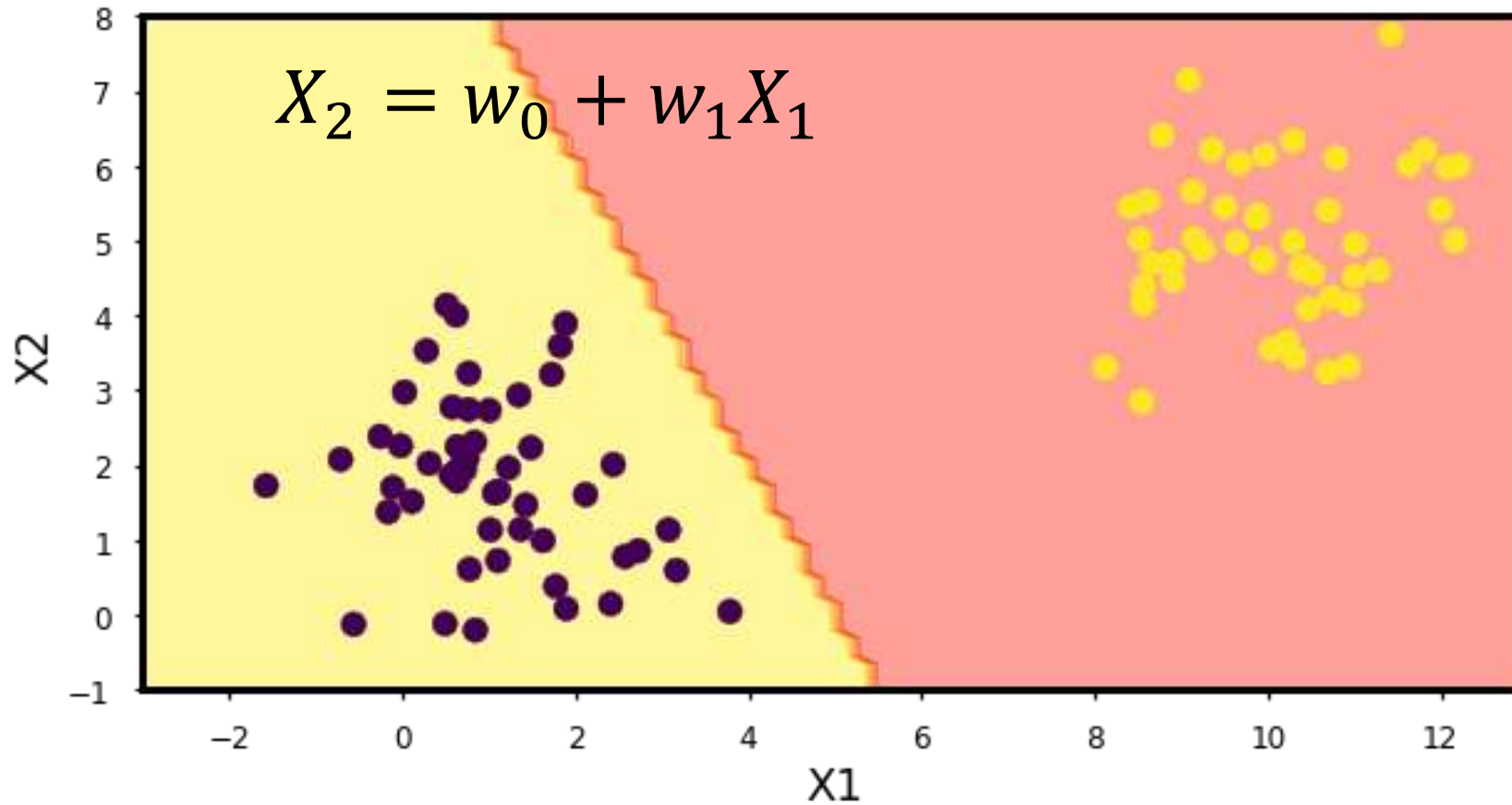
$$f_w(\vec{X}^i) = \bar{Y}$$

Parameters of the model

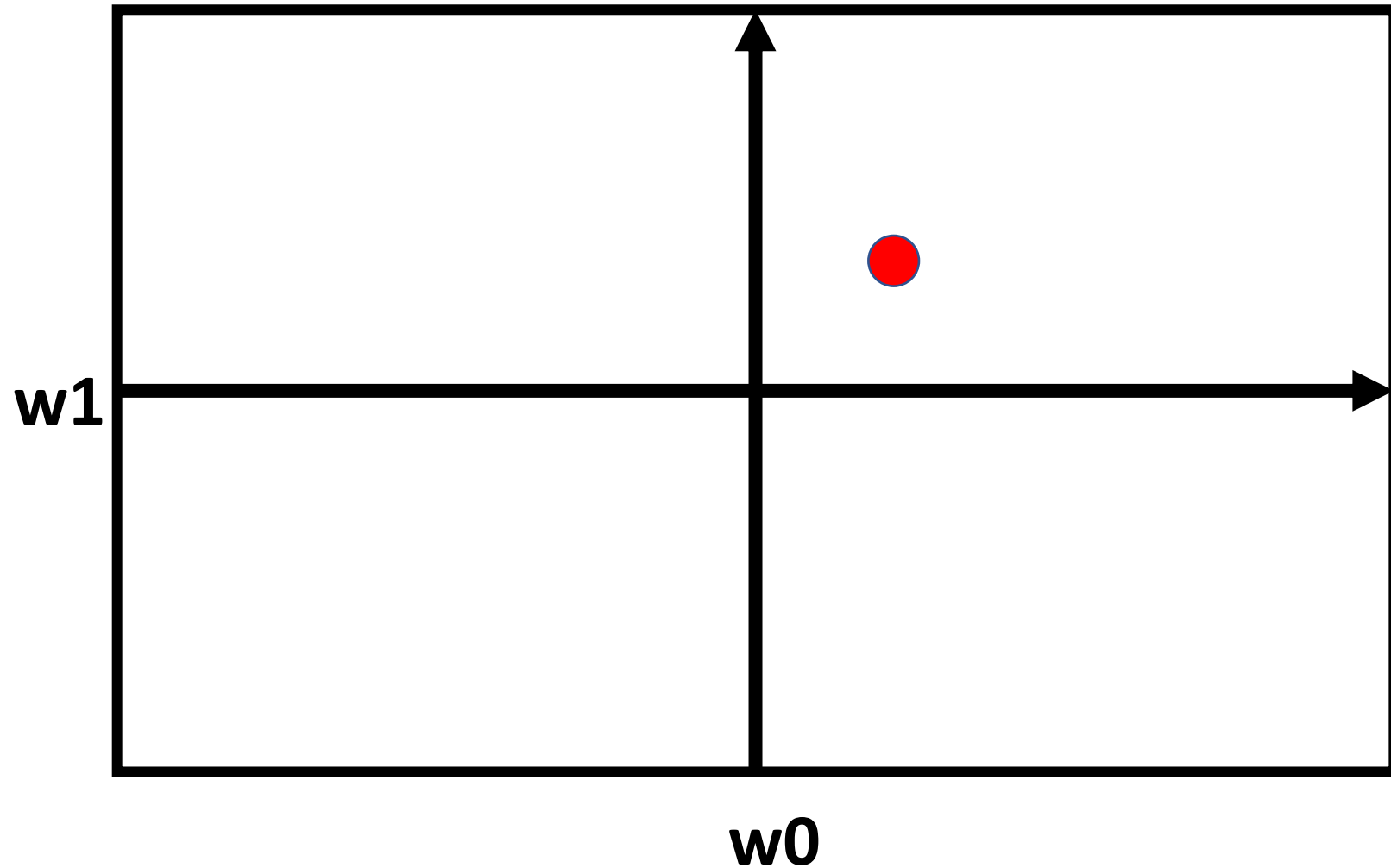
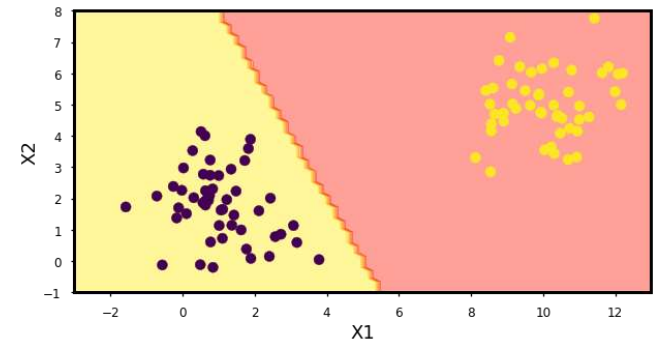
Features of the sample

