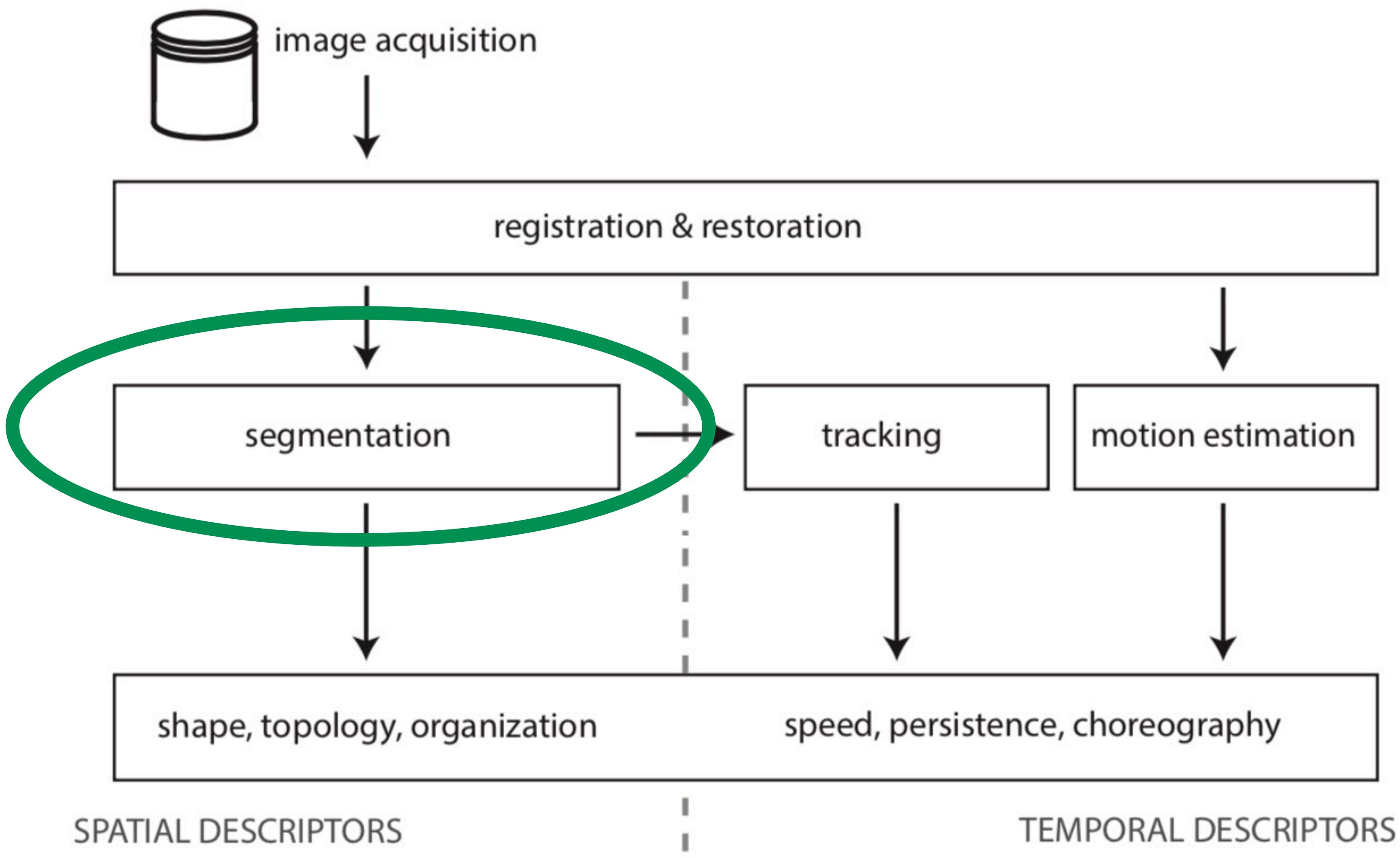
Principles of Image Processing & Quantification: **automatic** segmentation

Mauricio Cerda/Jorge Jara
ICBM, Faculty of Medicine
Universidad de Chile
mauriciocerda@med.uchile.cl
www.scian.cl

IMAGE SEGMENTATION: SUMMARY

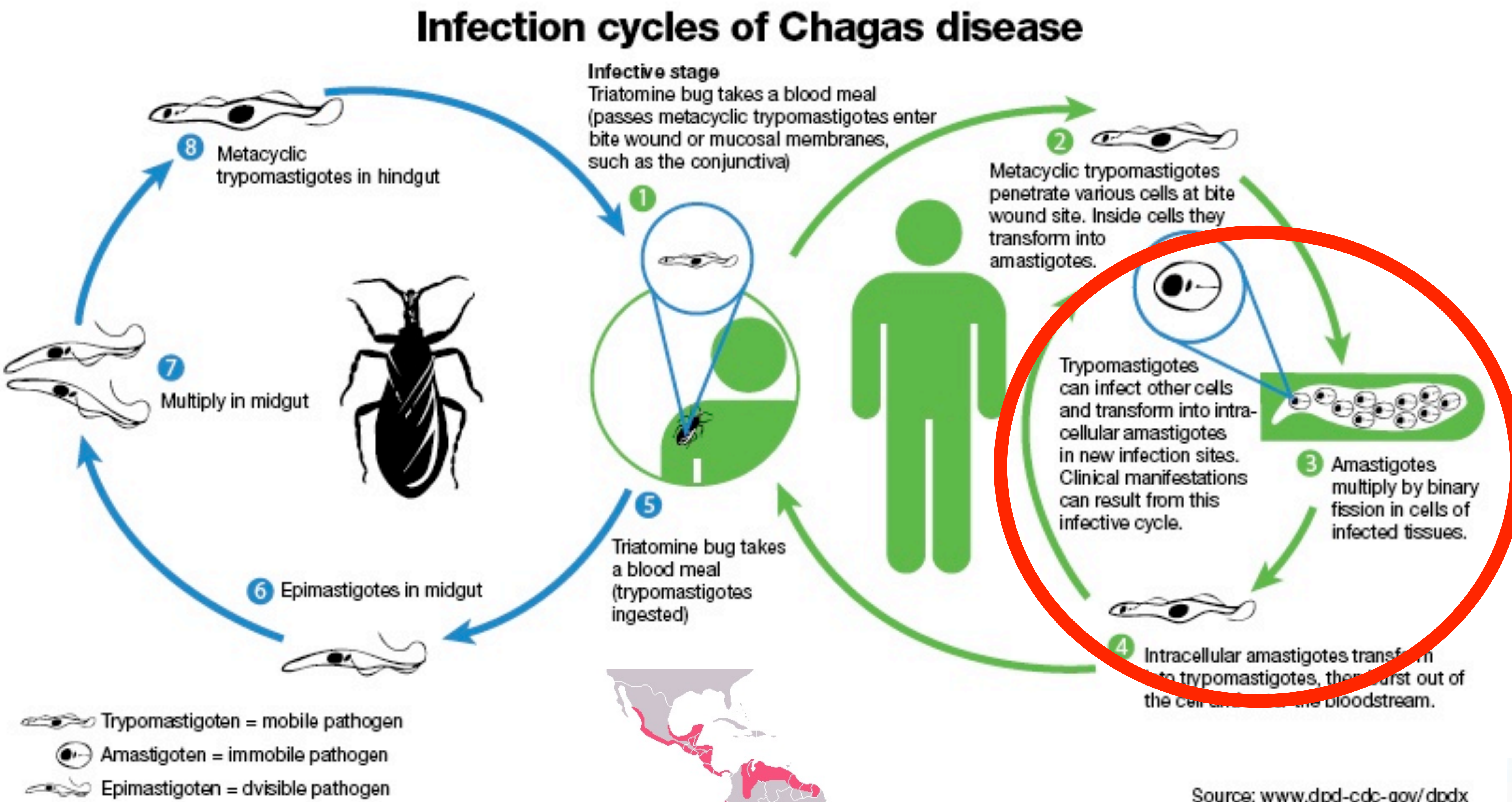


Computational Methods for Analysis of Dynamic Events In Cell Migration.
Castañeda V, Cerda M, Santibañez F, Jara J, Pulgar E, Palma K, Lemus CG, Concha M, Härtel S
Current Molecular Medicine, 14(2), 291–307.
Show Abstract

- Segmentation (clustering)
- Segmentation (random forest)

