Small data technique

Lecture 14

Changho Suh

January 26, 2024

DTs for regression, challenge of DTs & ensemble learning

Outline

1. Study DTs for **regression**.

2. Investigate a challenge that arises in DTs.

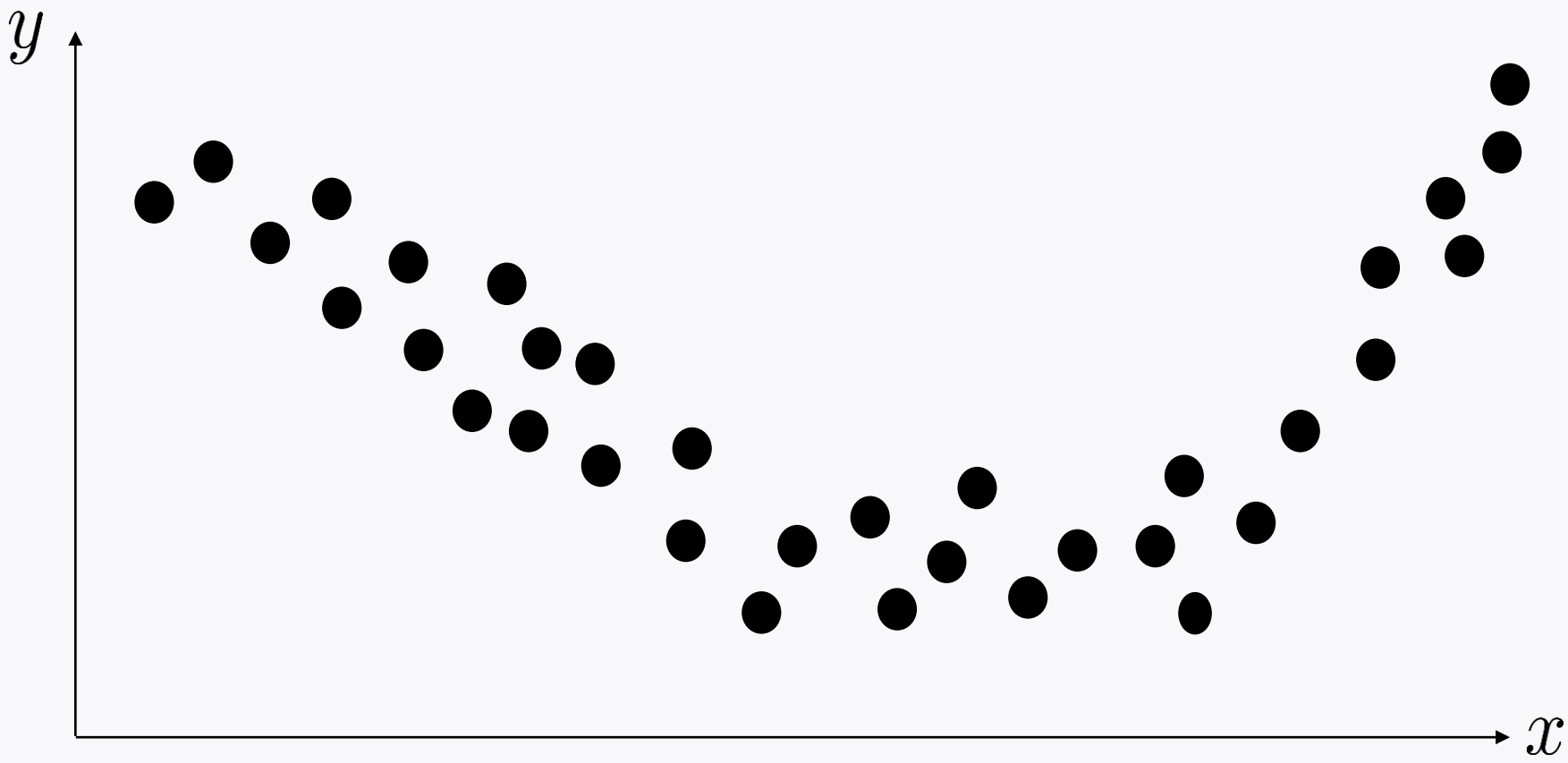
3. Explore a way to address the challenge:

Ensemble learning

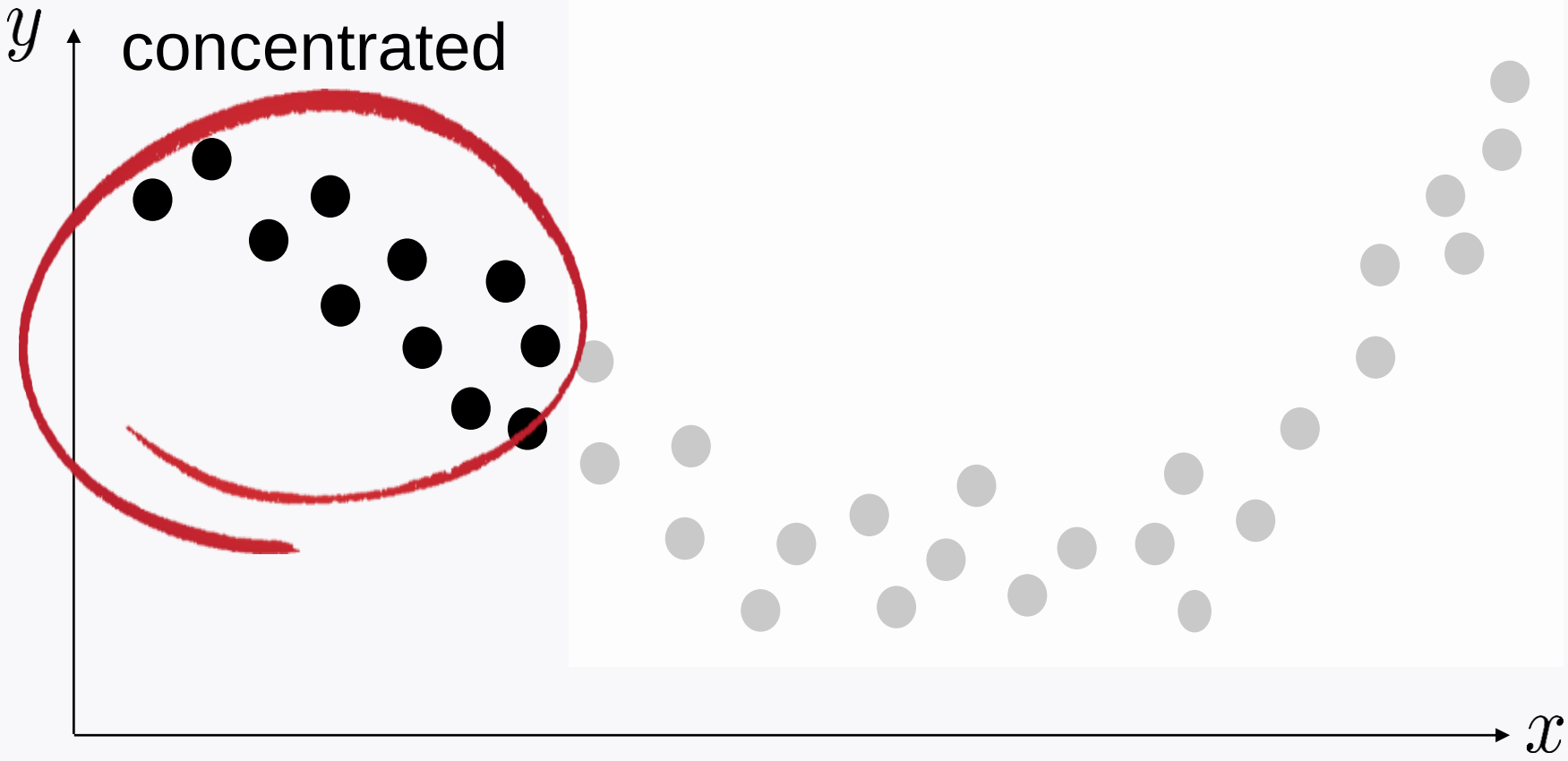
DTs for regression

A motivating example

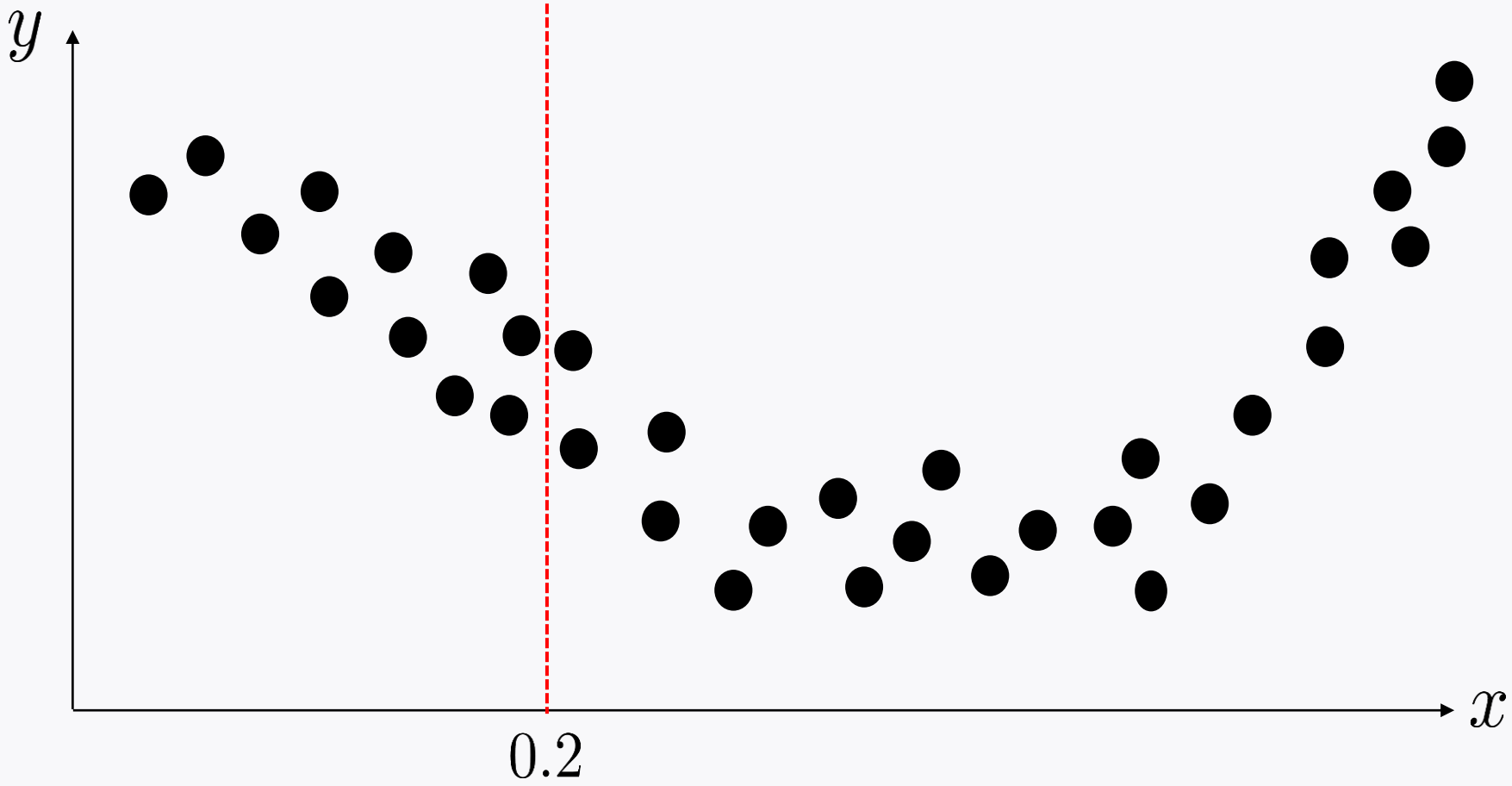
$x \in \mathbf{R} \quad y \in \mathbf{R}$