Label of the sample

Feature space and decision boundary



Model space



A Simple Model

Simplest model possible

$$f_w \left(\vec{X}^i \right) = \text{const.}$$

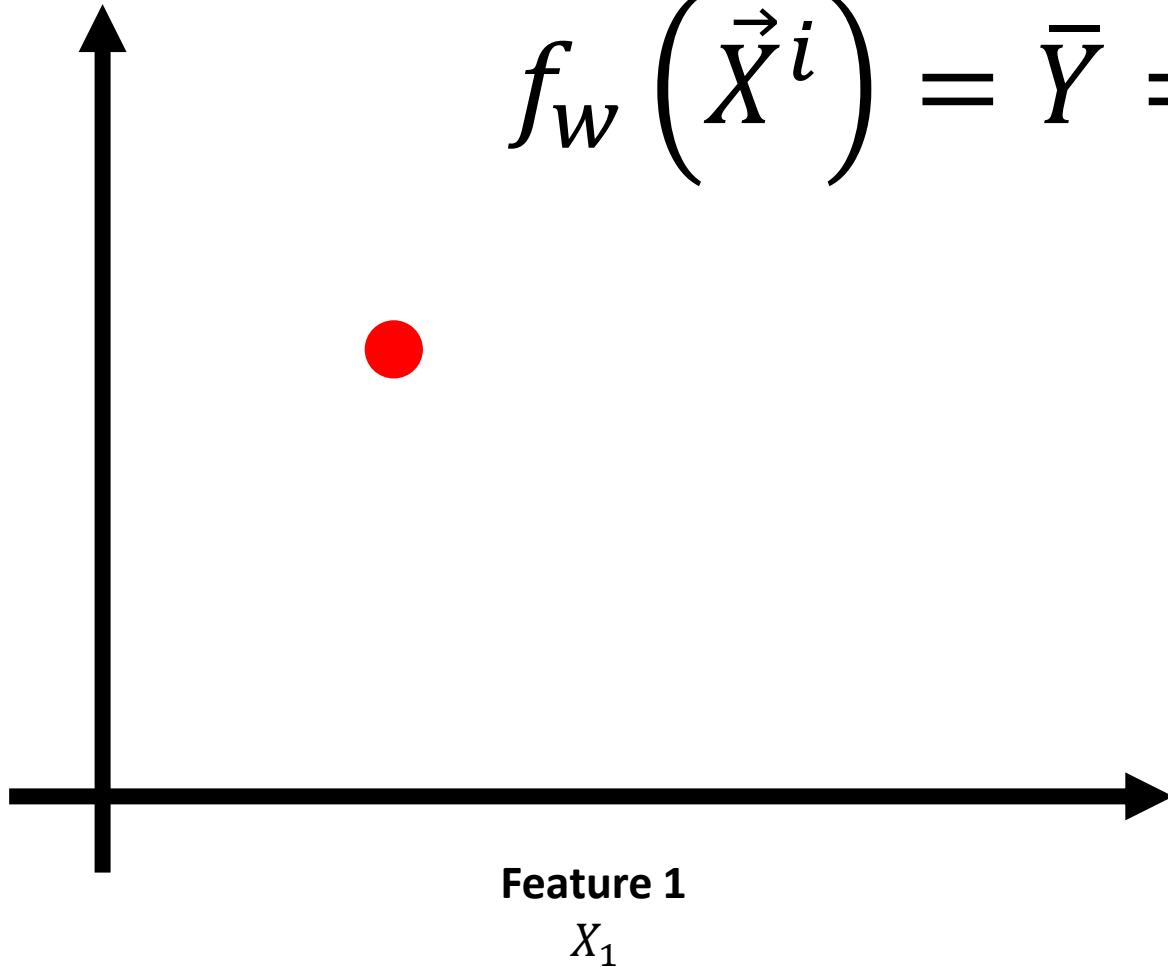
Simplest model possible: linear model

$$f_w(\vec{X}^i) = \sum_j w_j X_j^{(i)}$$

One sample

$$f_w(\vec{X}^i) = \bar{Y} = w_2 X_2 + w_1 X_1 + w_0$$

Feature 2
 X_2



Feature 1
 X_1

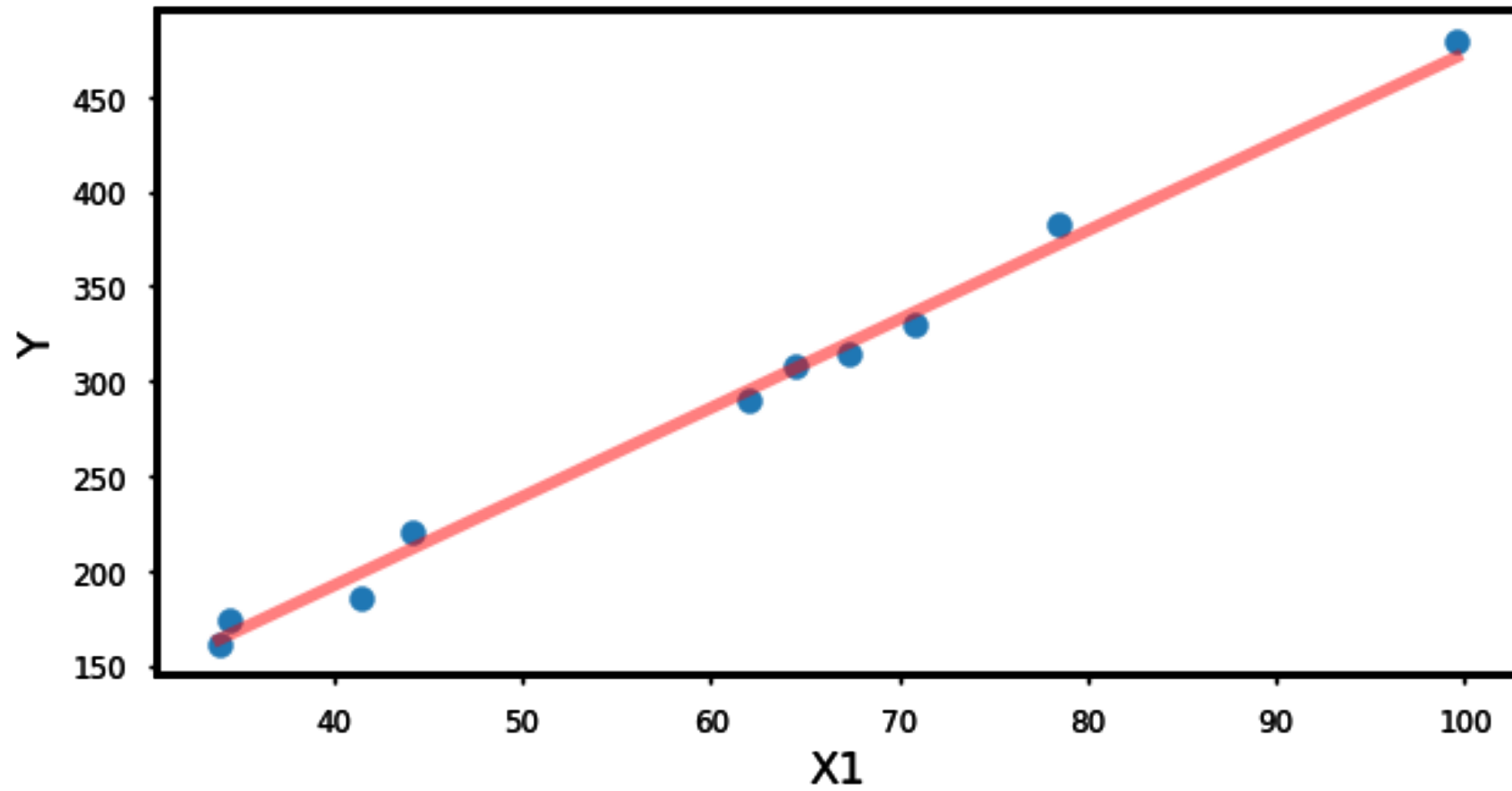
Vectorization

$$\begin{aligned} f_w \left(\vec{X}^i \right) &= \sum_j w_j X_j^{(i)} \\ &= \vec{w} \cdot \vec{X}^i \end{aligned}$$

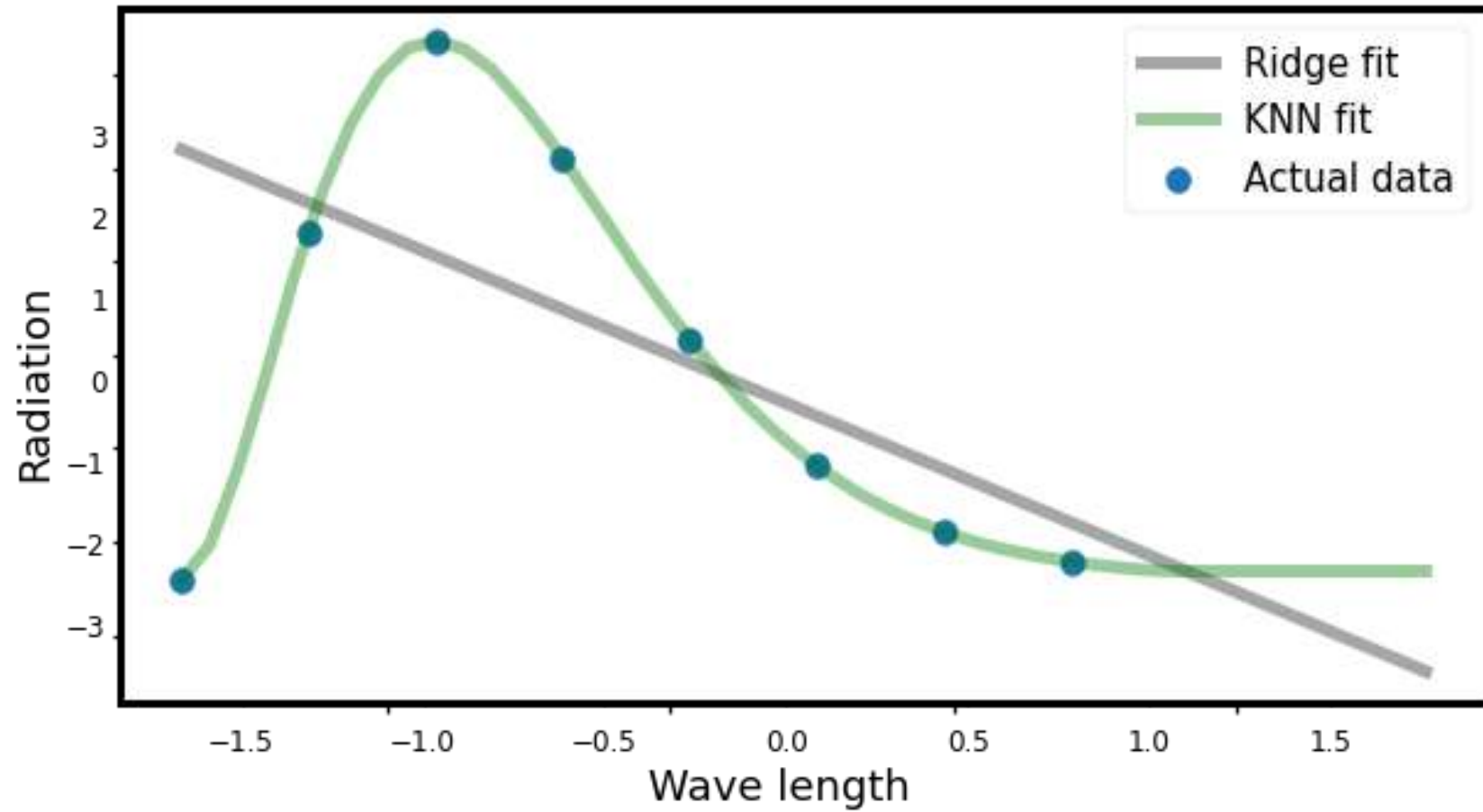
$$\vec{w} = (w_0, w_1, \cdots w_{n_f})$$