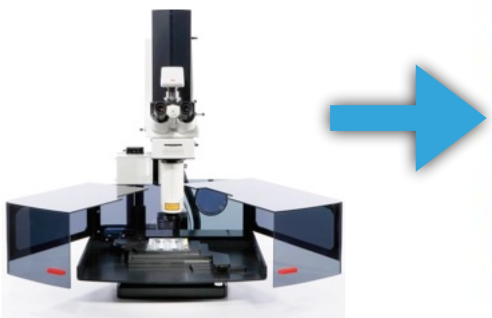
- http://fiji.sc/

IMAGE PROCESSING: CHAGAS



Source: www.dpd-cdc-gov/dpdx

IMAGE PROCESSING: PARASITES



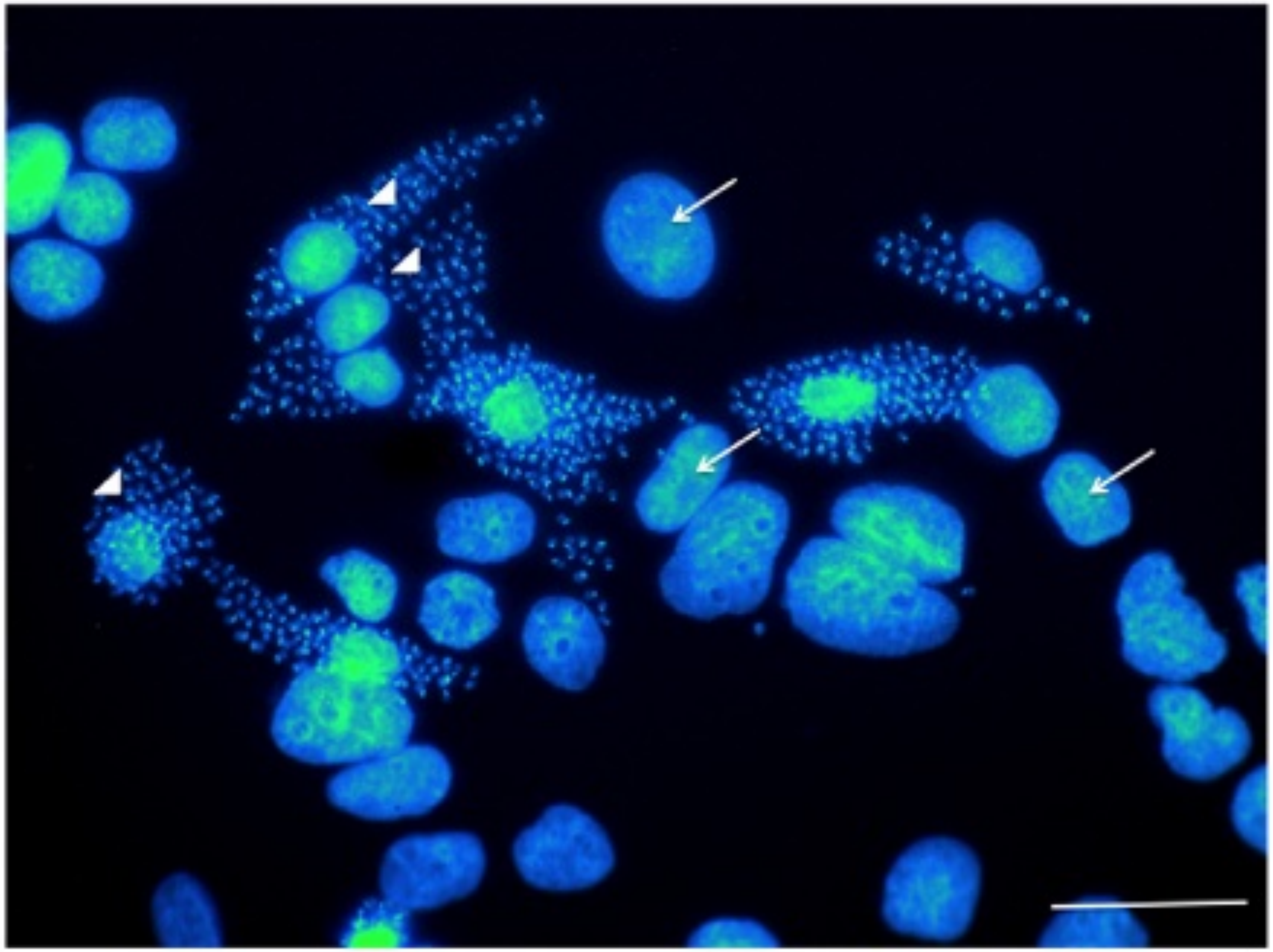
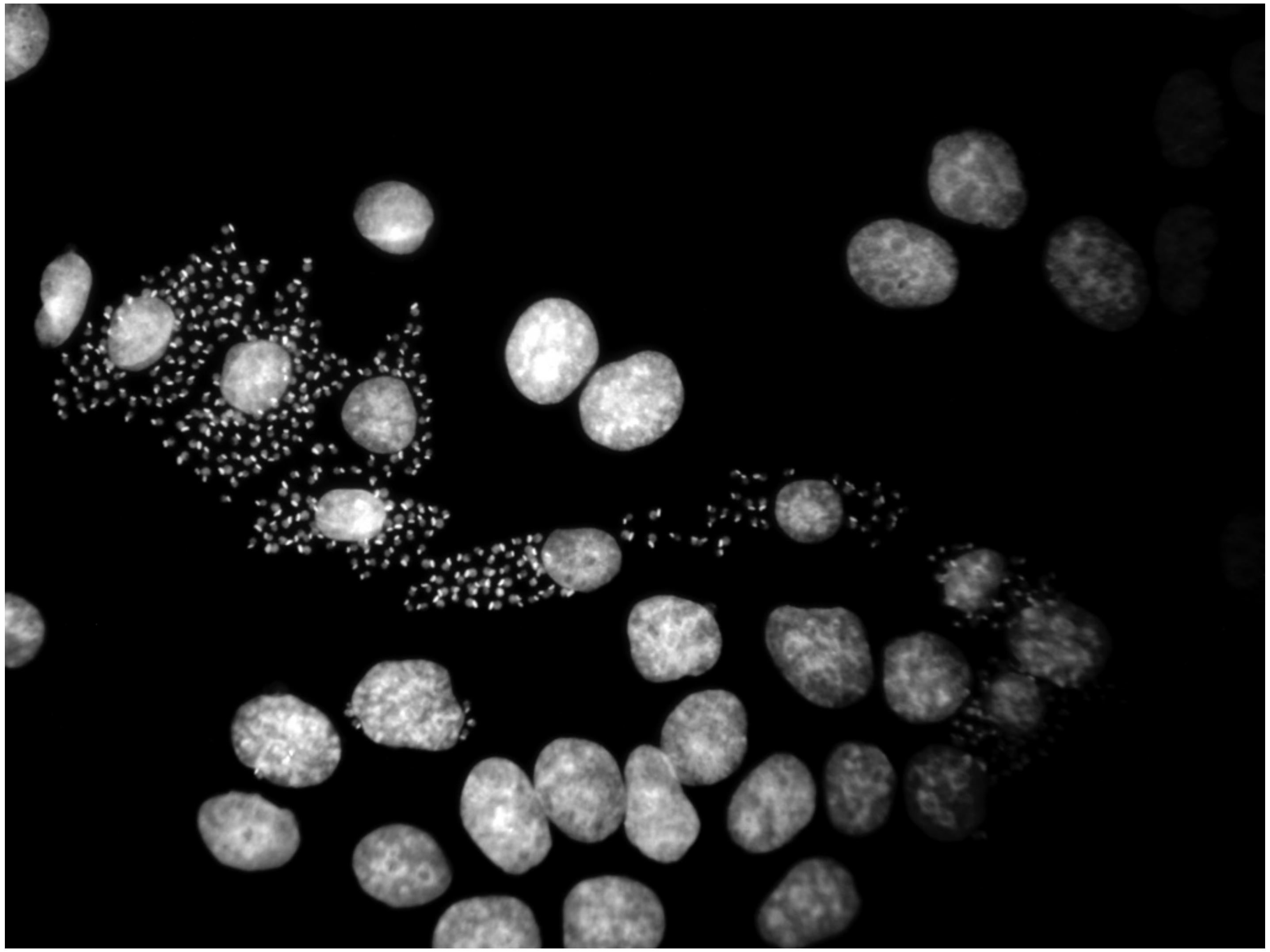


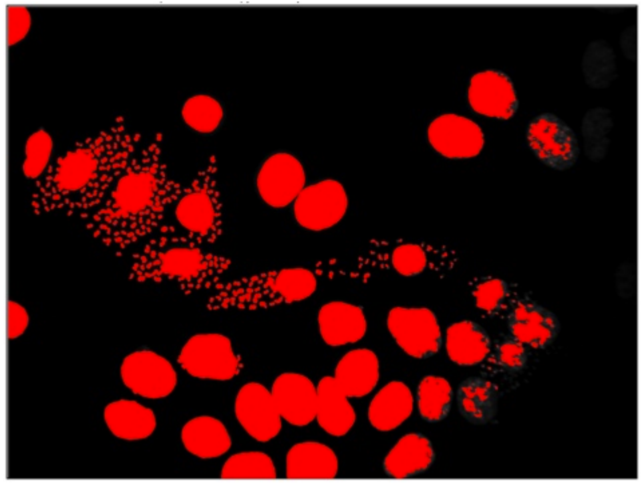
Fig. 1. Infection of BeWo cells with *T. cruzi* amastigotes. BeWo cells were challenged with *T. cruzi* Ypsilon strain trypomastigotes at a parasite:cell ratio of 1:1 for 24 h and were processed for DAPI staining after 48 h. The arrows show BeWo cell nuclei, and the arrowheads show intracellular amastigotes. Scale bar: 10 µm.

- Pregnancy?