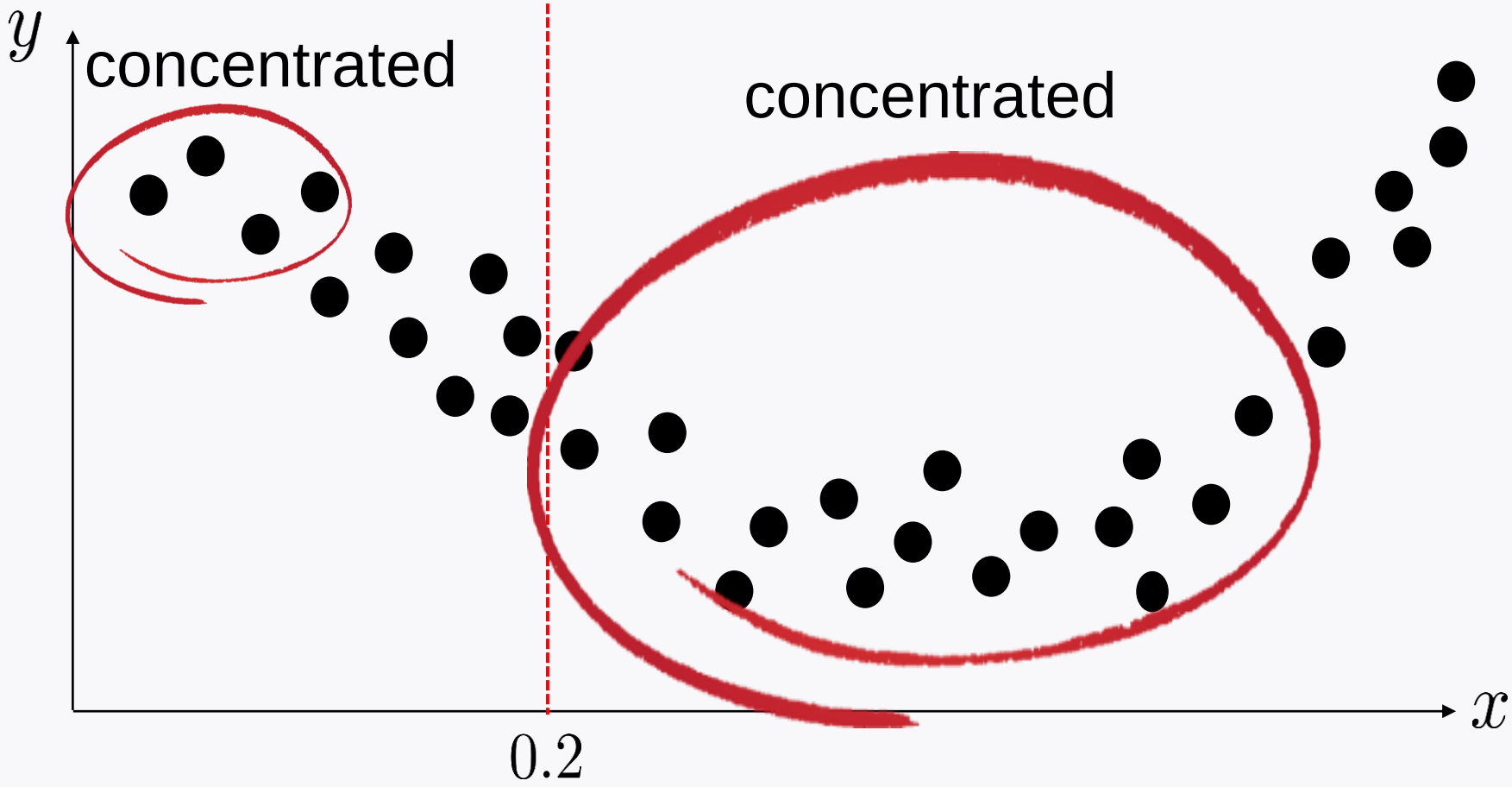
Observation



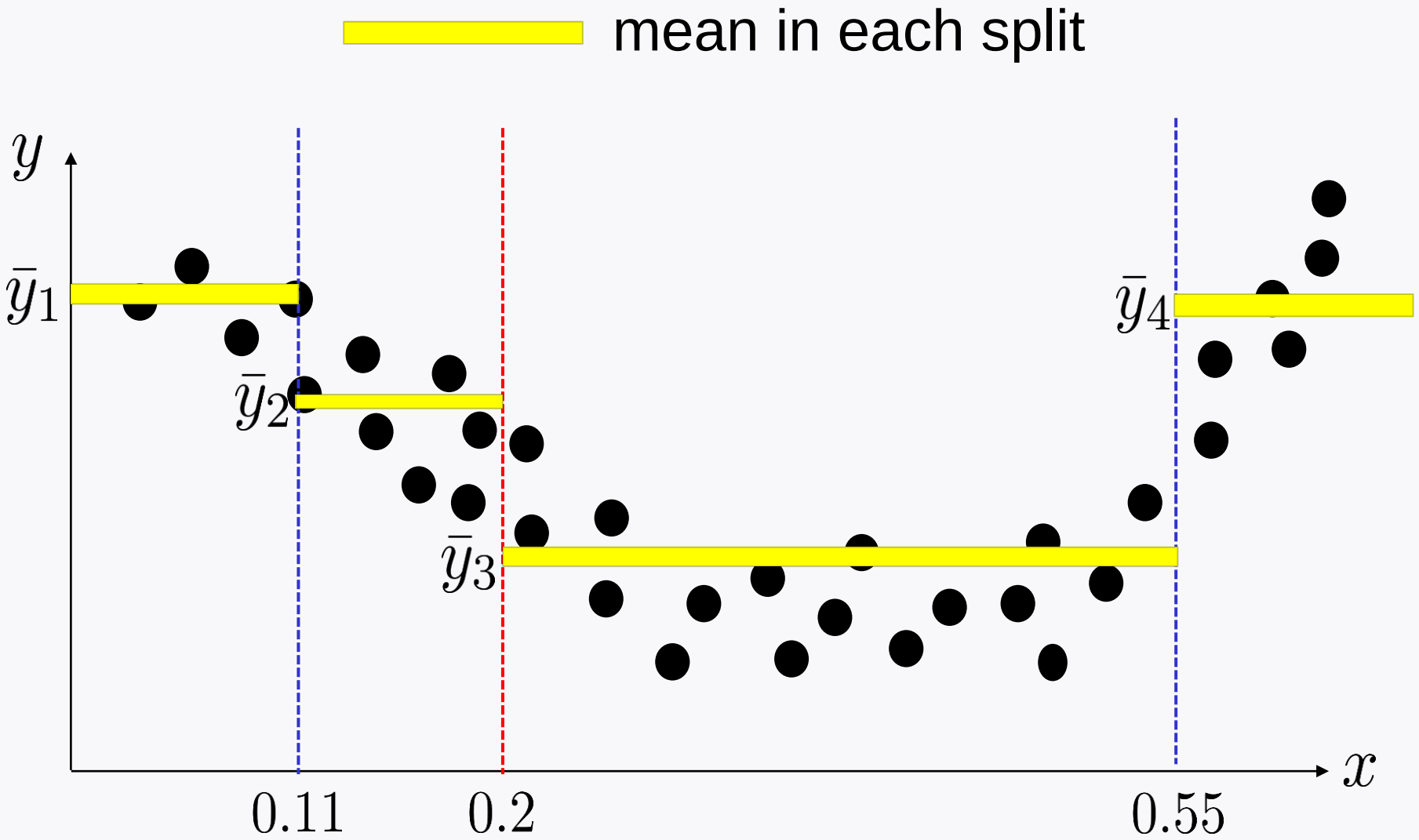
A natural attempt for separation