$$\vec{X} = (1, X_1, \cdots X_{n_f})$$

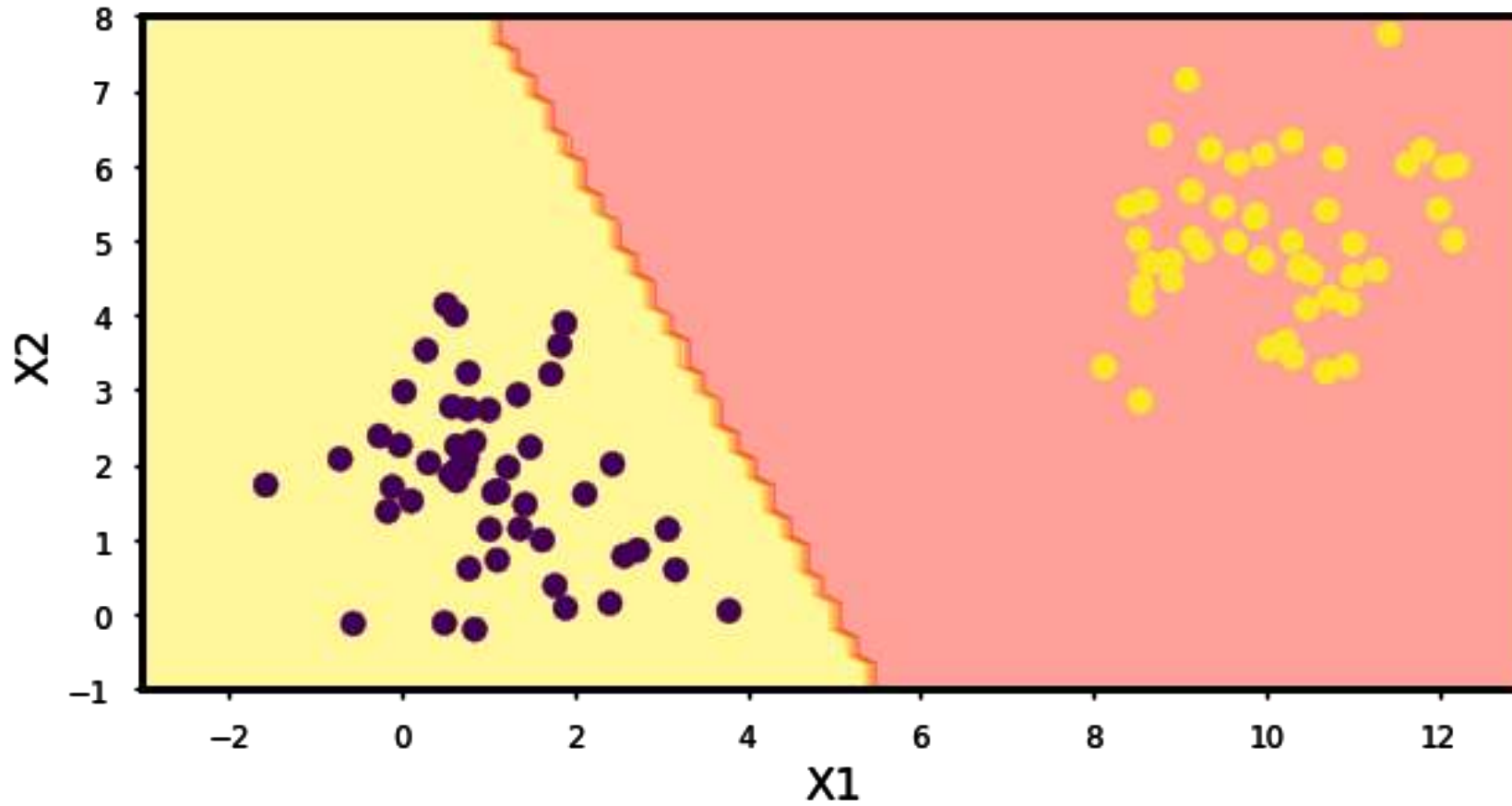
Regression



Regression

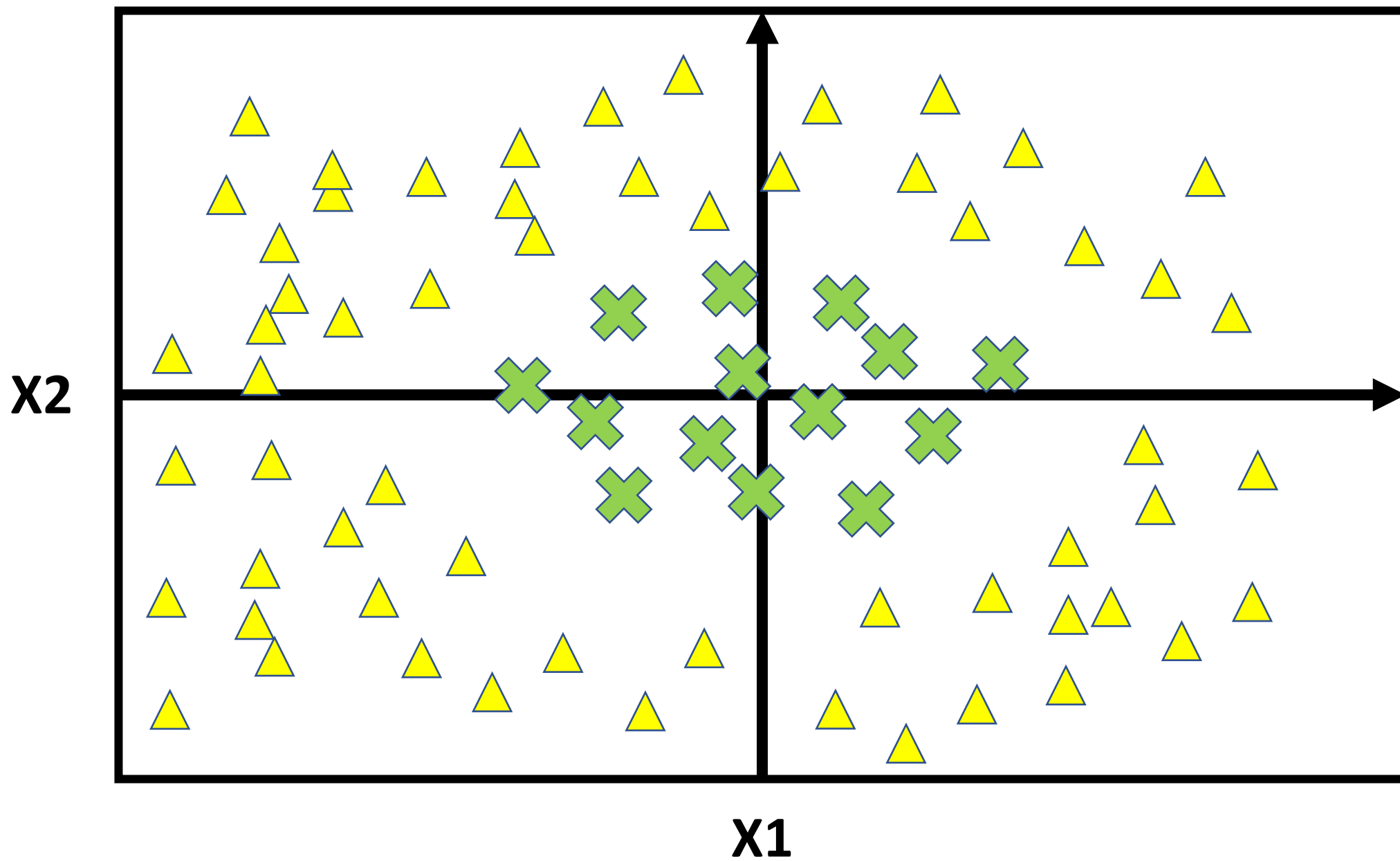


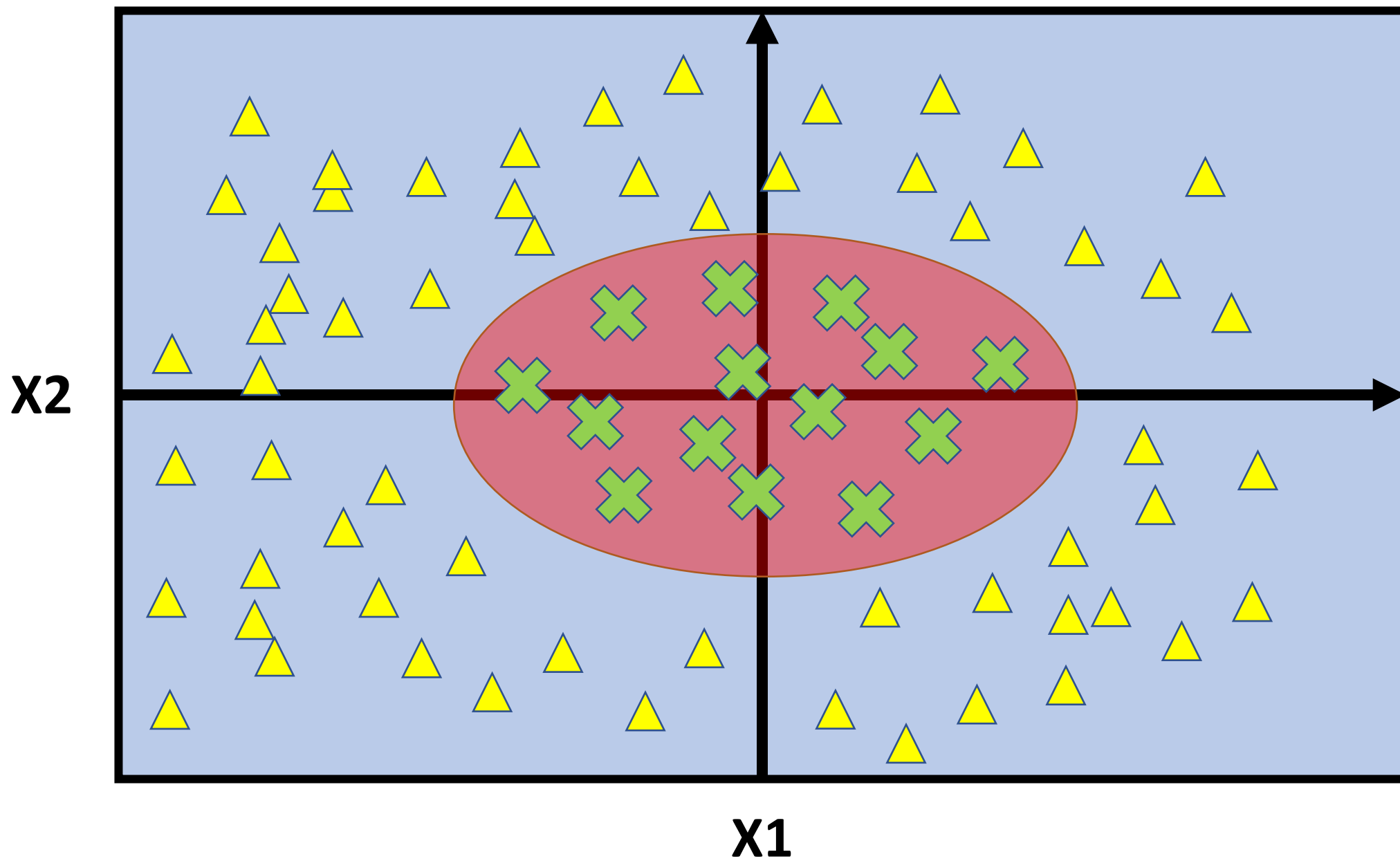
Classification



What is a good fit?

Linear
vs.
Non-linear





Polynomial models

How can we make a quadratic model?

Example:

$$\vec{X} = (X_1, X_2)$$

$$f_w(\vec{X}^i) = w_0 + w_1X_1 + w_2X_2$$

$$+ w_3X_1^2 + w_4X_1X_2 + w_5X_2^2$$

How?

1. Feature Transformation

$$\vec{X} \Rightarrow \Phi(X)$$

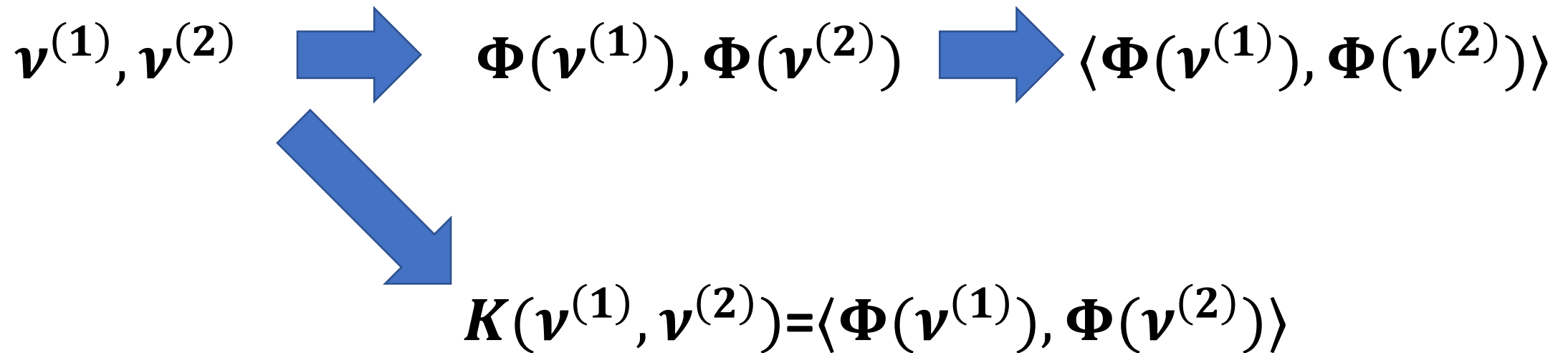
Example: Quadratic model

$$\vec{X} = (1, X_1, X_2) \Rightarrow \Phi(X) = (1, X_1, X_2, X_1^2, X_2^2, X_1X_2)$$

How?