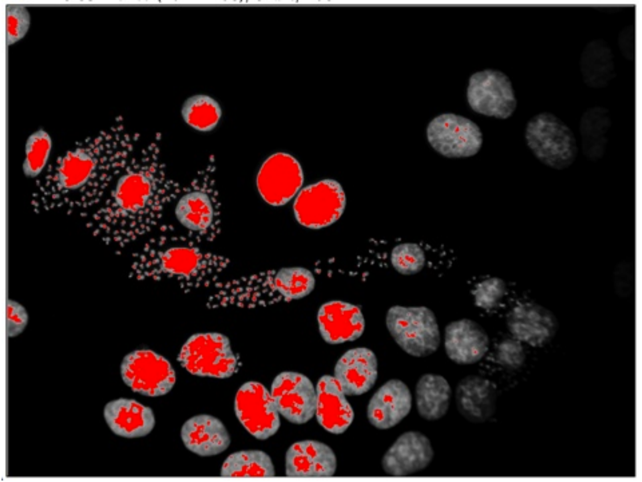
- The simplest segmentation… a manual global threshold



raw image

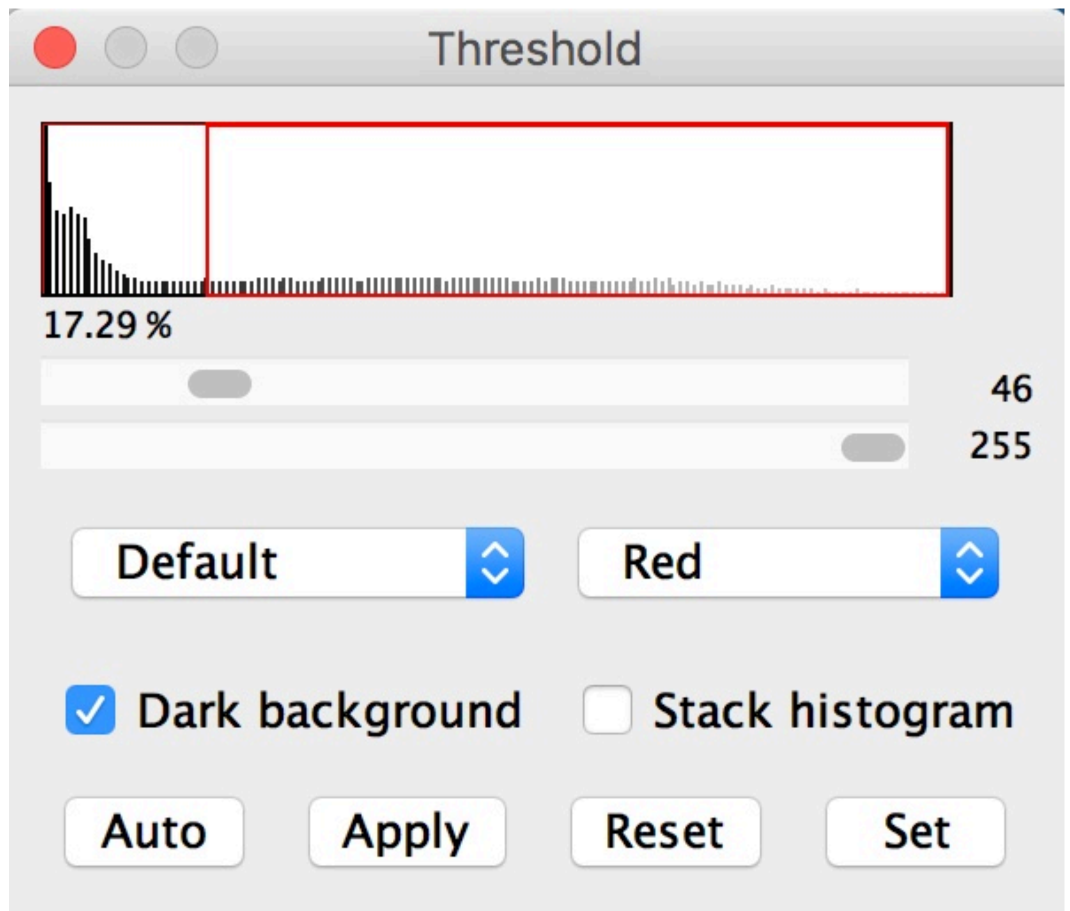


segmentation (>46)

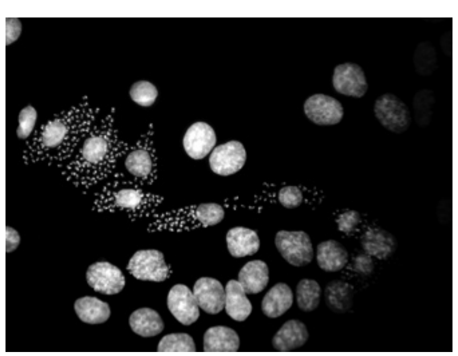


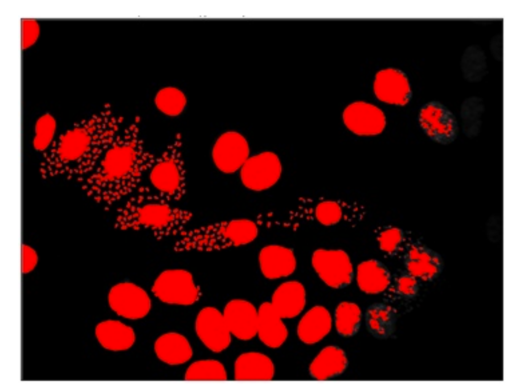
segmentation (>158)

- How to define the threshold ? …



FIJI interface

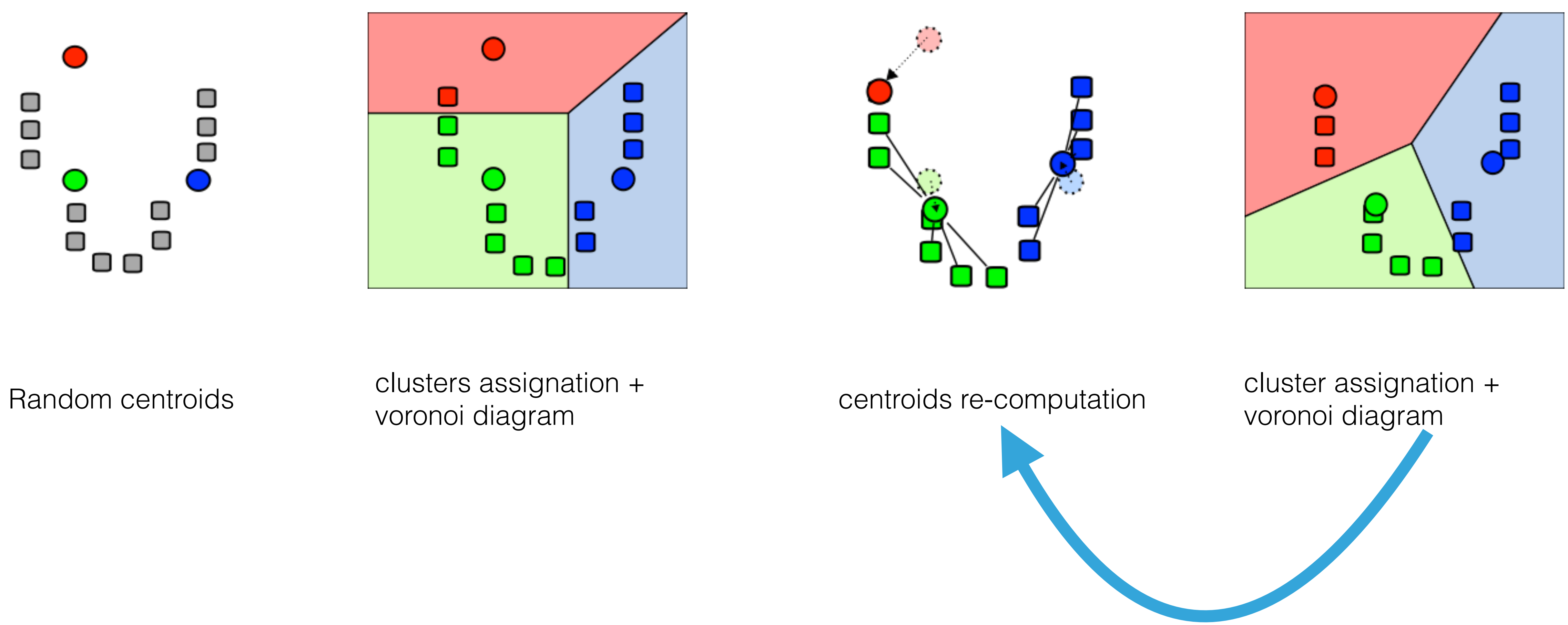




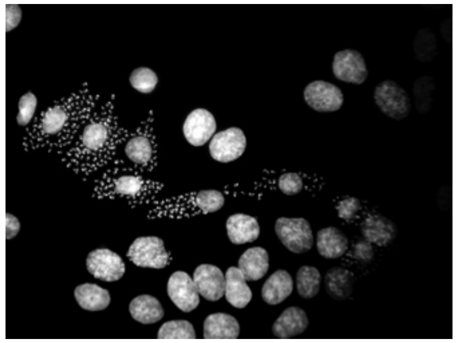
- <u>We have free parameters</u> (!)
- BUT, we know there are two groups of pixels: cells, and background.