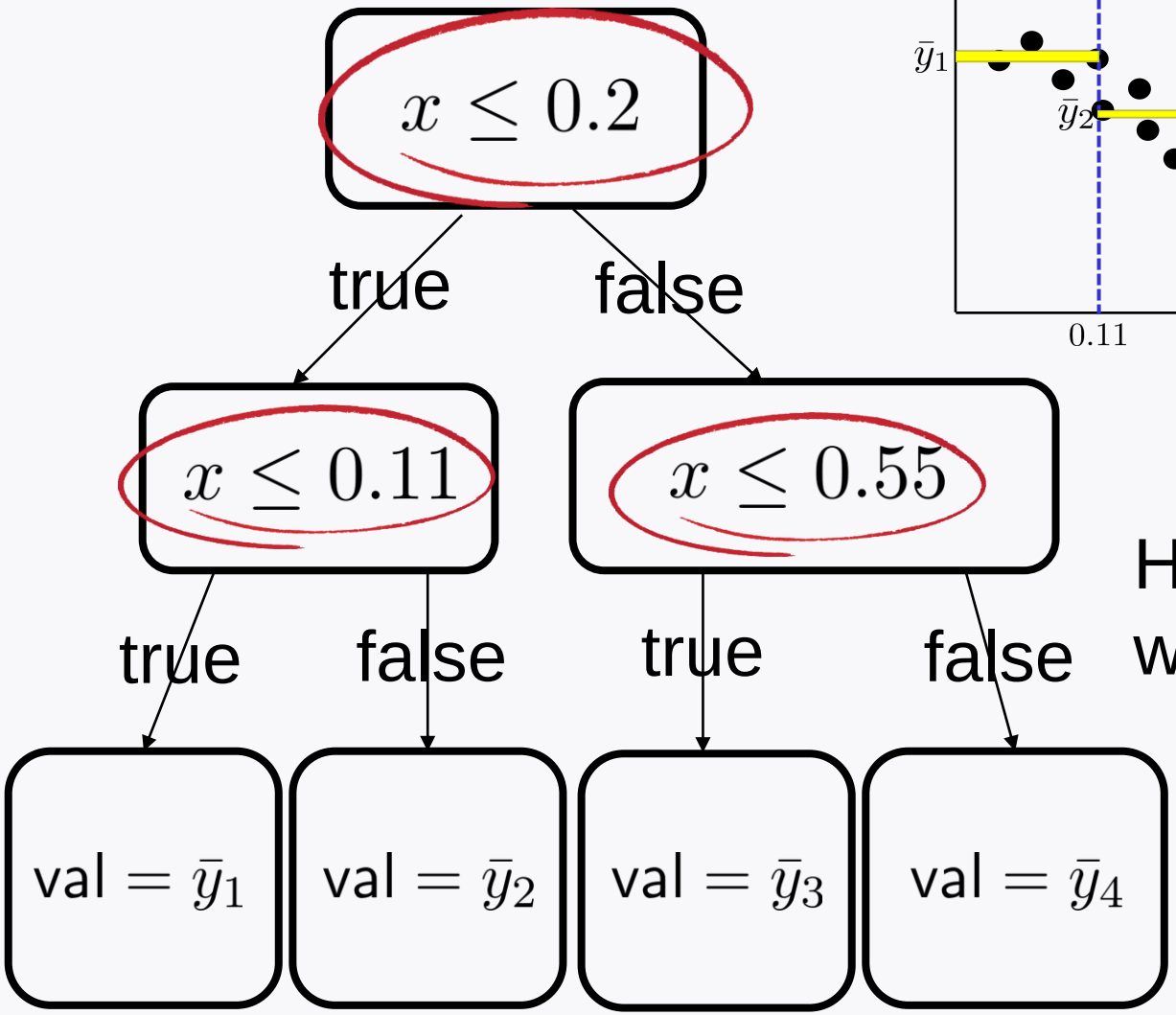
Observation in each split

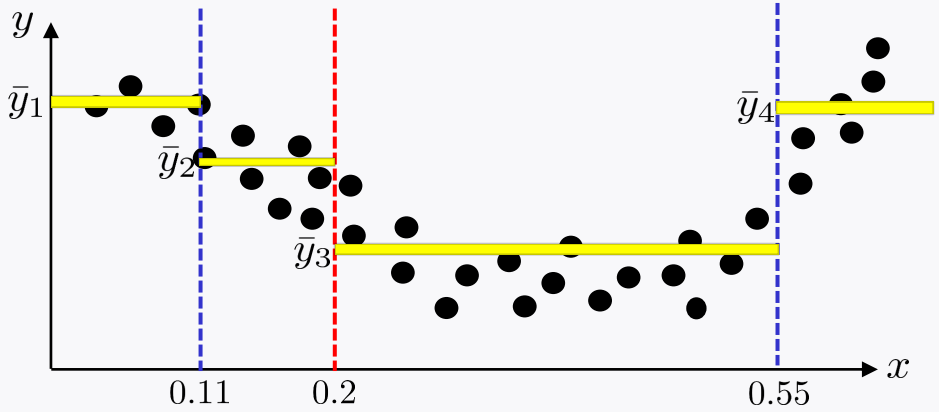