2. Kernel

Often we are interested in a scalar product $\langle \mathbf{v}^{(1)}, \mathbf{v}^{(2)} \rangle$



Example:

$$\Phi \left(\begin{bmatrix} X_1 \\ X_2 \end{bmatrix} \right) = \begin{bmatrix} X_1^2 \\ X_2^2 \\ \sqrt{2}X_1X_2 \end{bmatrix}$$

$$\left\langle \Phi \left(\begin{bmatrix} X_1 \\ X_2 \end{bmatrix} \right), \Phi \left(\begin{bmatrix} Z_1 \\ Z_2 \end{bmatrix} \right) \right\rangle$$

Example:

$$\Phi \left(\begin{bmatrix} X_1 \\ X_2 \end{bmatrix} \right) = \begin{bmatrix} X_1^2 \\ X_2^2 \\ \sqrt{2}X_1X_2 \end{bmatrix}$$

$$\left\langle \Phi \left(\begin{bmatrix} X_1 \\ X_2 \end{bmatrix} \right), \Phi \left(\begin{bmatrix} Z_1 \\ Z_2 \end{bmatrix} \right) \right\rangle = \left\langle \begin{bmatrix} X_1^2 \\ X_2^2 \\ \sqrt{2}X_1X_2 \end{bmatrix}, \begin{bmatrix} Z_1^2 \\ Z_2^2 \\ \sqrt{2}Z_1Z_2 \end{bmatrix} \right\rangle$$

Example:

$$\Phi \left(\begin{bmatrix} X_1 \\ X_2 \end{bmatrix} \right) = \begin{bmatrix} X_1^2 \\ X_2^2 \\ \sqrt{2}X_1X_2 \end{bmatrix}$$

$$\begin{aligned} \left\langle \Phi \left(\begin{bmatrix} X_1 \\ X_2 \end{bmatrix} \right), \Phi \left(\begin{bmatrix} Z_1 \\ Z_2 \end{bmatrix} \right) \right\rangle &= \left\langle \begin{bmatrix} X_1^2 \\ X_2^2 \\ \sqrt{2}X_1X_2 \end{bmatrix}, \begin{bmatrix} Z_1^2 \\ Z_2^2 \\ \sqrt{2}Z_1Z_2 \end{bmatrix} \right\rangle \\ &= X_1^2Z_1^2 + X_2^2Z_2^2 + 2X_1X_2Z_1Z_2 \end{aligned}$$

Example:

$$K(X, Z) = \langle X, Z \rangle^2$$

$$\langle X, Z \rangle^2 = \left\langle \begin{bmatrix} X_1 \\ X_2 \end{bmatrix}, \begin{bmatrix} Z_1 \\ Z_2 \end{bmatrix} \right\rangle^2$$

Example:

$$K(X, Z) = \langle X, Z \rangle^2$$

$$\langle X, Z \rangle^2 = \left\langle \begin{bmatrix} X_1 \\ X_2 \end{bmatrix}, \begin{bmatrix} Z_1 \\ Z_2 \end{bmatrix} \right\rangle^2 = (X_1 Z_1 + X_2 Z_2)^2$$

Example:

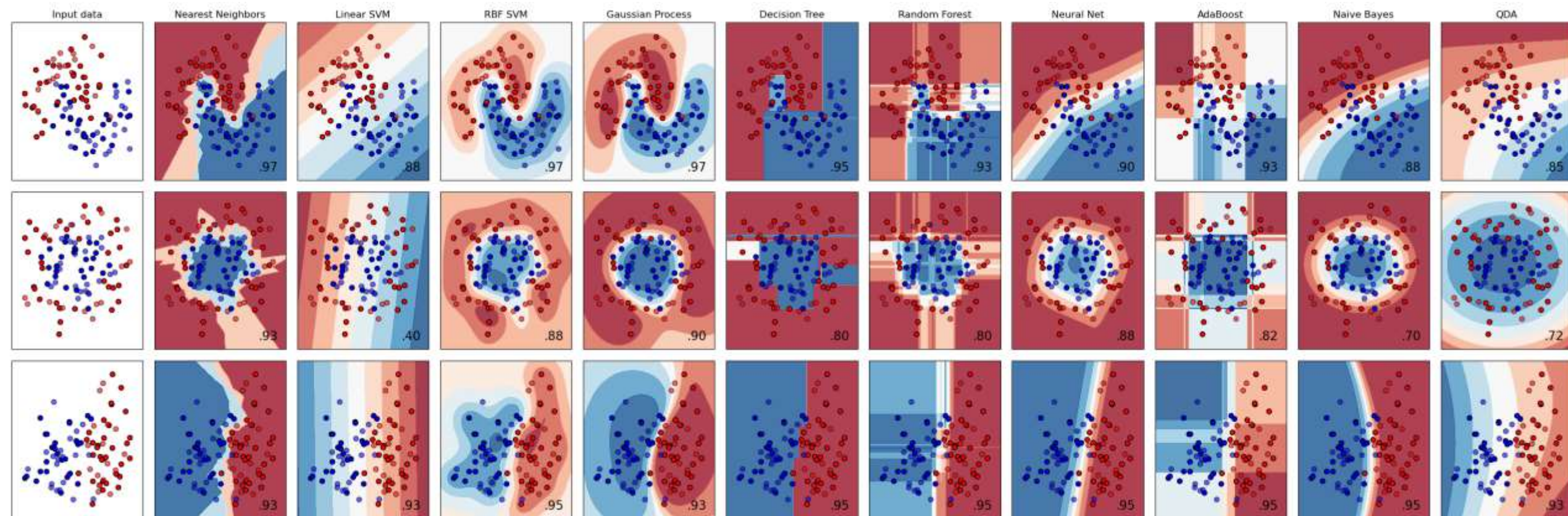
$$K(X, Z) = \langle X, Z \rangle^2$$

$$\langle X, Z \rangle^2 = \left\langle \begin{bmatrix} X_1 \\ X_2 \end{bmatrix}, \begin{bmatrix} Z_1 \\ Z_2 \end{bmatrix} \right\rangle^2 = (X_1 Z_1 + X_2 Z_2)^2$$

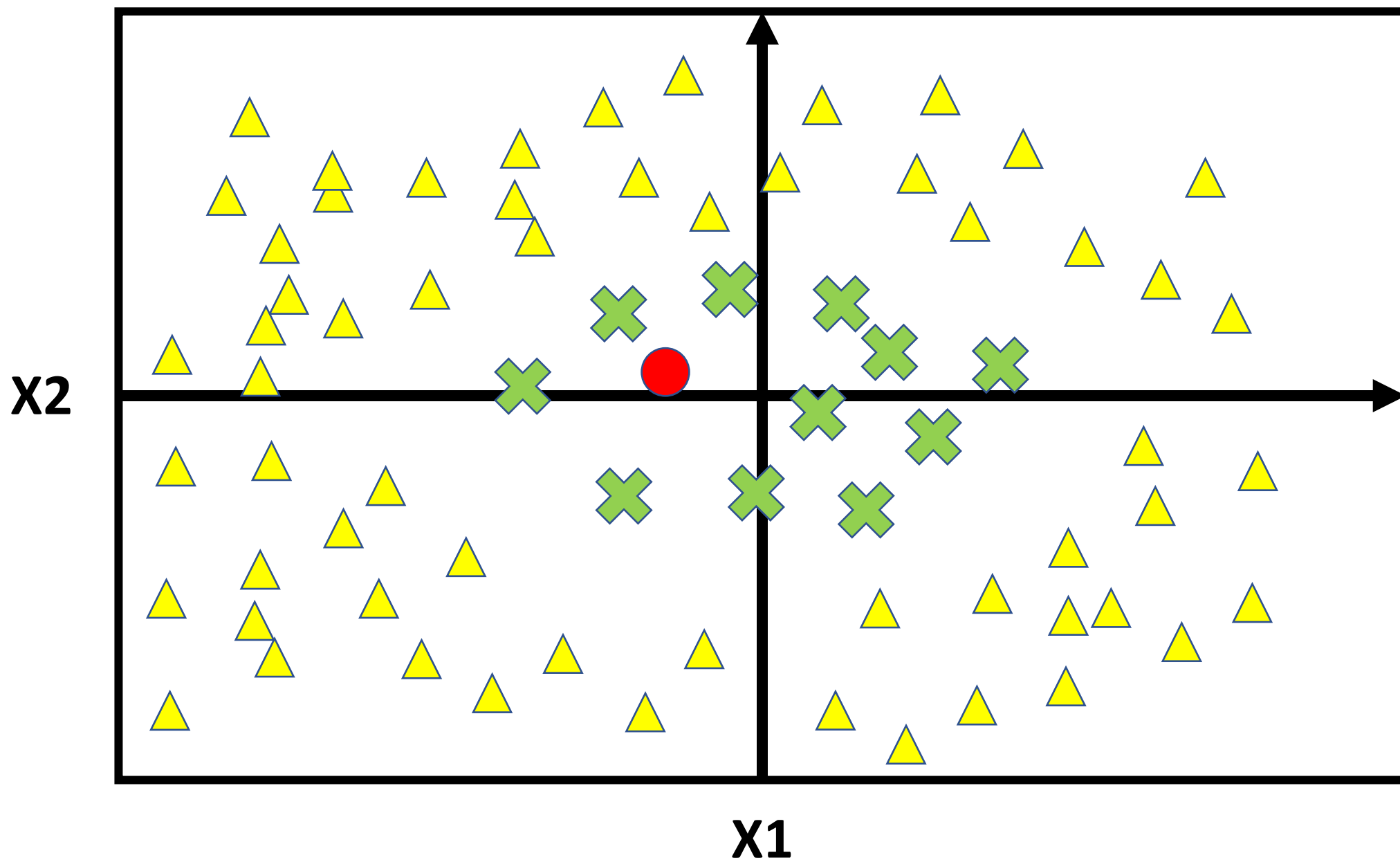
$$= X_1^2 Z_1^2 + X_2^2 Z_2^2 + 2X_1 X_2 Z_1 Z_2 = \left\langle \Phi \left(\begin{bmatrix} X_1 \\ X_2 \end{bmatrix} \right), \Phi \left(\begin{bmatrix} Z_1 \\ Z_2 \end{bmatrix} \right) \right\rangle$$