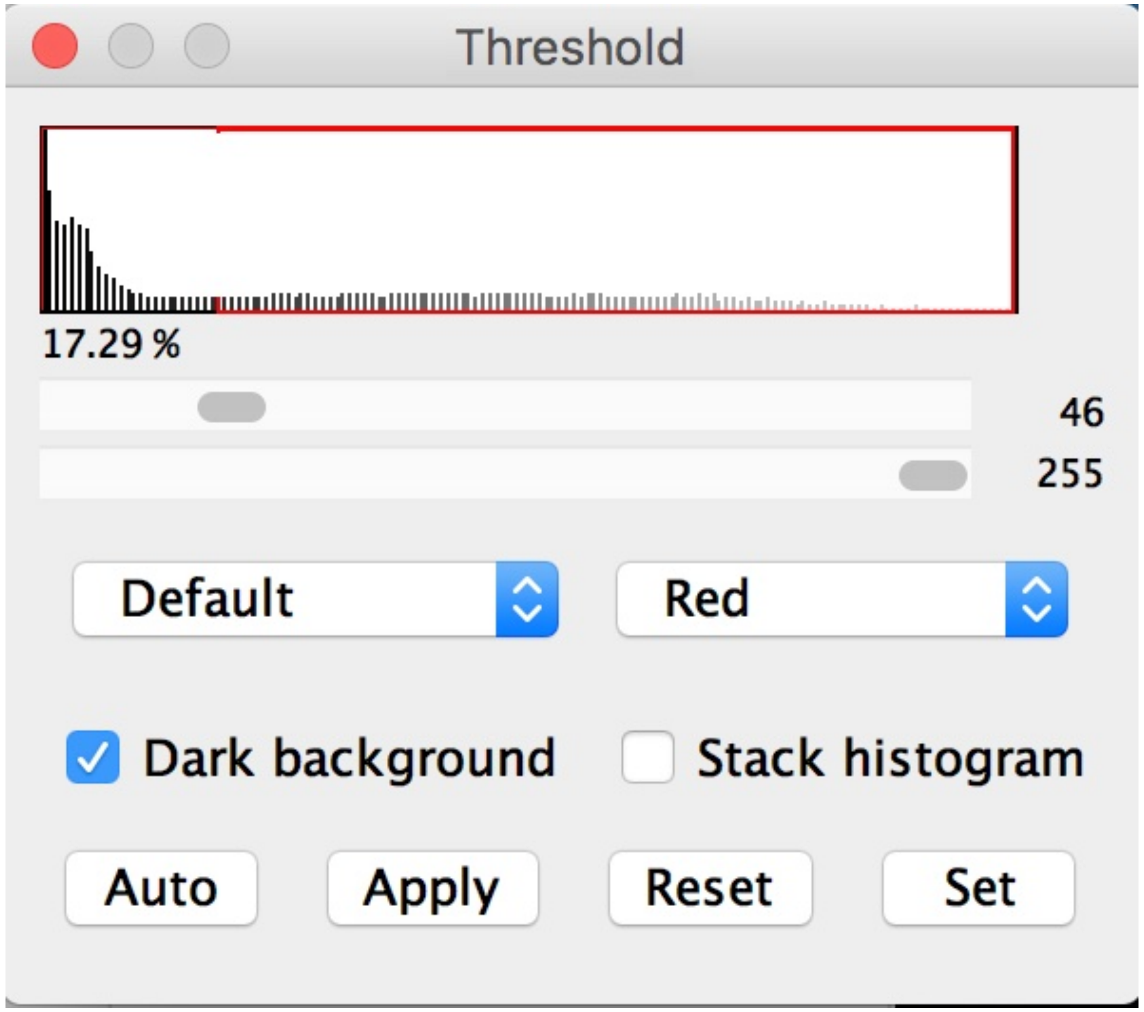
- A kind of (statistical) learning problem!
  - **clustering**
  - **classification**

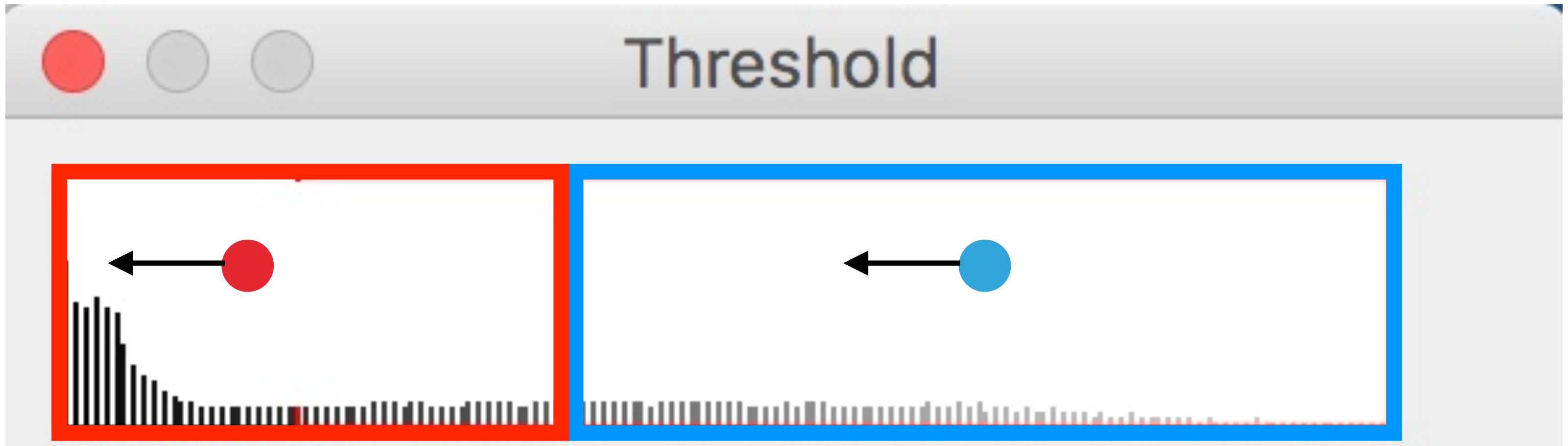
- We can model threshold selection as how to discover the best k groups or clusters at a pixel level.
- K-means clustering (k=3):



- K-means for our image…
- Using the histogram:



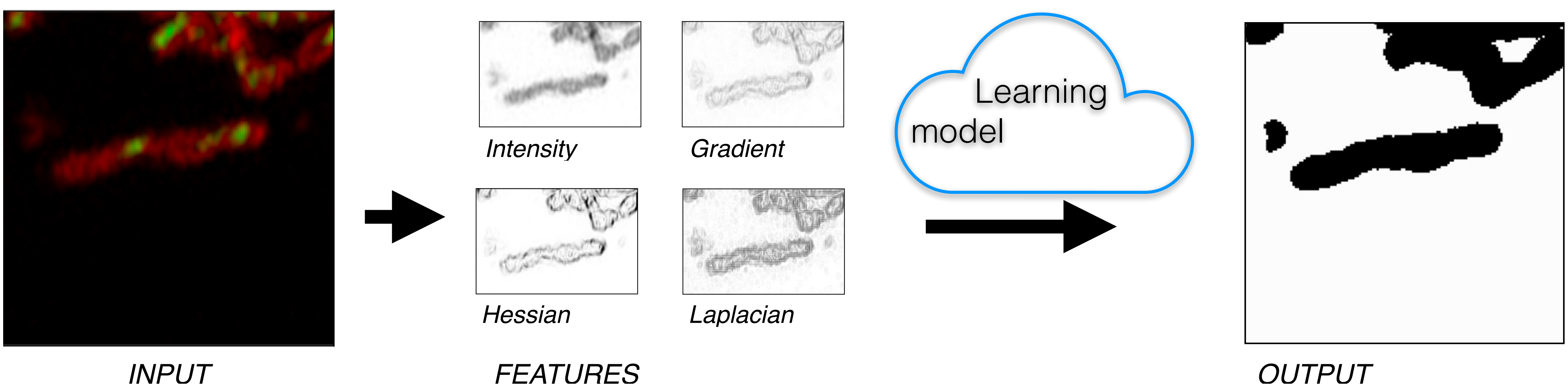


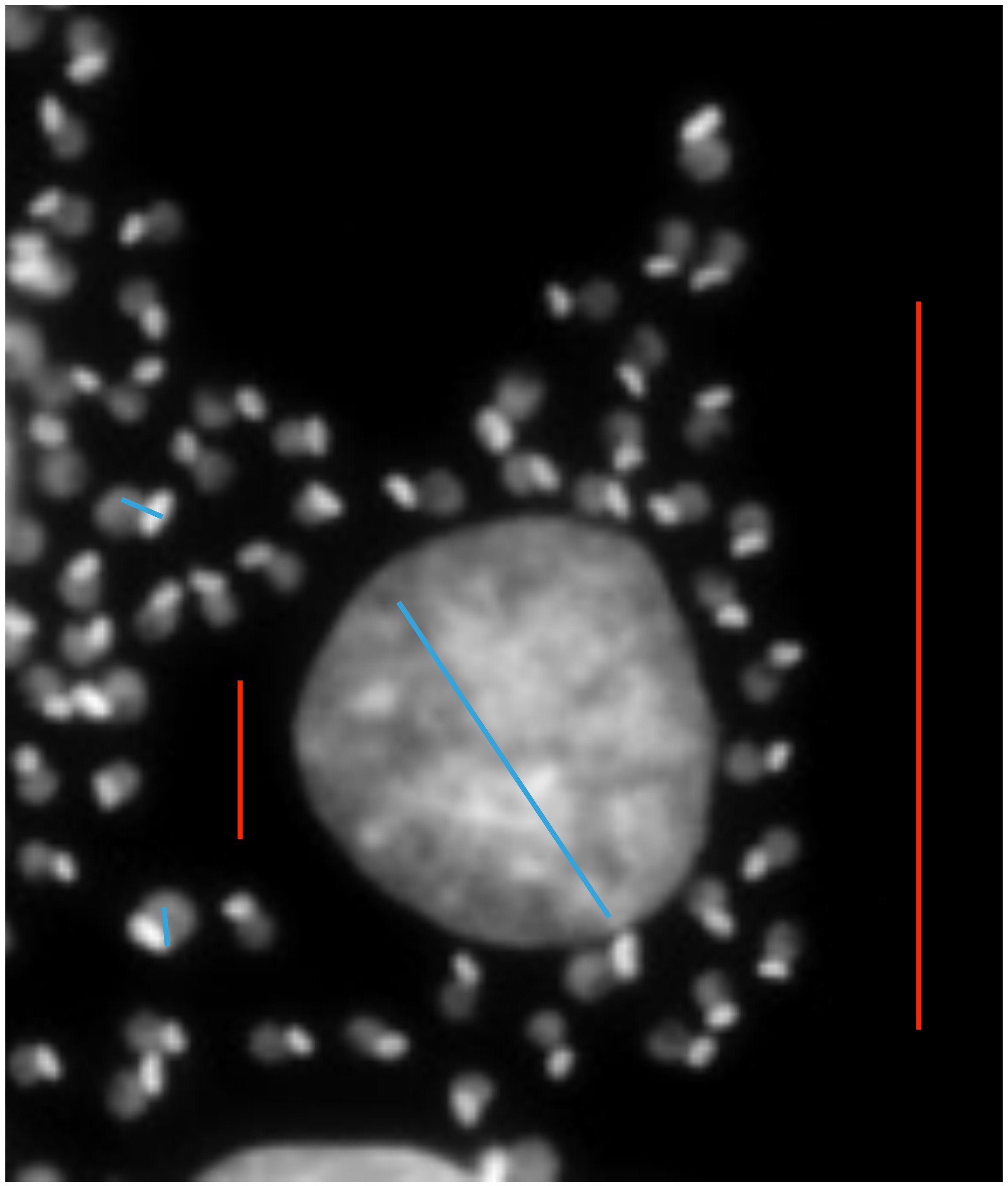


- 1.- Start by guessing 2 centroids.
- 2.- Associate each intensity to 1 centroid.
- 3.- Recompute centroids
- 4.- **Repeat step 2.**

IMAGE SEGMENTATION: SUPERVISED APPROACH

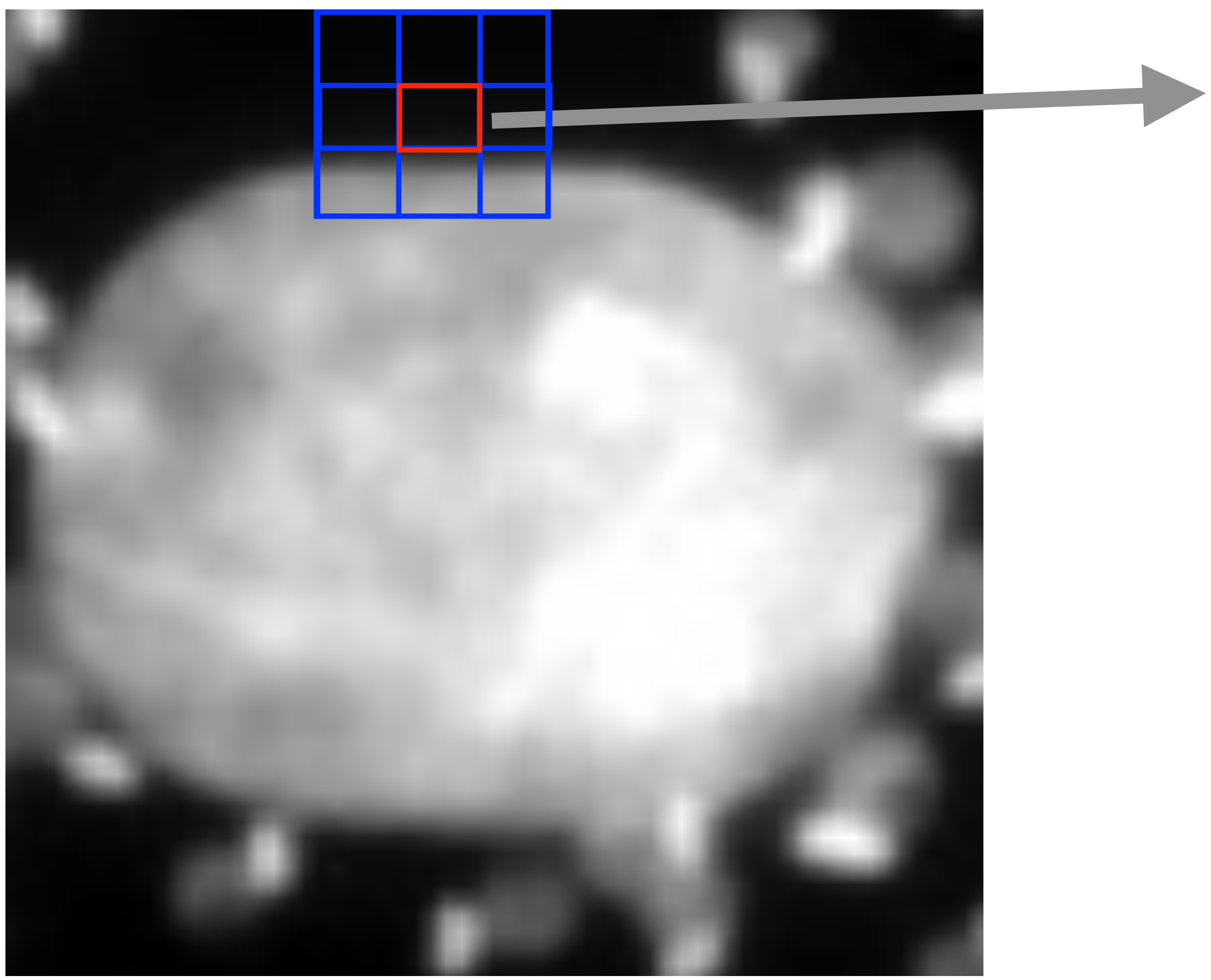
- If we know input and **expected output**: supervised learning.
- Features: voxel intensity, color, shape, size.
- Learning = **how** to identify classes using features.





- We may not have examples of segmentation, BUT we can quickly build examples.
- Class A (background)
- Class B (objects)

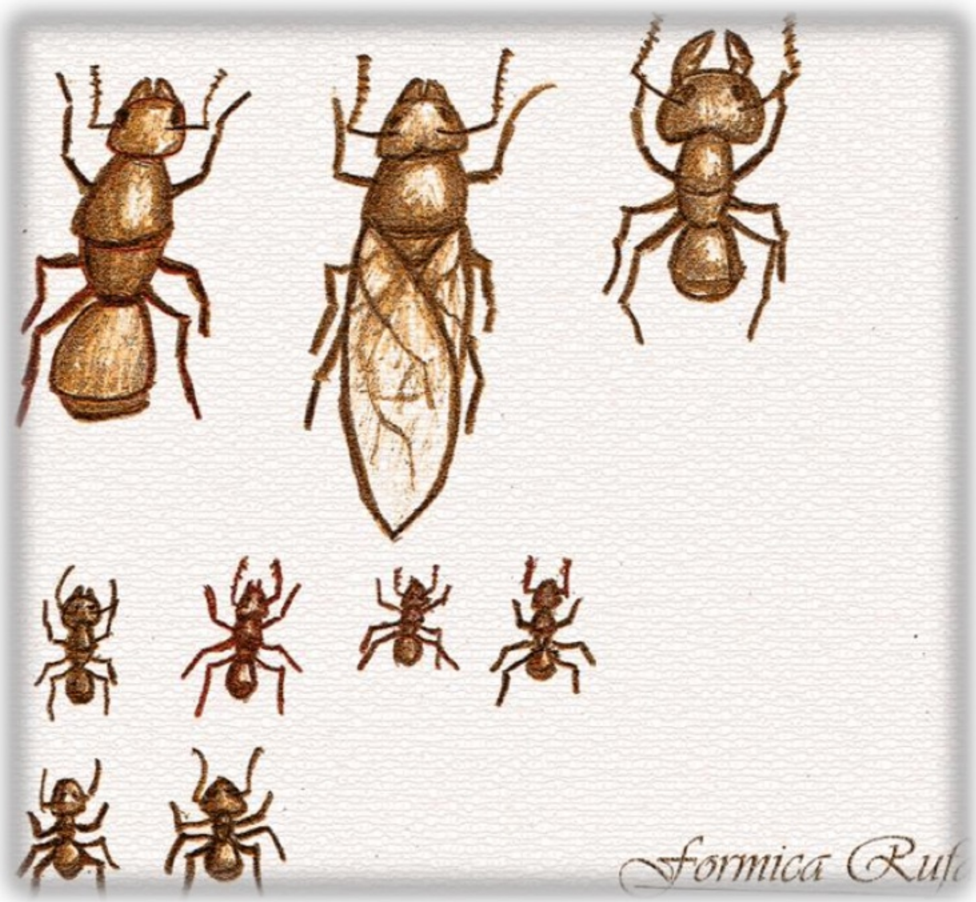
- We can understand pixels in higher dimensions.