A follow-up natural attempt



Decision tree



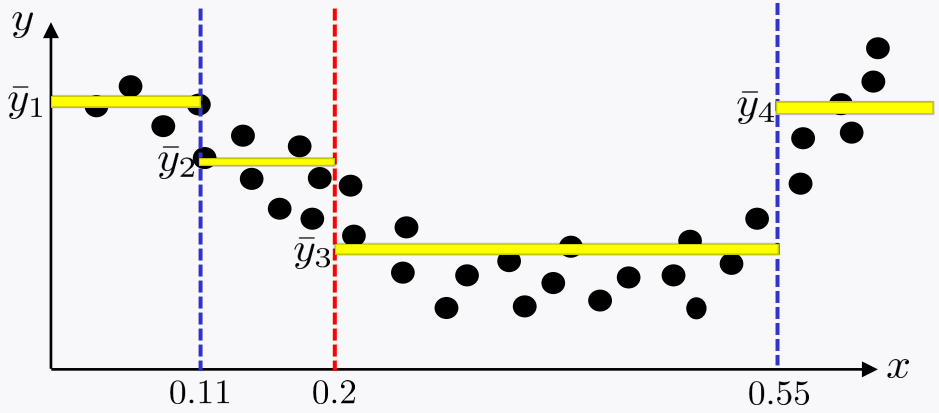


How to come up with the conditions?