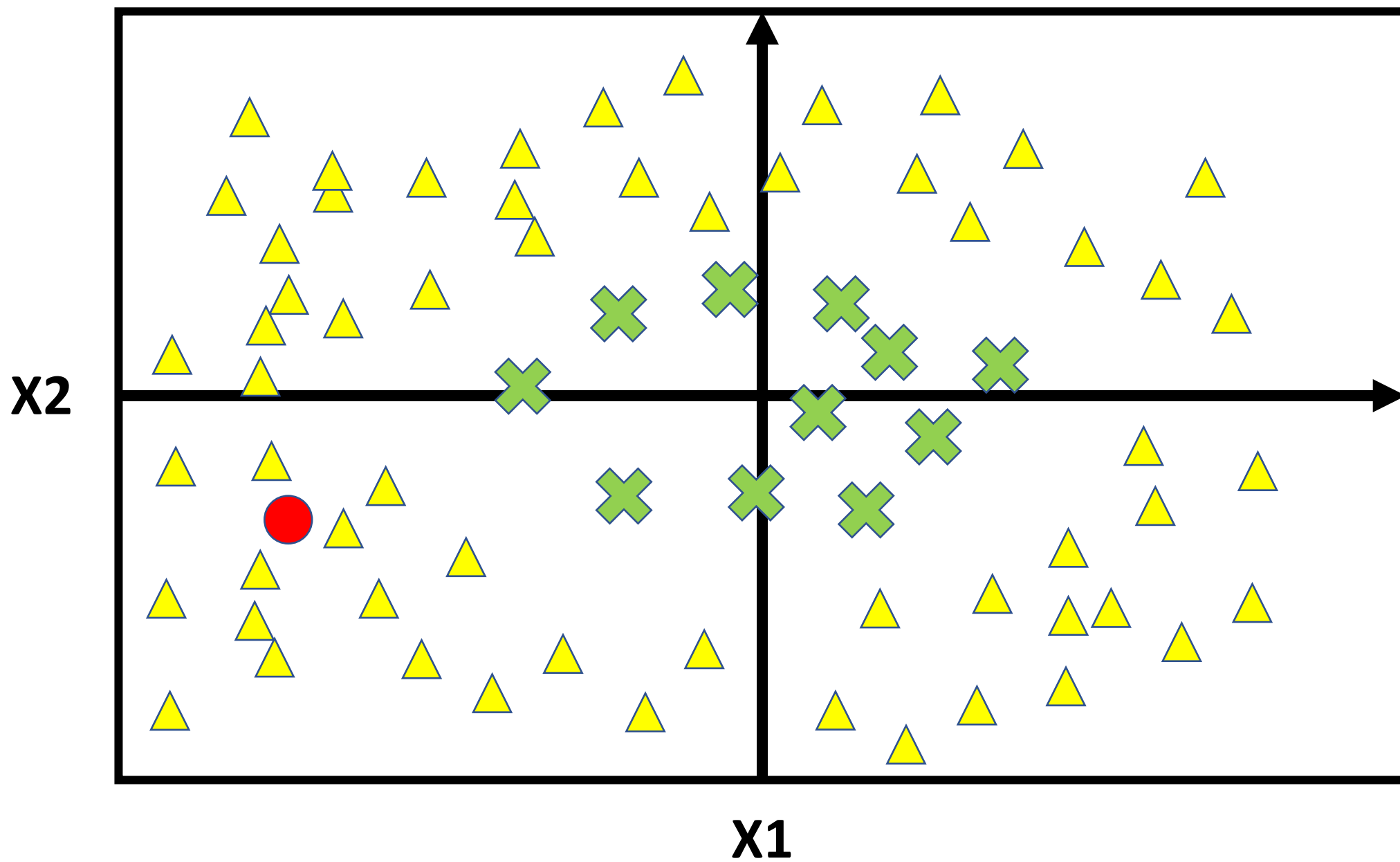
**Inherently
non-linear
models**

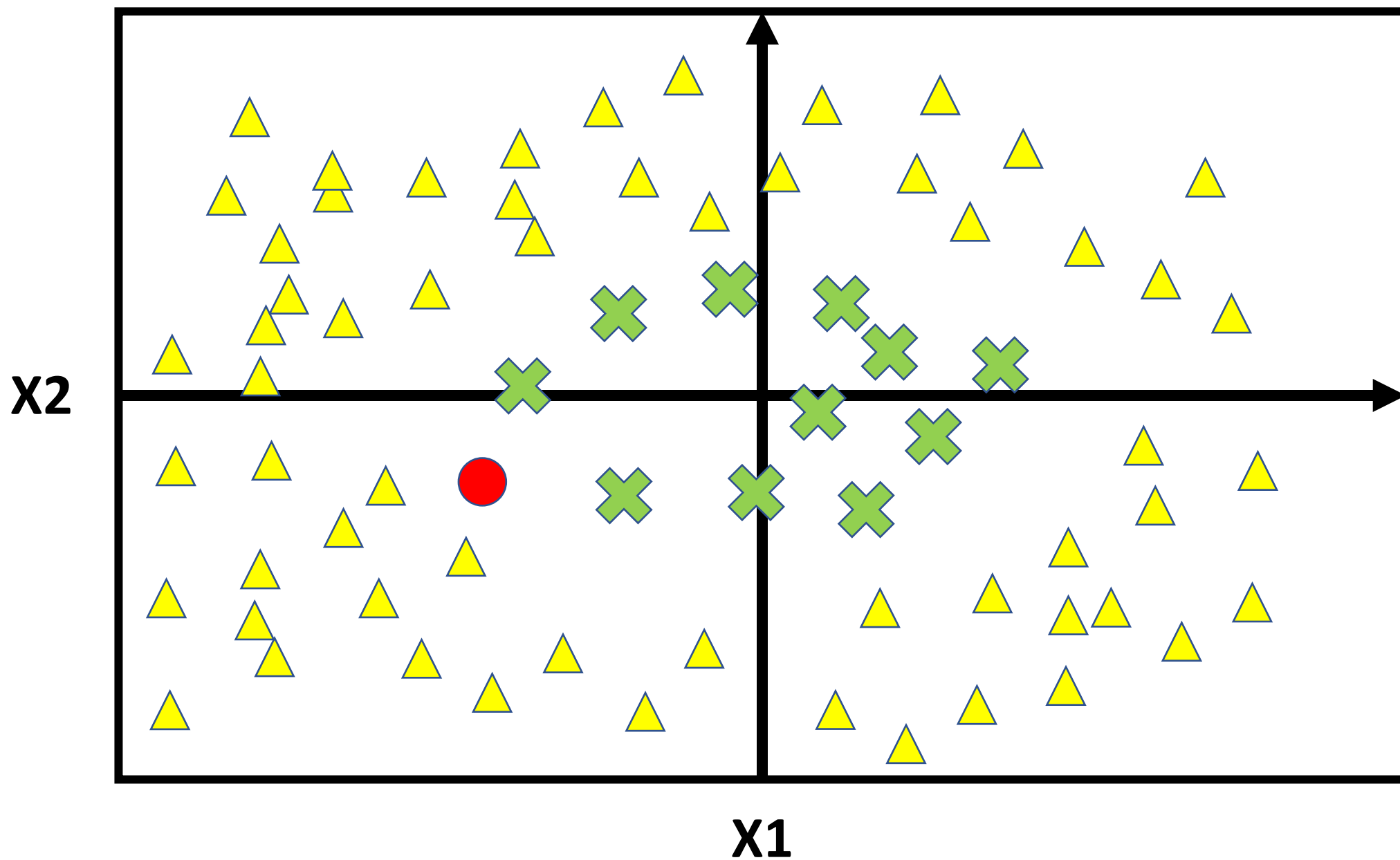
There are many different techniques ...



**K Nearest
neighbours
KNN**







Variables of the KNN model

- How many neighbours?
- Policy?
 - Majority
 - Weighted distance
- Metric

Learning vs memorizing

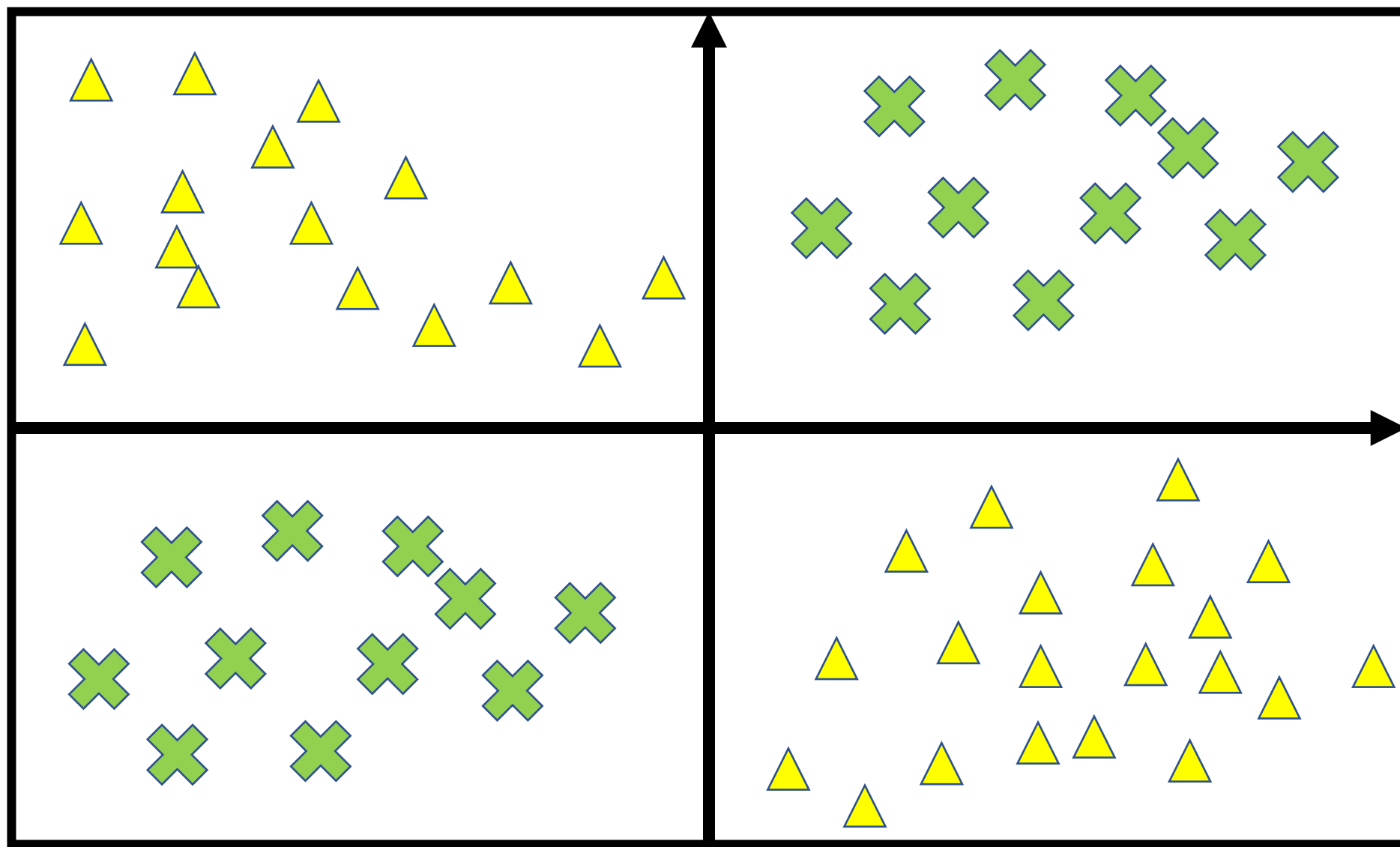
- Training is roughly equivalent to storing all the data points
- Prediction:
 - Cross-checking the input with the stored data points.

What happens if

- $k \rightarrow n_s$?
- $k \rightarrow 1$?