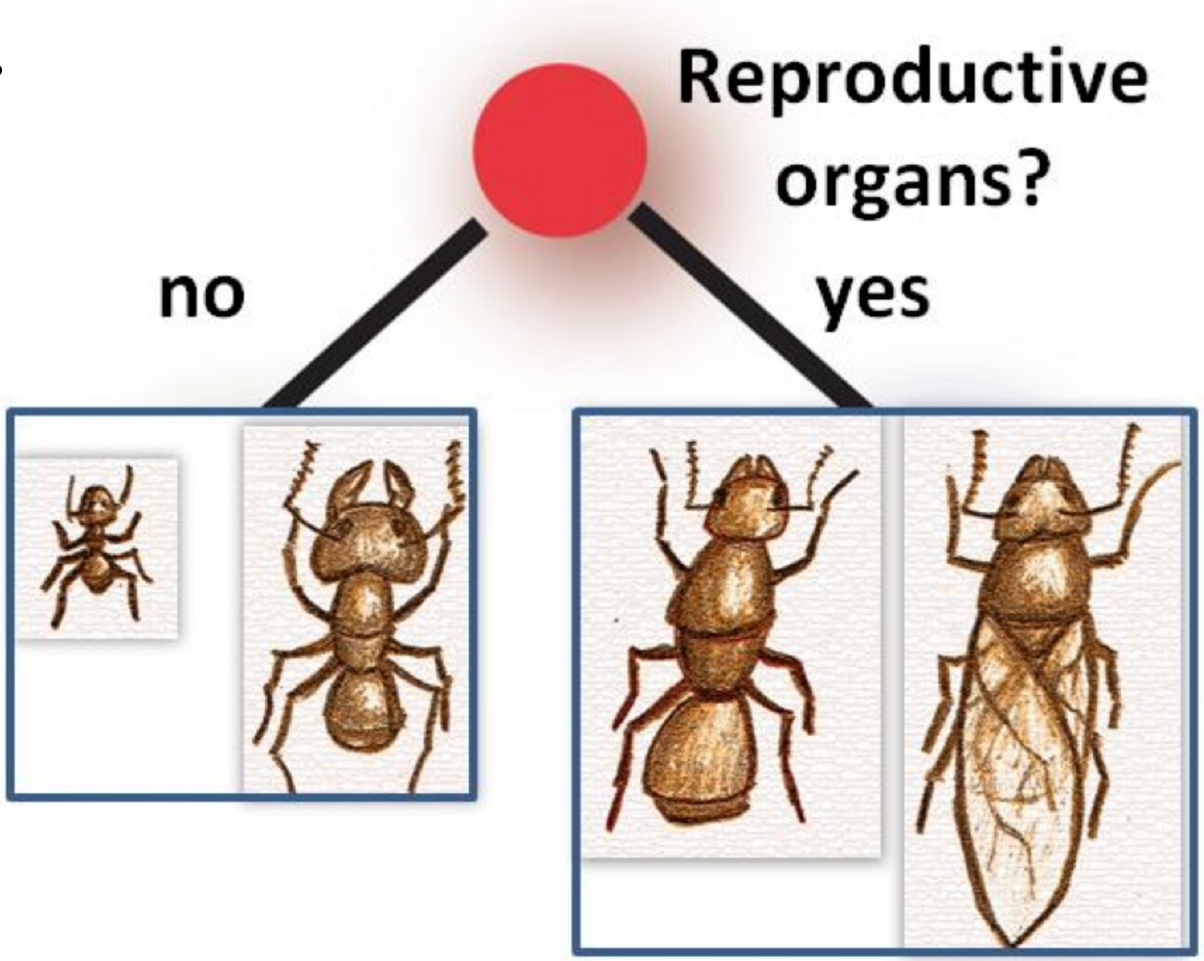
- Intensity (0)
- Variance 3x3
- Mean 3x3
- Sobel 3x3
- ...

- How to build rules to **identify cells and parasites**?

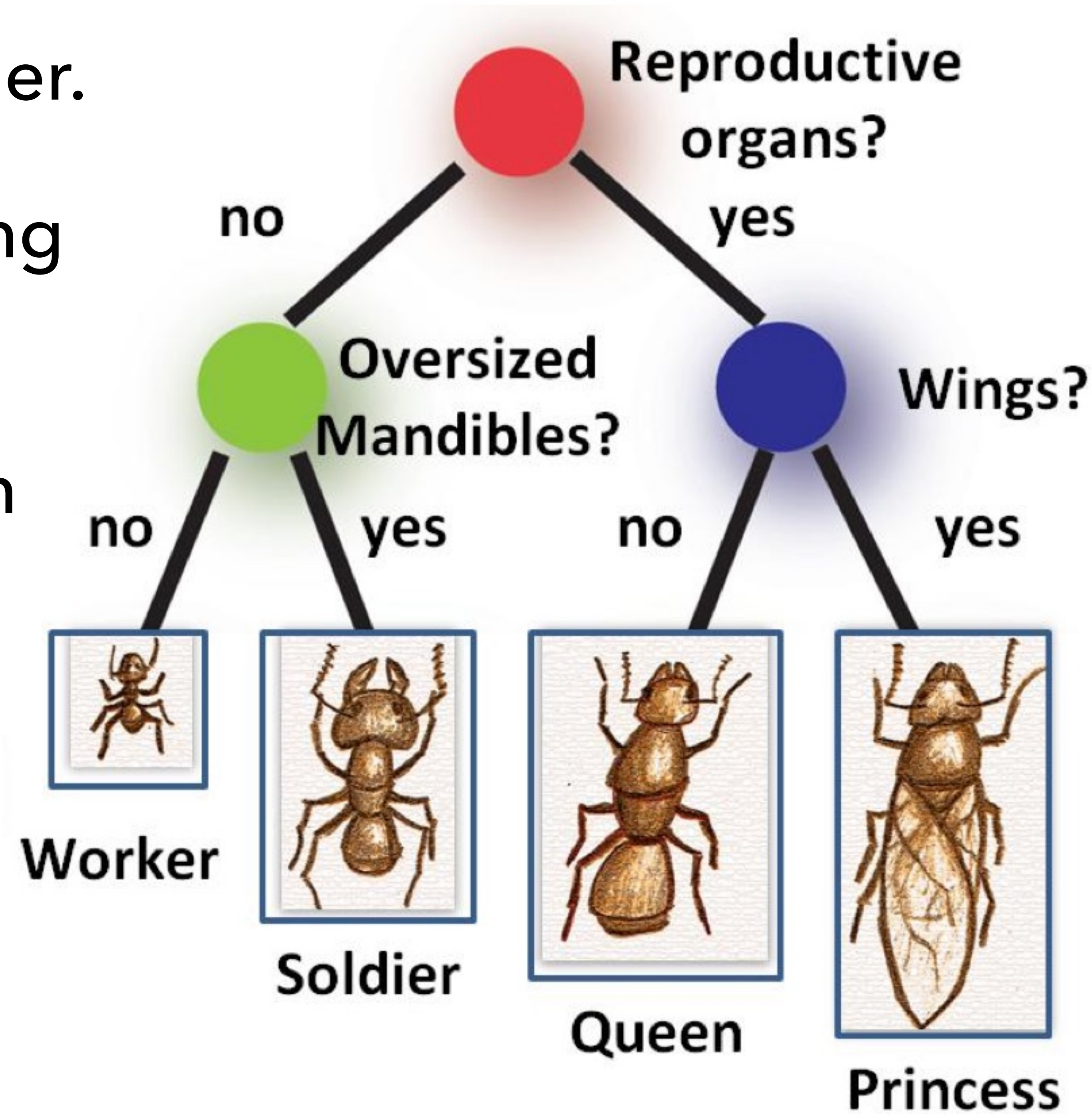
- To classify pixels we can use many methods (SVM, Neural networks), even in high dimensions. For instance decision trees.
- We can think pixels information as “ants” that we want to classify.