CART algorithm

k : feature index

t_k : threshold



Step 1: Find (k, t_k) such that $J(k, t_k)$ is minimized.

$$J(k, t_k) = \frac{m_{\text{left}}}{m}\text{MSE}_{\text{left}} + \frac{m_{\text{right}}}{m}\text{MSE}_{\text{right}}$$

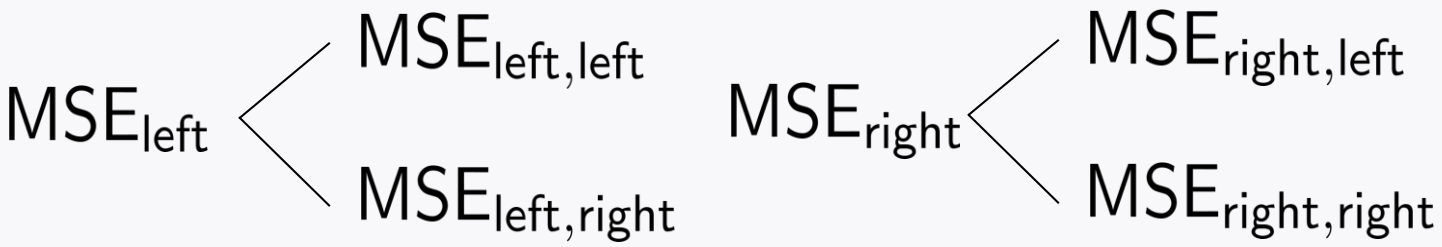
$$\text{MSE}_{\text{left}} := \frac{1}{m_{\text{left}}} \sum_{i \in \text{left}} (y^{(i)} - \bar{y}_{\text{left}})^2 \qquad \bar{y}_{\text{left}} = \frac{1}{m_{\text{left}}} \sum_{i \in \text{left}} y^{(i)}$$

CART algorithm

Step 1: Find (k, t_k) such that $J(k, t_k)$ is minimized.

$$J(k, t_k) = \frac{m_{\text{left}}}{m}\text{MSE}_{\text{left}} + \frac{m_{\text{right}}}{m}\text{MSE}_{\text{right}}$$

Step 2: Repeat Step 1 for each split:



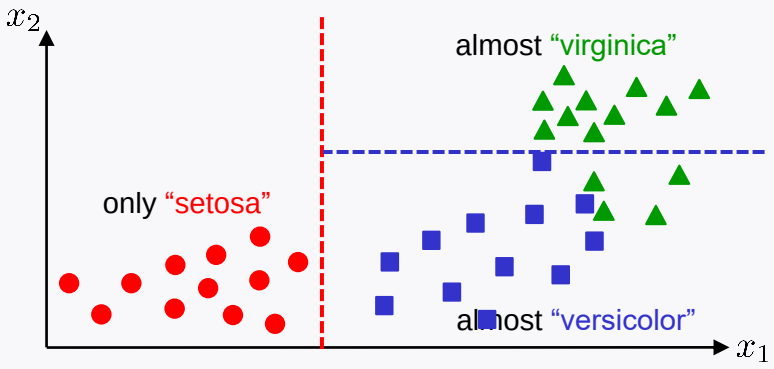
Stopping criteria & hyperparameters are the same as those of classification.

Challenge of DTs

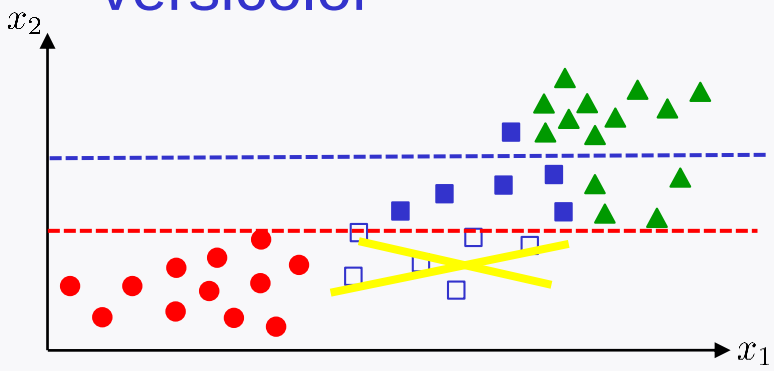
Challenge

Sensitive to **small variations** of training data.

Example:



remove some points of versicolor



A solution to address variation sensitivity

Turns out:

Ensemble learning is a solution.

For the rest:

1. Study what **ensemble learning** is.
2. Study ond powerful ensemble method:

Random forests (RFs)

Ensemble learning

Debate on a decision

How to decide when we have *diverse* opinions?

Often rely on **majority voting**.

Wisdom of the crowd: An aggregated decision is often better than even an expert's answer.

Can expect in the predictor context:

An aggregating prediction based on many predictors

→ A better prediction relative to the best predictor.