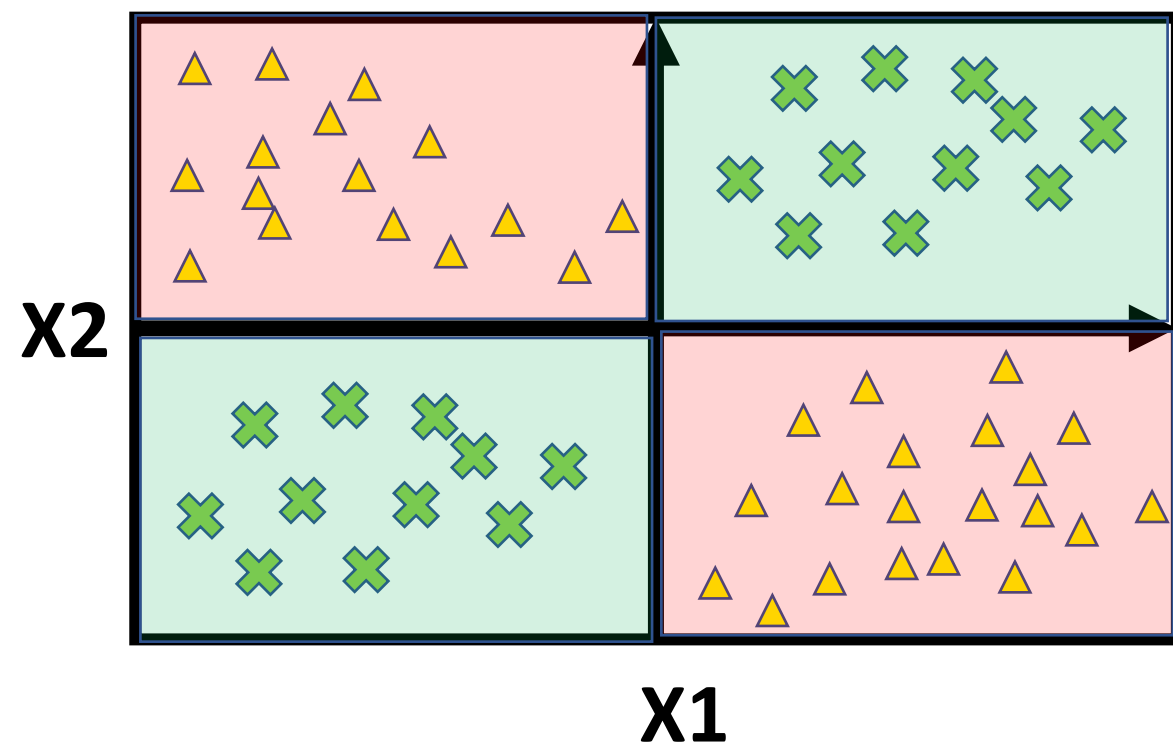
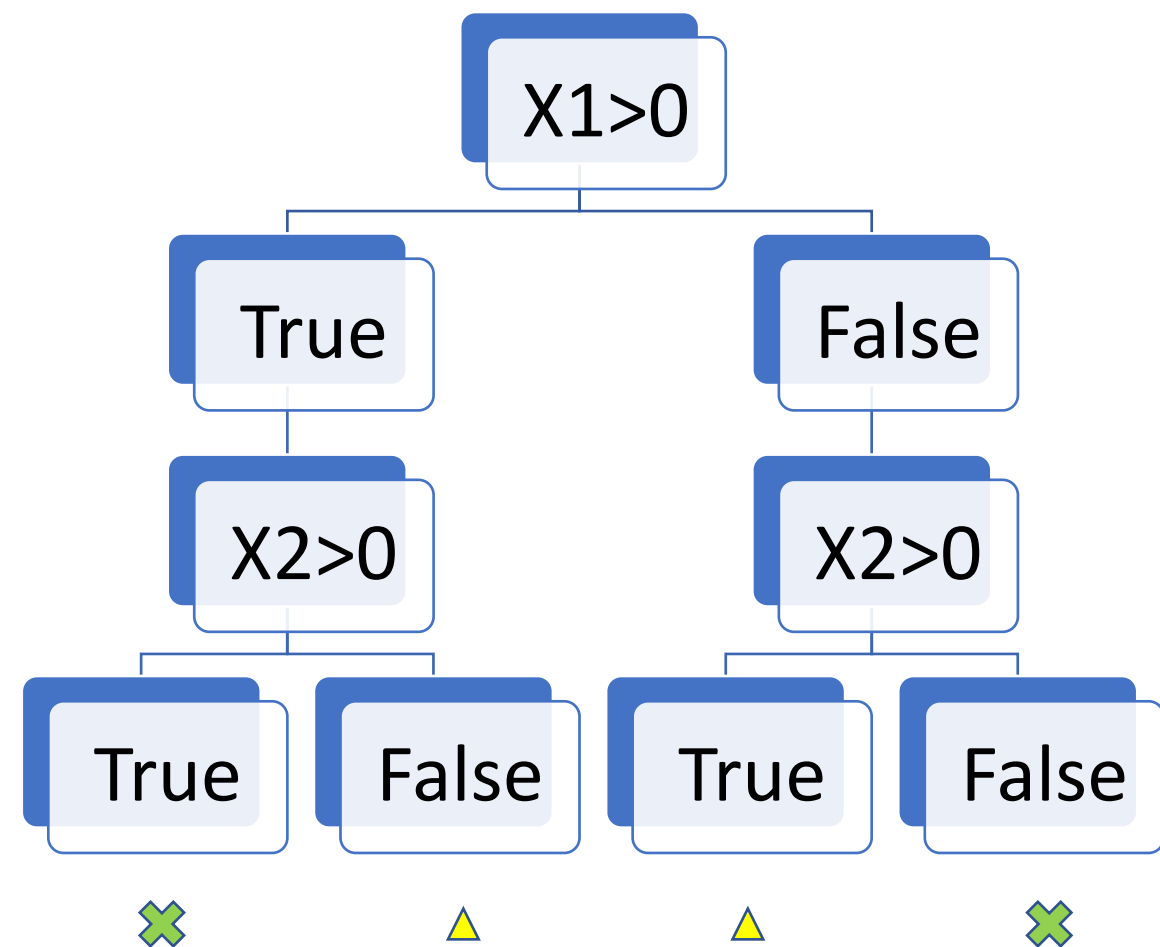
Decision Trees

X2

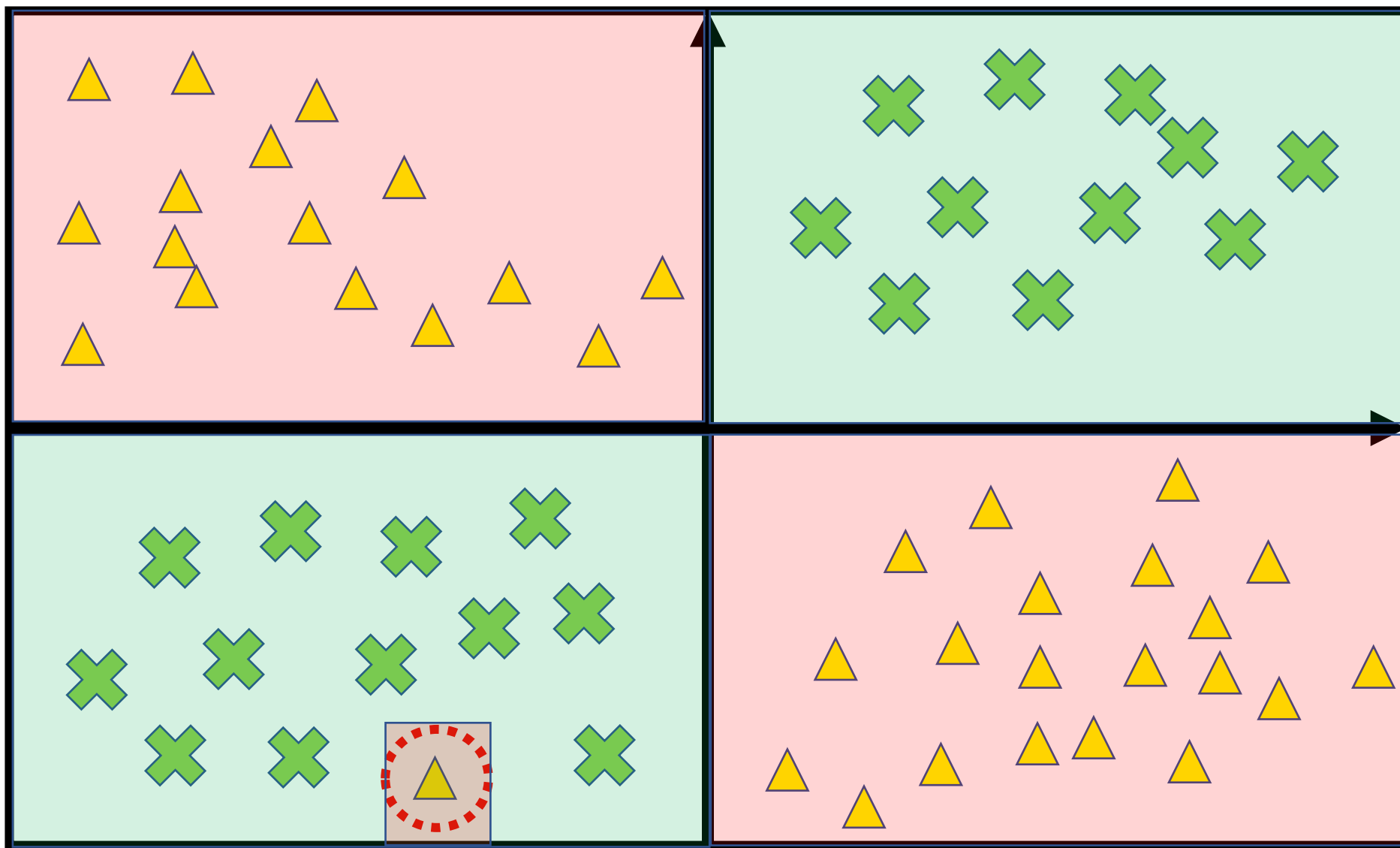


X1

Decision tree



X2



X1

What is the main variable?

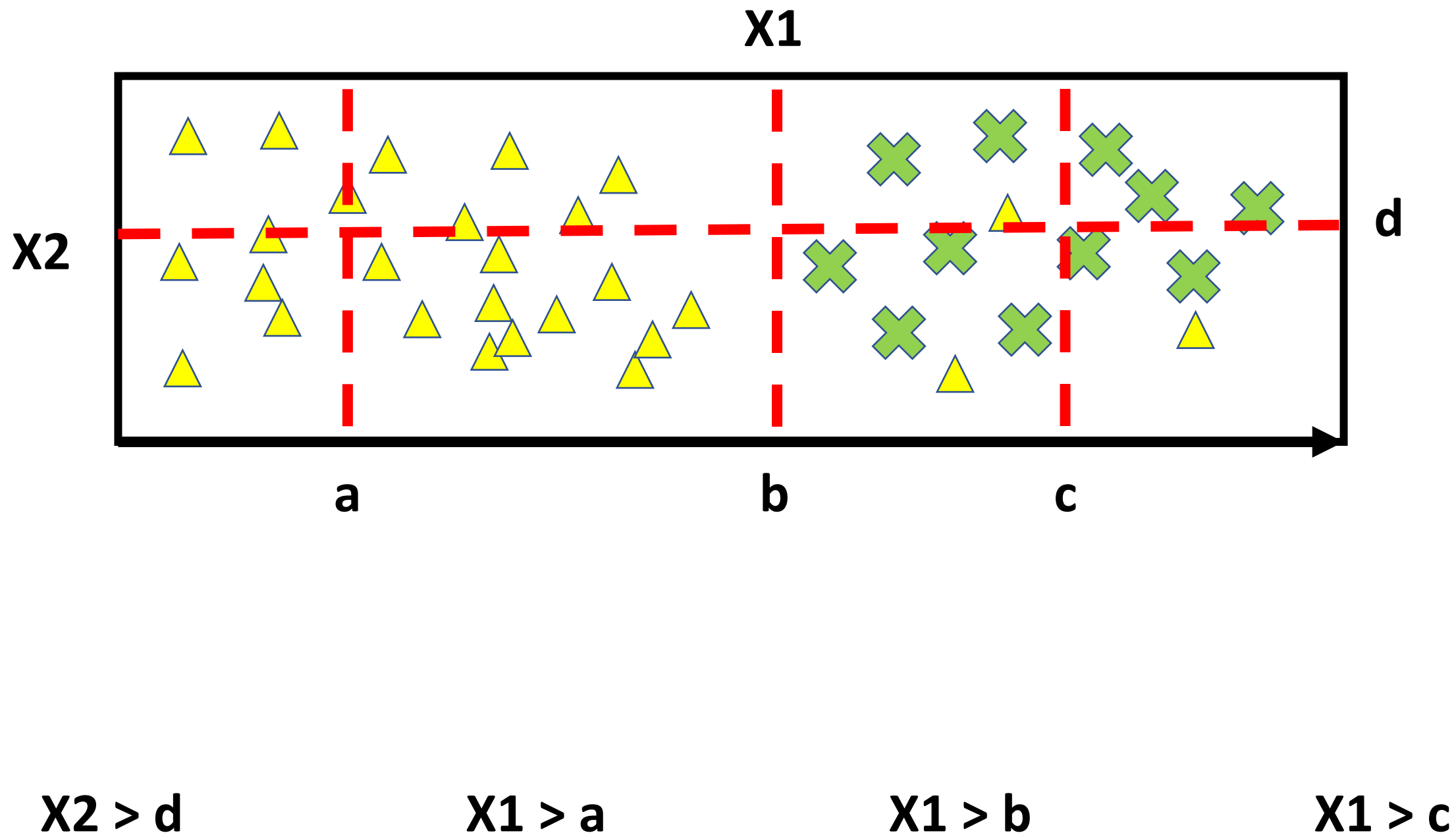
Depth of the tree

What happens if $depth \rightarrow \infty$?

What are we optimizing?