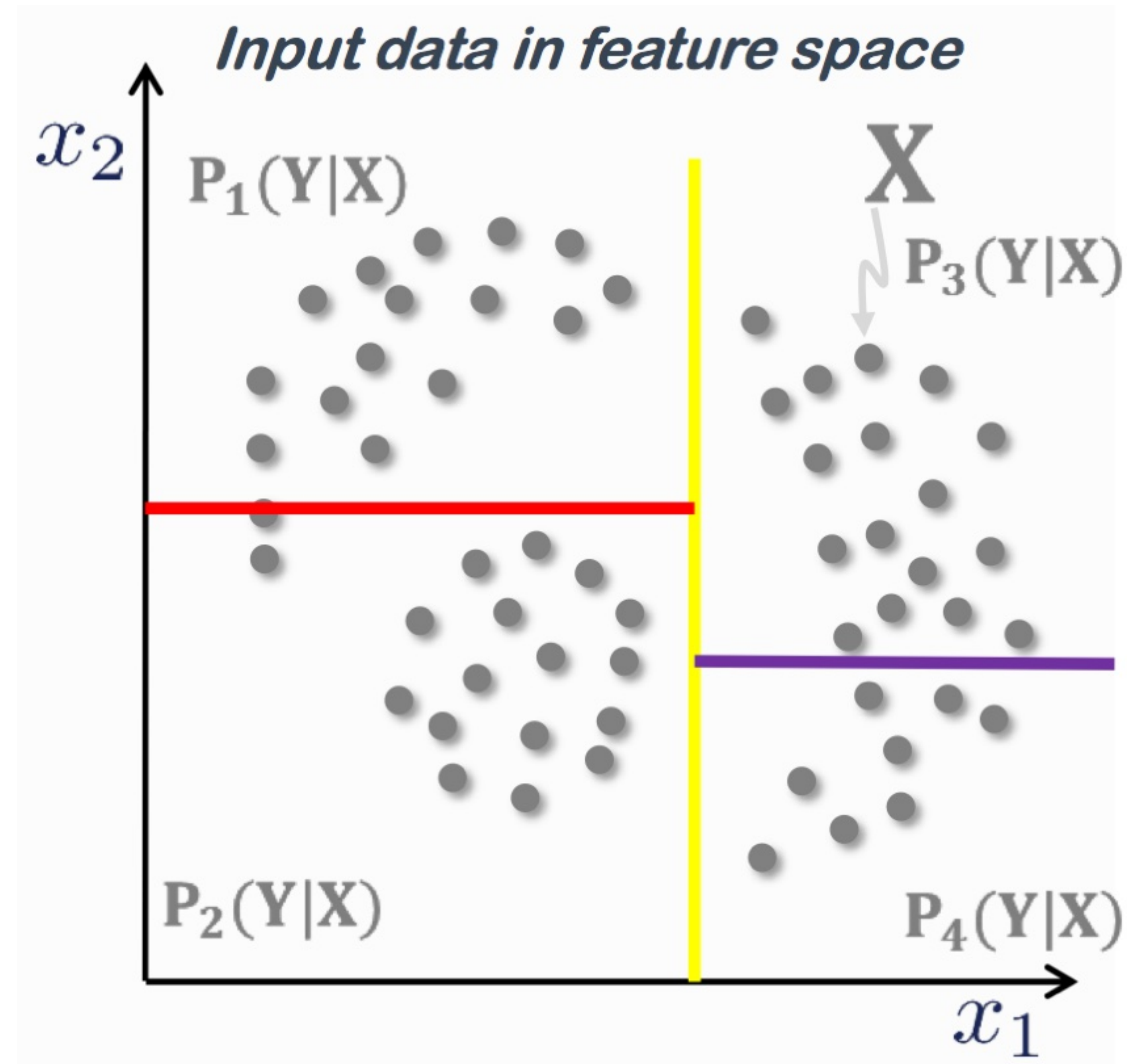
- Main idea: divide & conquer.
  - (1) Divide examples using simple rules.
  - (2) Conquer: repeat with subgroups



- Main idea: divide & conquer.
  - (1) Divide examples using simple rules.
  - (2) Conquer: repeat with subgroups



- **Objetive**: From observations (X), identify the probability of class (Y), or P(Y|X).
  - (1) Partition: each rule divide feature space.
  - (2) Model: Compute P(Y|X) per partition.
- The tree estimates P(Y|X) by parts (partitions).



- We still need to know:
  - How to build the tree?
  - How to measure how good the tree is?

IMAGE SEGMENTATION: HOW TO BUILD TREES?

Node training

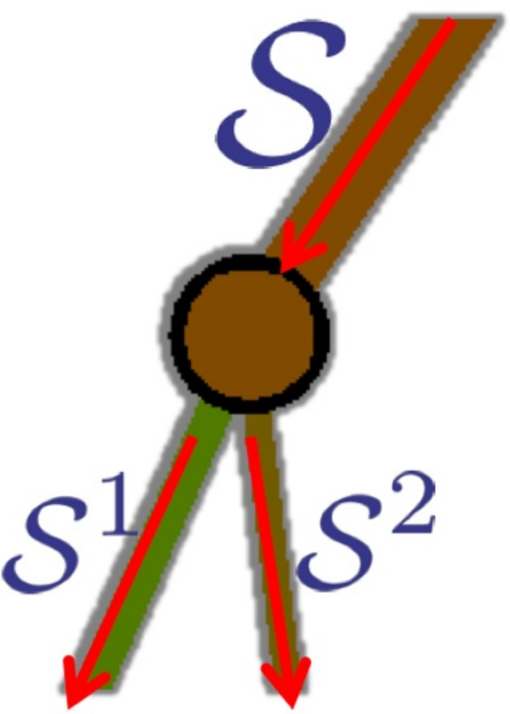
$$\boldsymbol{\theta}^* = \arg\max_{\boldsymbol{\theta}\in\mathcal{T}} I$$

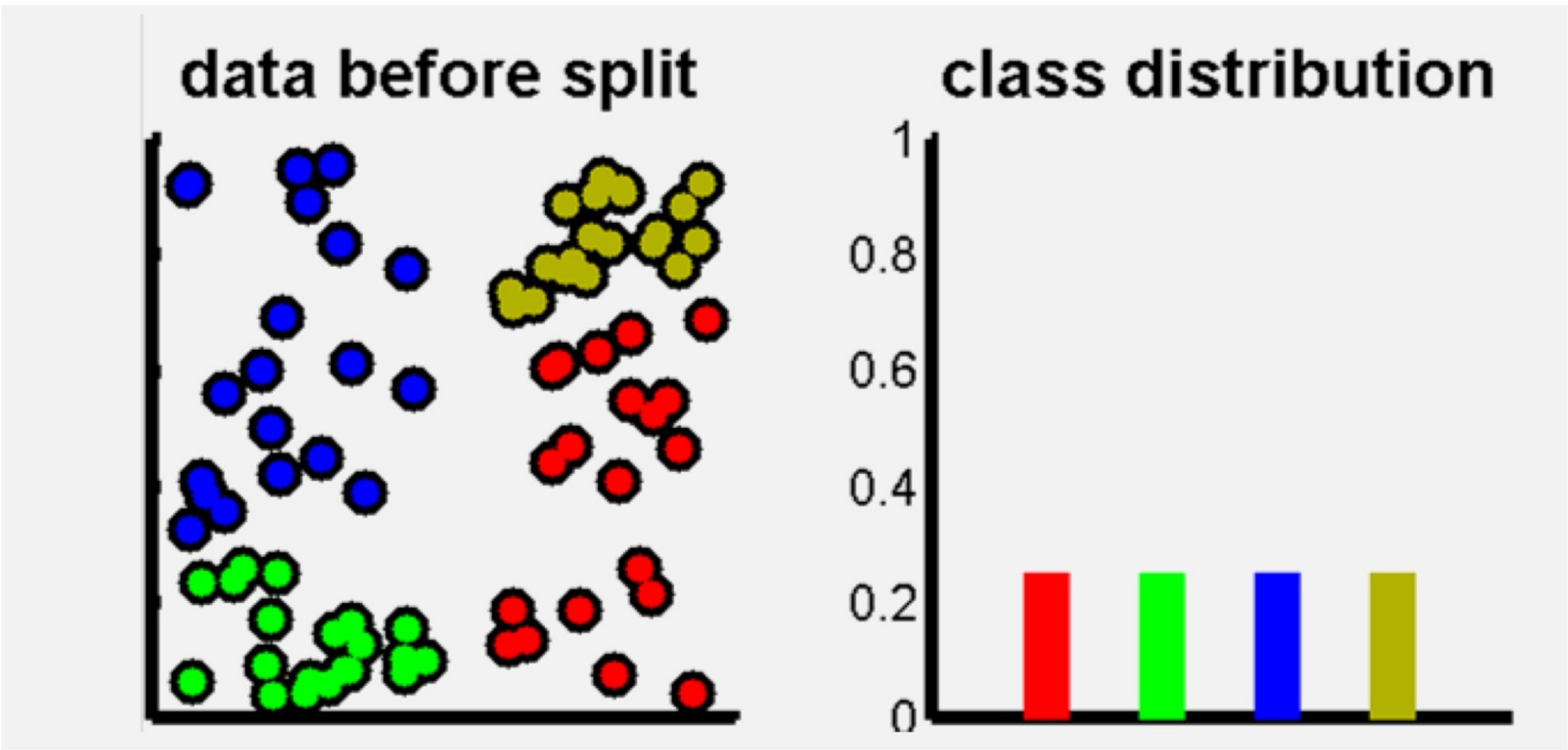
Information gain

$$I = H(\mathcal{S}) - \sum_{i\in\{1,2\}} \frac{|\mathcal{S}^i|}{|\mathcal{S}|} H(\mathcal{S}^i)$$