Ensemble learning

Ensemble: A *group* of predictors

Ensemble learning:
A technique that aggregates predictions of the ensemble.

Hard voting: Declare the one that gets **most votes**.

Soft voting: Declare the one with **highest probability** averaged over predictors

A way to obtain ensemble

Train each predictor on a **different subset** of the training set.

How to construct different subsets?

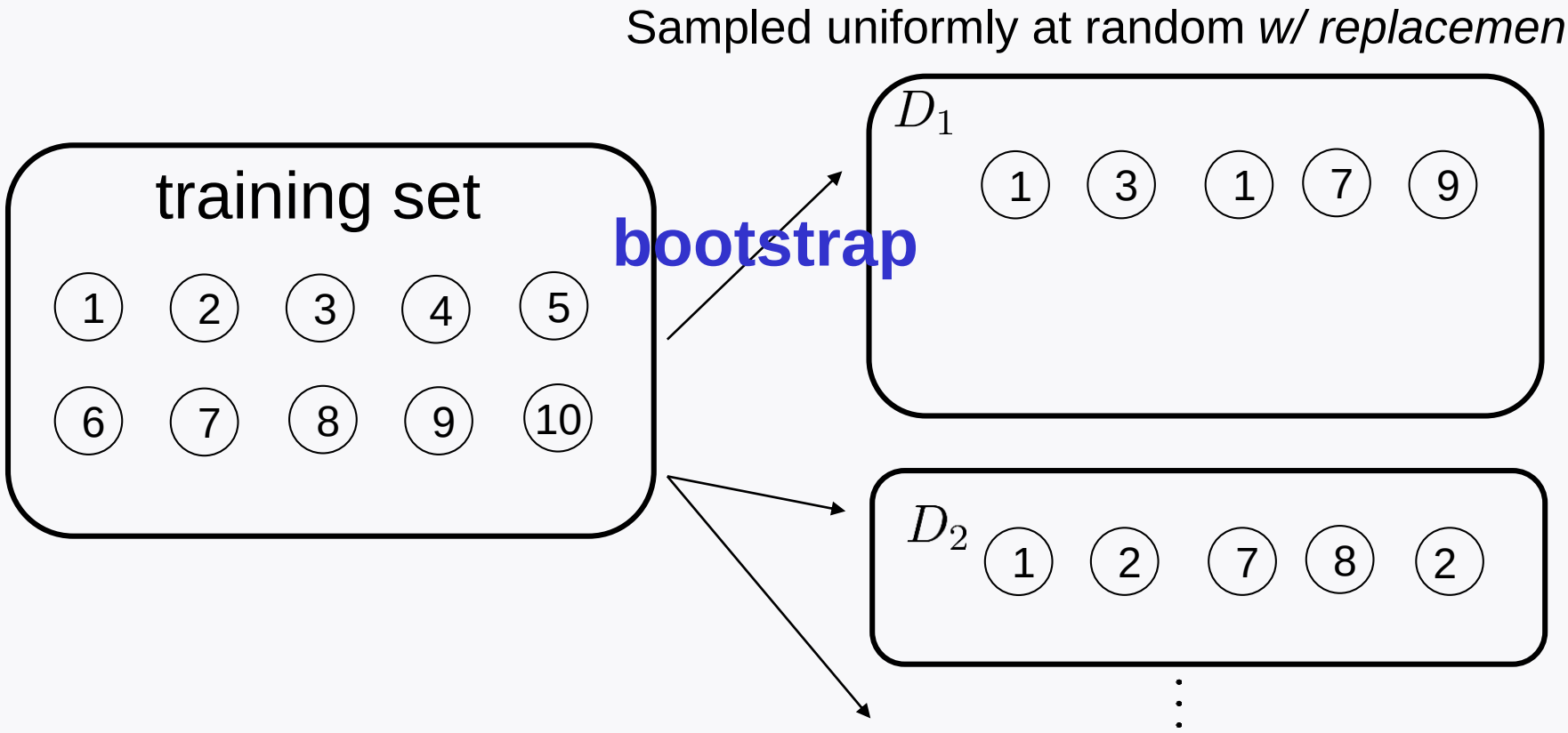
1. A way to choose *partial examples*:

   Bootstrap

2. A way to choose *partial features*:

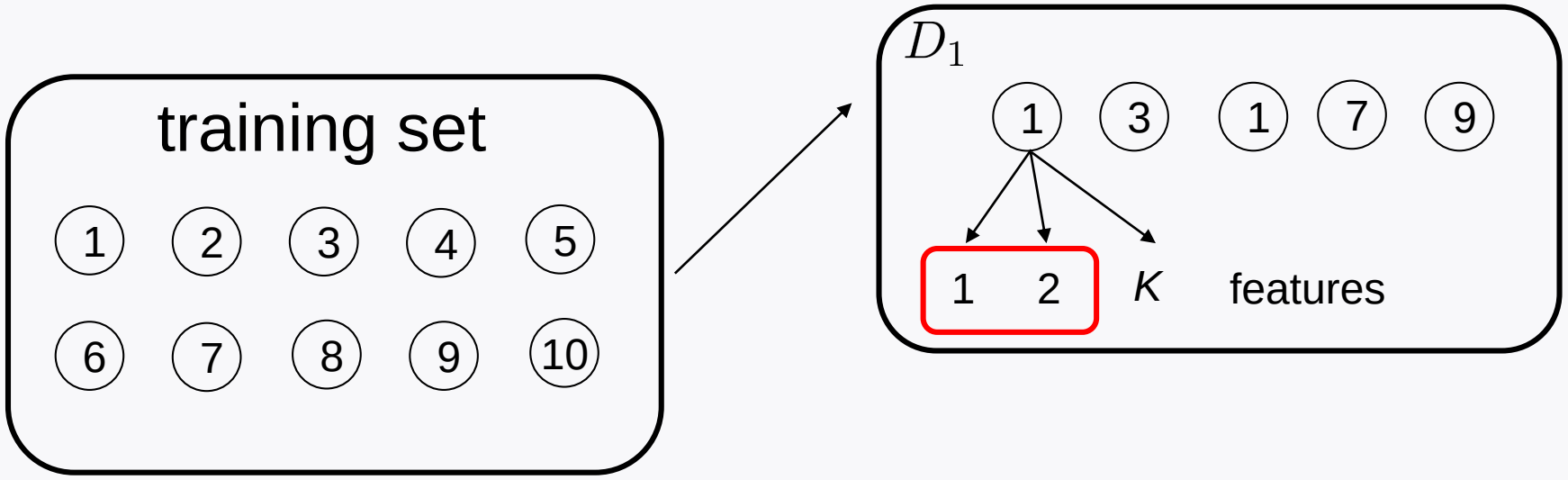
   Random subspace method

RF = Bootstrap + random subspace

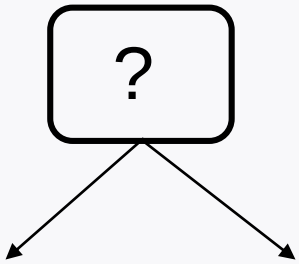


RF = Bootstrap + random subspace

Sampled uniformly at random *w/ replacement*

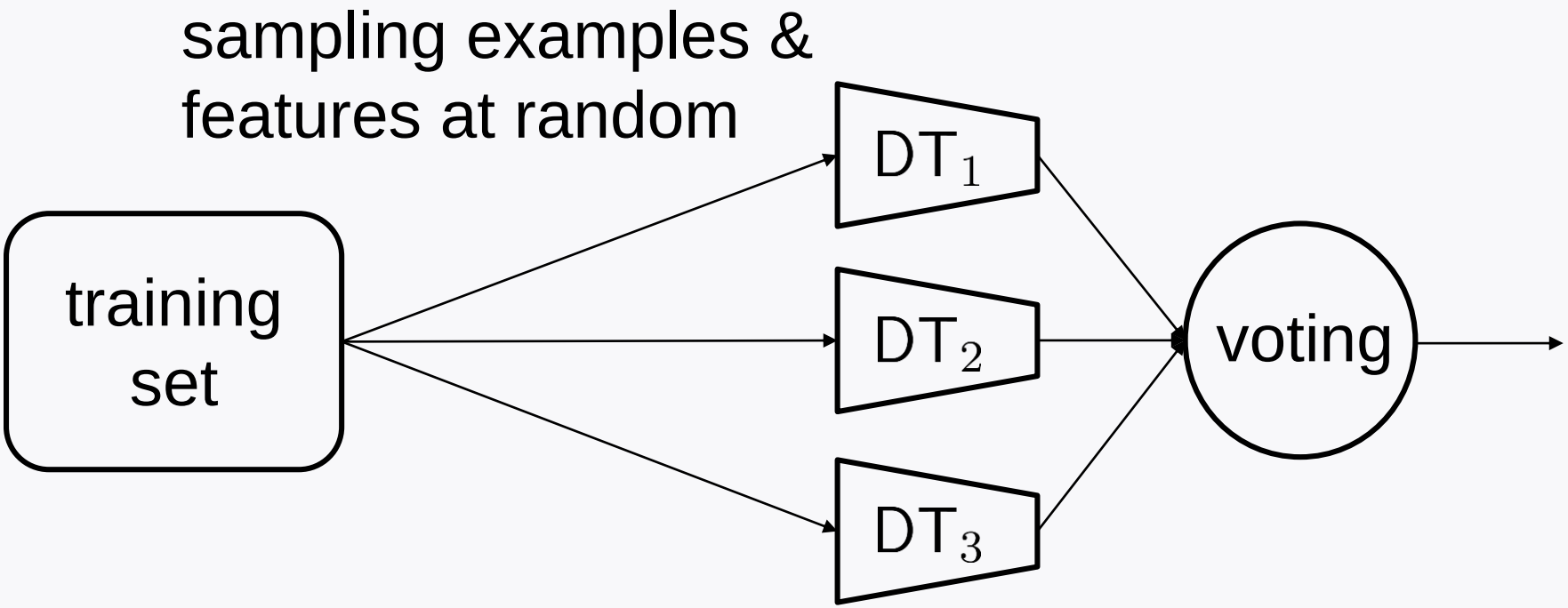


Decision Tree (DT) w/ D_1



Split a node considering a **random subset of features**.

RF in picture



Look ahead

Study **RF** in depth:

1. Investigate **hyperparameters**;

2. Study a measure for model *interpretation*:
   Feature Importance