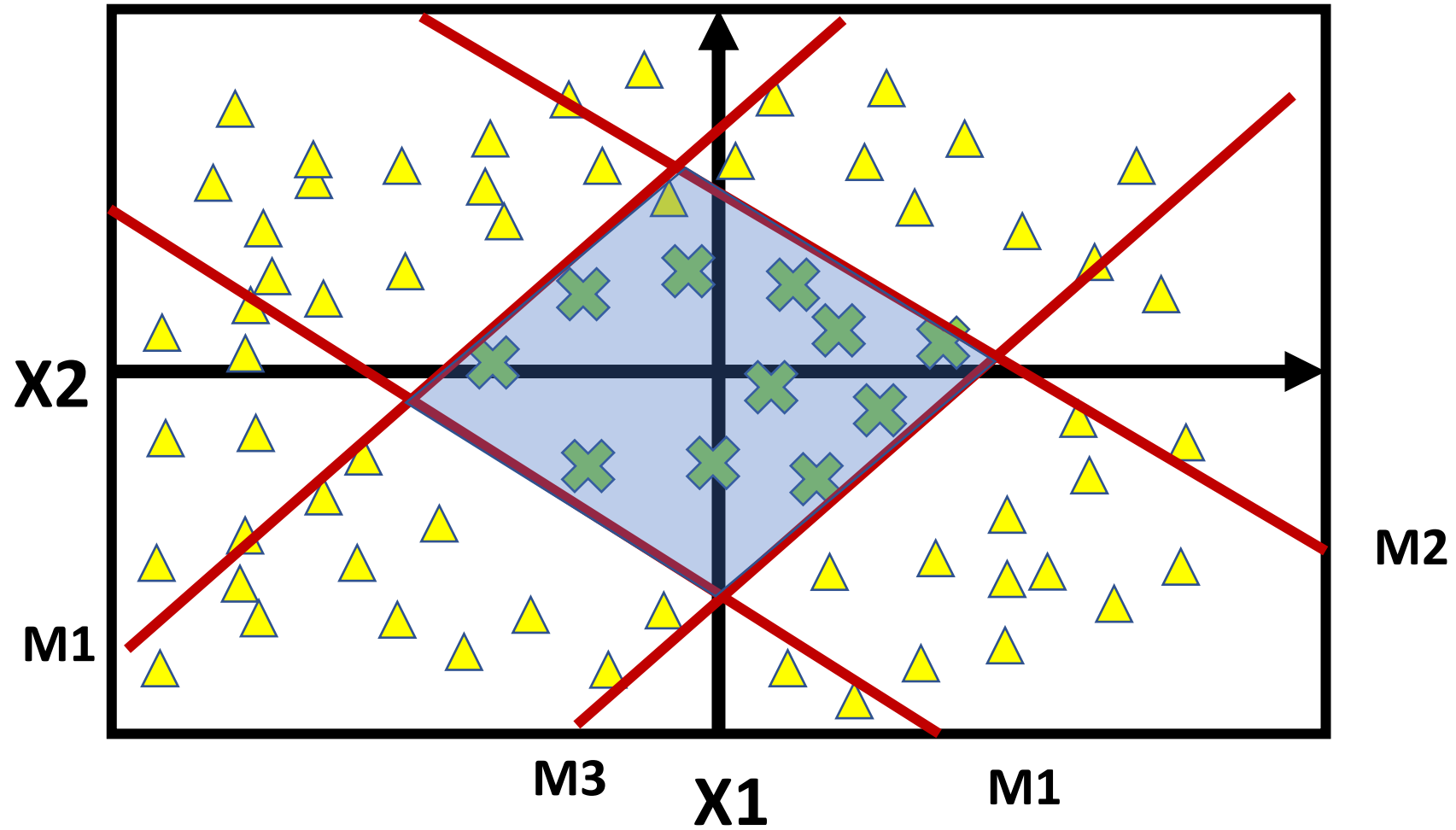
What is the objective?

- What is the most informative questions to ask?
 - Information gain
 - Variance reduction

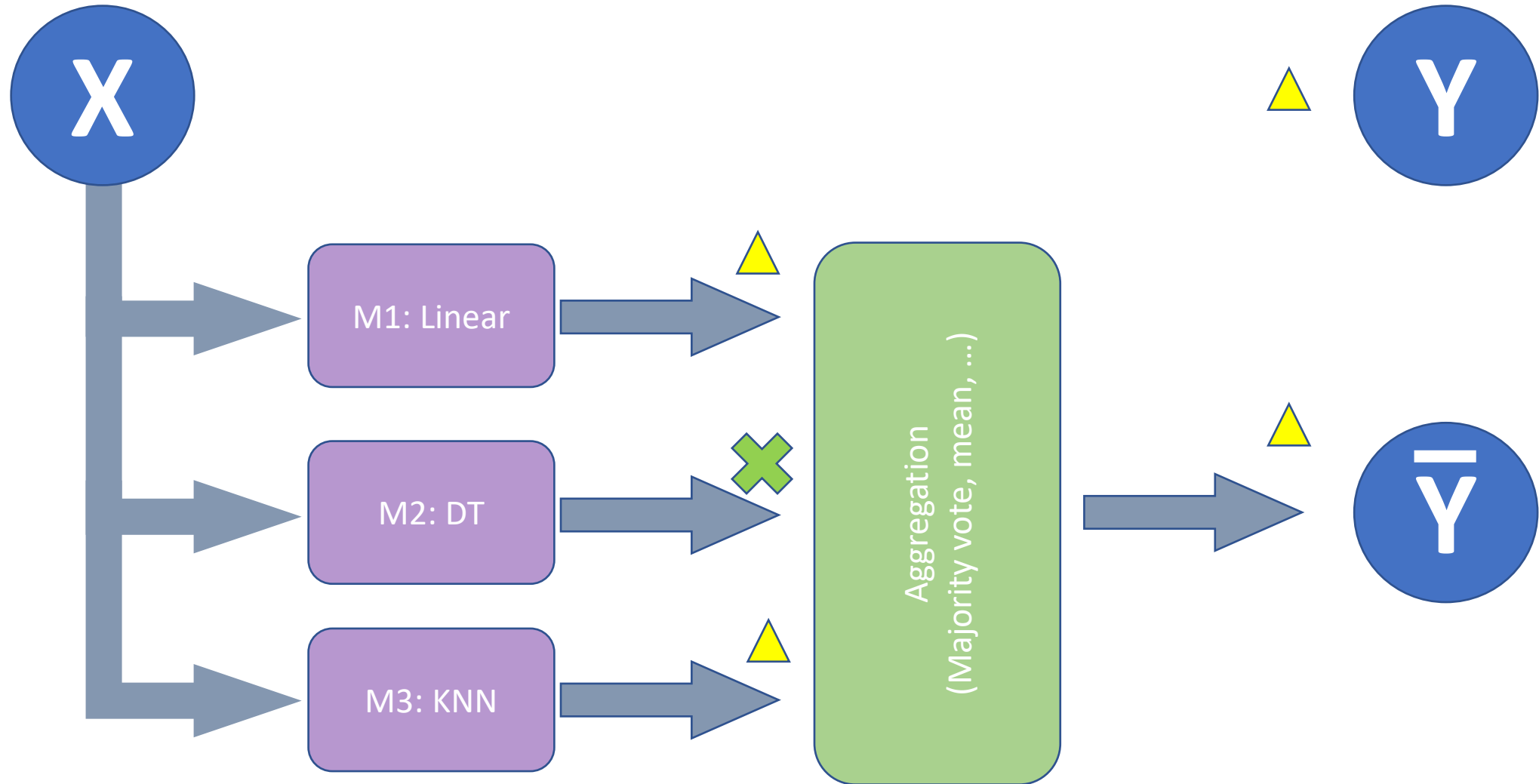


Ensemble Techniques

Intuition

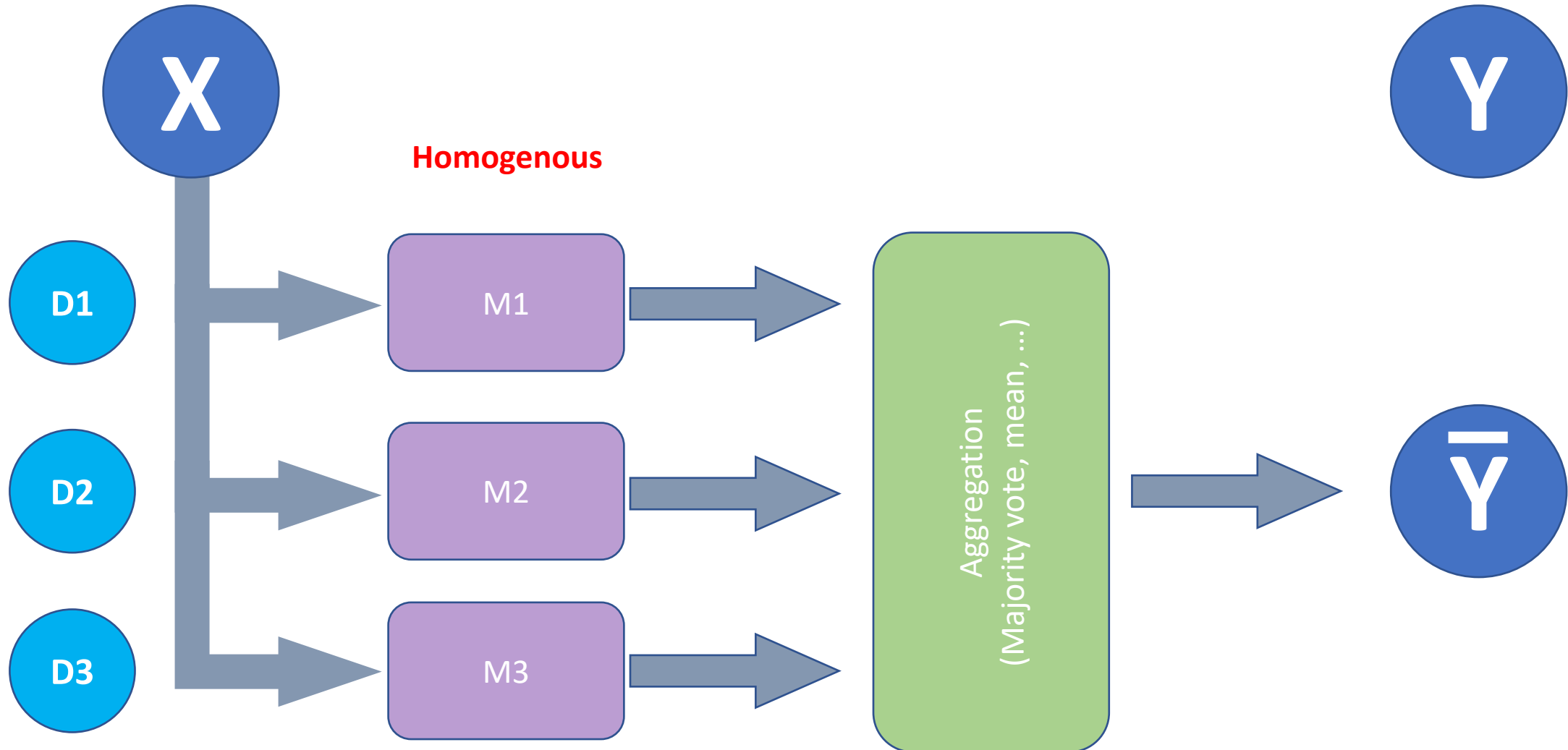


Simplest way to aggregate models



Bagging (Bootstrap aggregating)

D_i is a subset of X .





Example of Bagging: Random Forest

Boosting

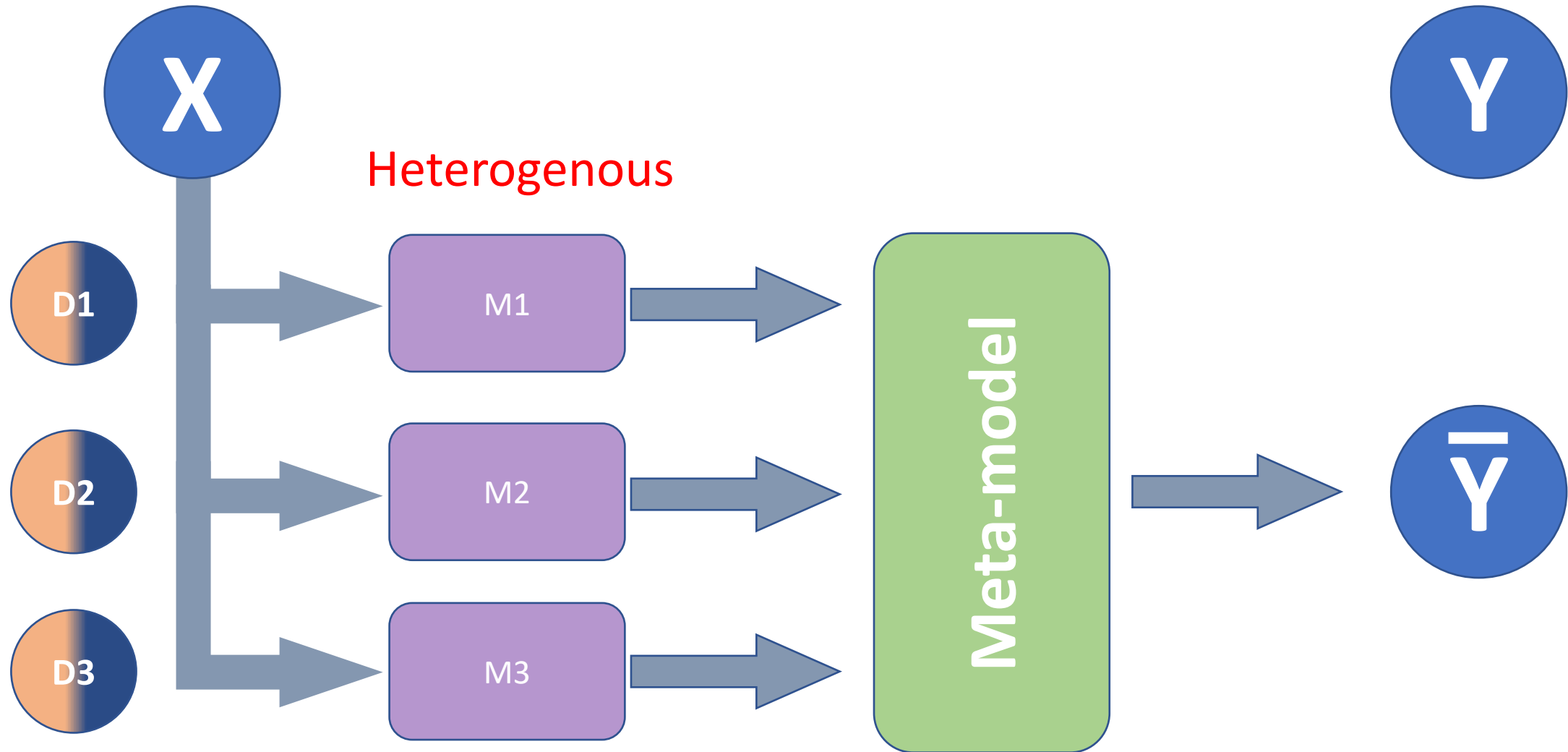
- Sequential improvement

Put more emphasis on the mislabelled samples and try to force the models to do better.

- Examples:

- Adaboost
- Gradient boosting

Stacking

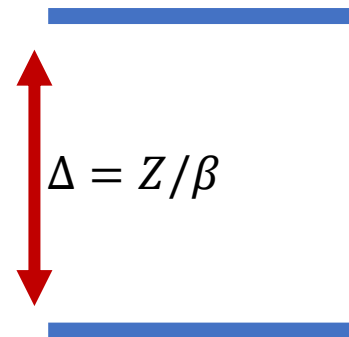


**Activation
Function:**

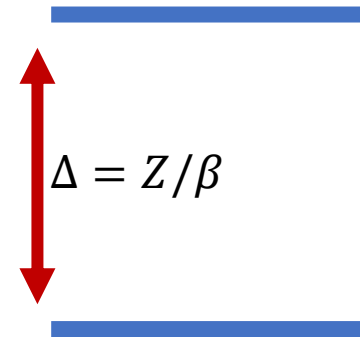
Logistic Regression

$$f_w \left(\vec{X}^i \right) = \text{Sigmoid} \left(\sum_j w_j X_j^{(i)} \right)$$

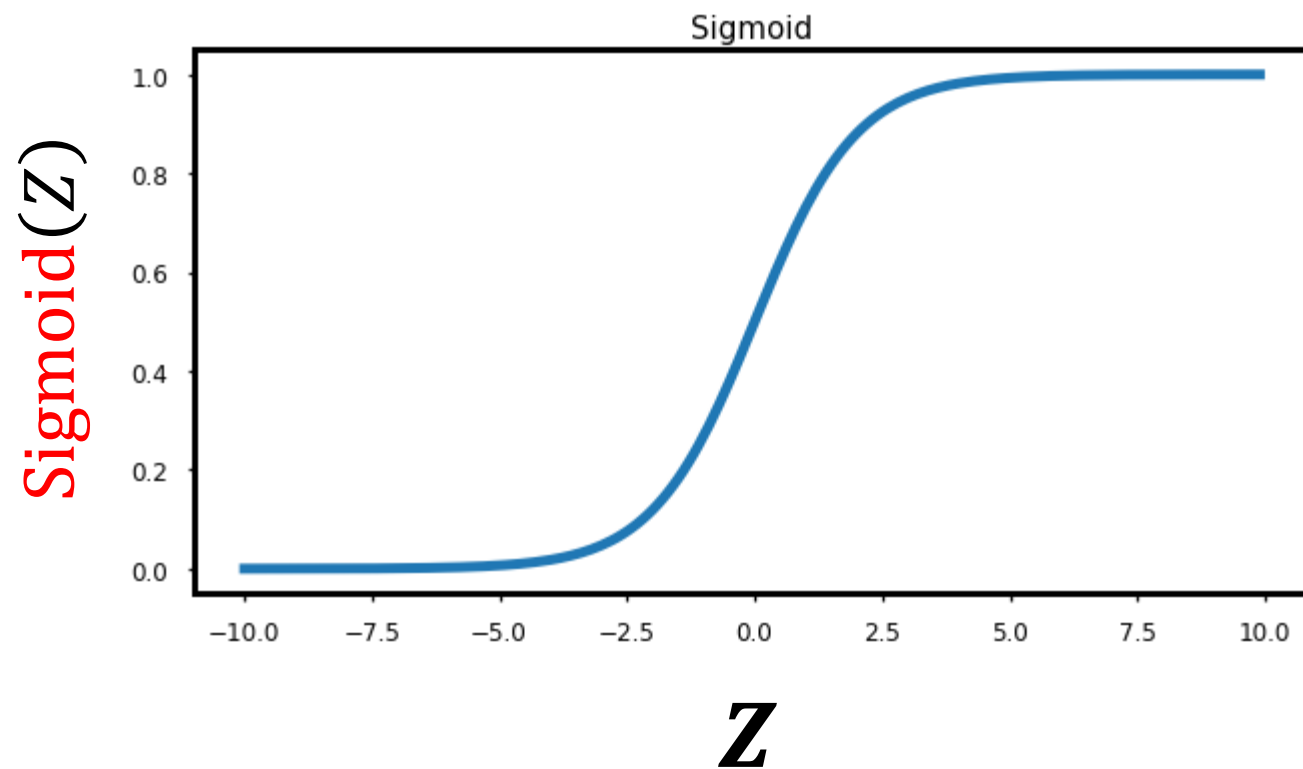
$$\text{Sigmoid}(Z) = \frac{1}{1 + e^{-Z}}$$



$$\text{Sigmoid}(Z) = \frac{1}{1 + e^{-Z}}$$



A diagram illustrating a two-level system. Two horizontal blue lines represent energy levels. A red double-headed vertical arrow between them is labeled $\Delta = Z/\beta$, representing the energy gap.



Is this a linear or non-linear model?

How can we use
activation functions
to make
non-linear Models?

Importance of Features

How can we define the importance of features?

Which one is more important?

Some ordering in the significance of different features

Linear models

$$f_w(\vec{X}^i) = \sum_j w_j X_j^{(i)}$$



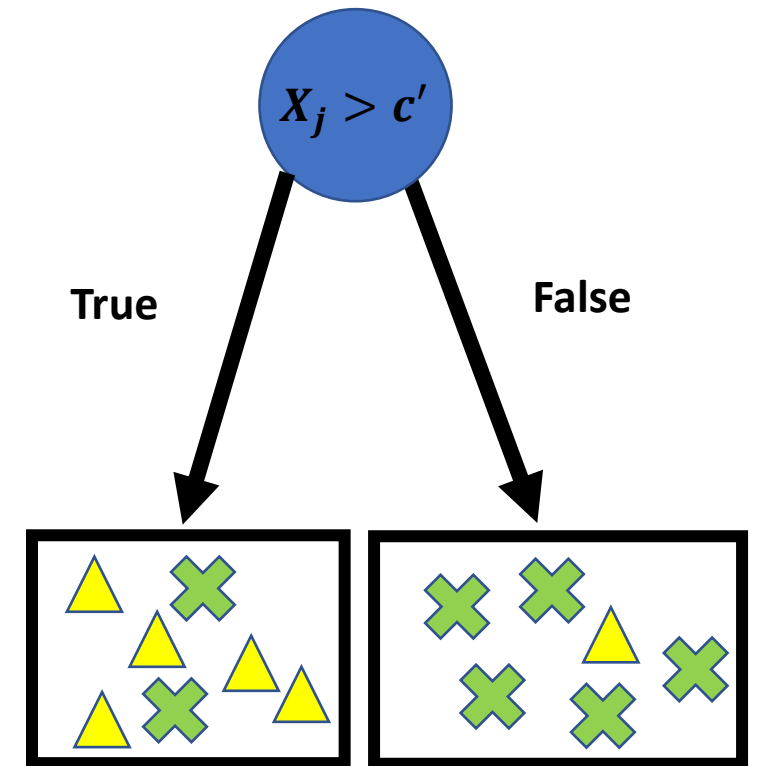
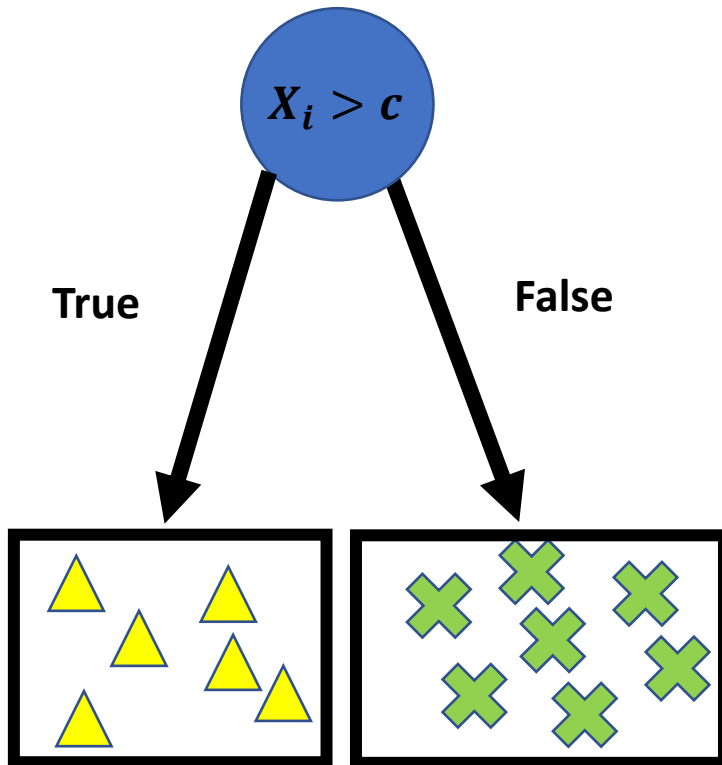
Decision trees

How can we define feature importance?

Information gain/Change in variance

Information gain/Change in variance

Which is better?



KNN

How would you define feature importance for KNN?

Which Model?

Comparison

	Linear	DT	KNN	Random Forest
Performance				
Idea				
Training time				
Prediction time				
Explainability/ Interpretability				

So far ...