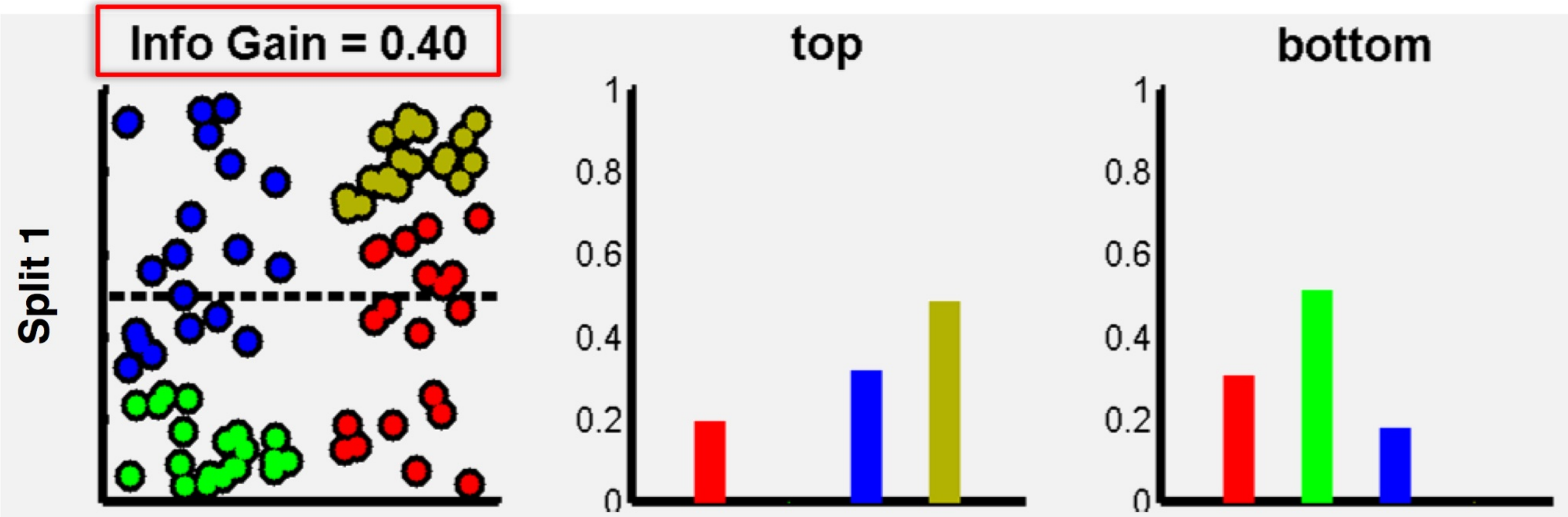
Shannon's entropy

$$H(\mathcal{S}) = -\sum_{c\in\mathcal{C}} p(c)\log(p(c))$$

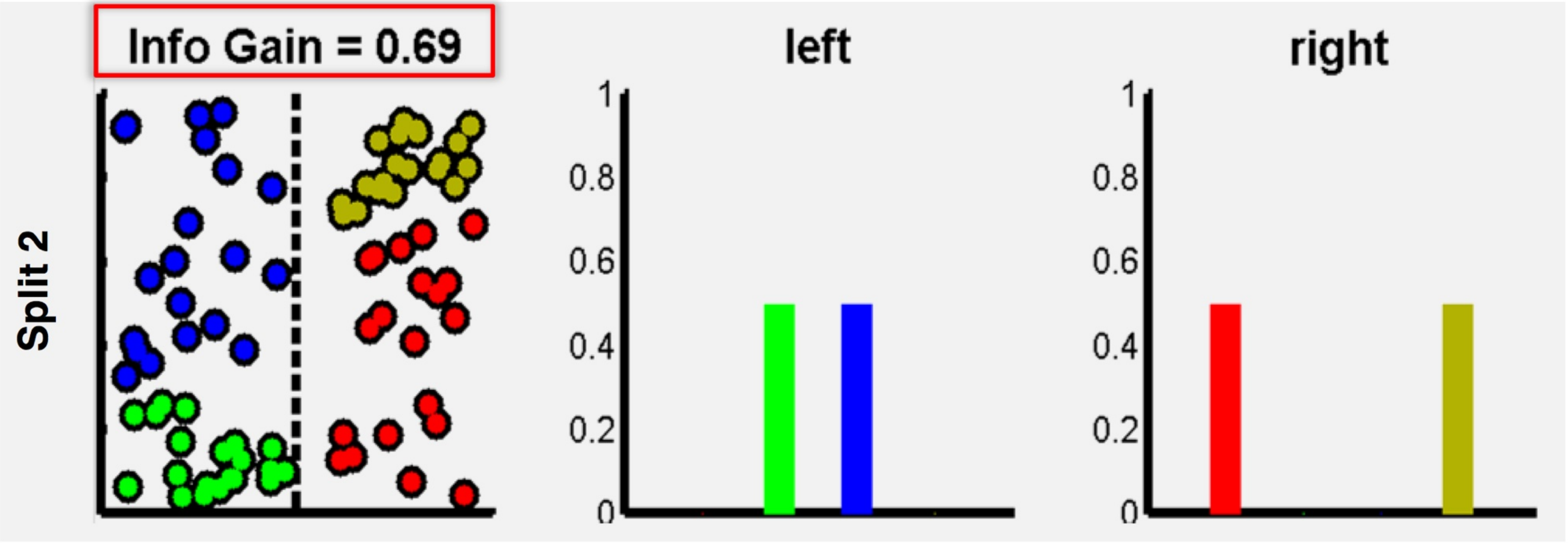




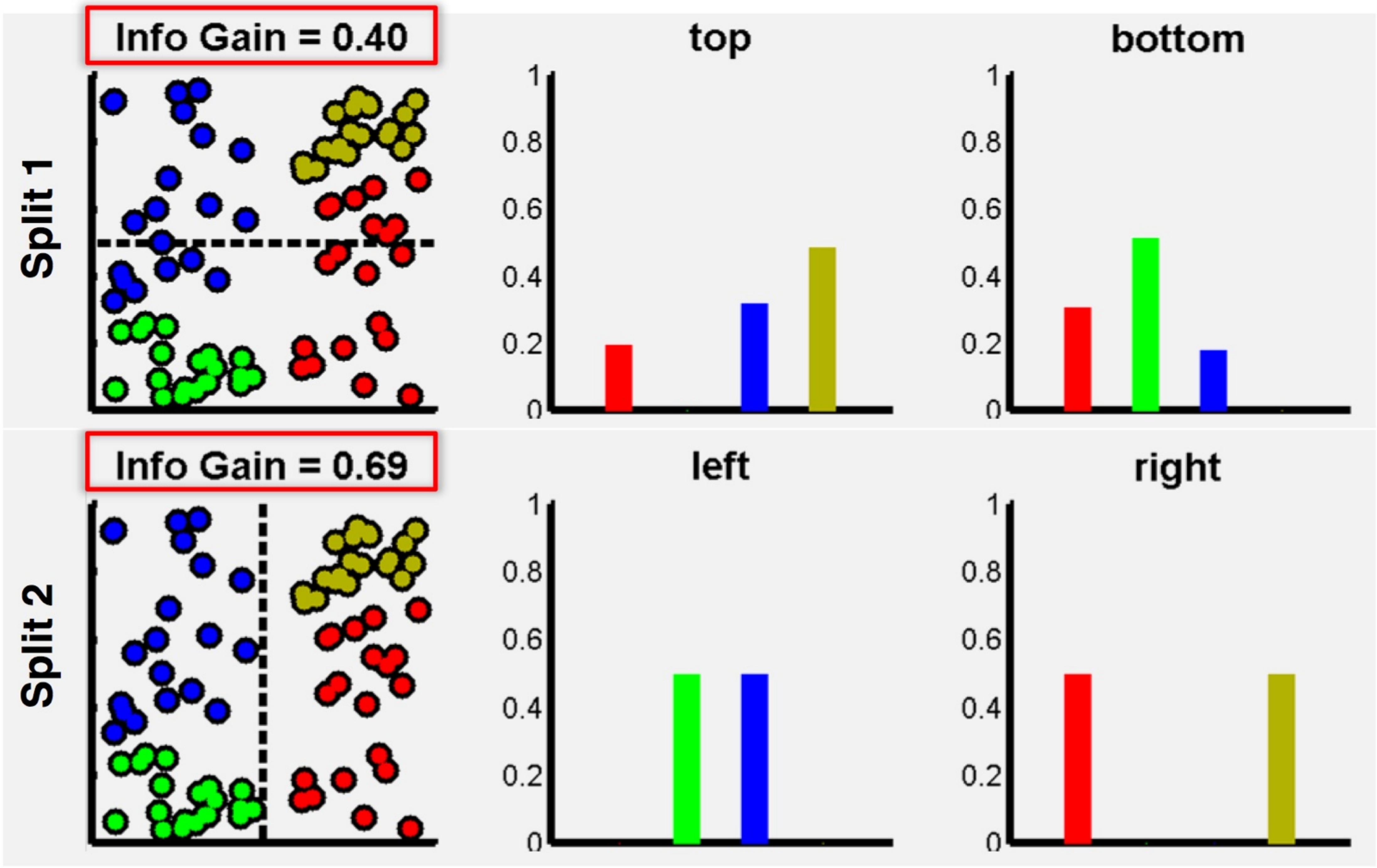
- We will compare two rules: **horizontal** or vertical partition.



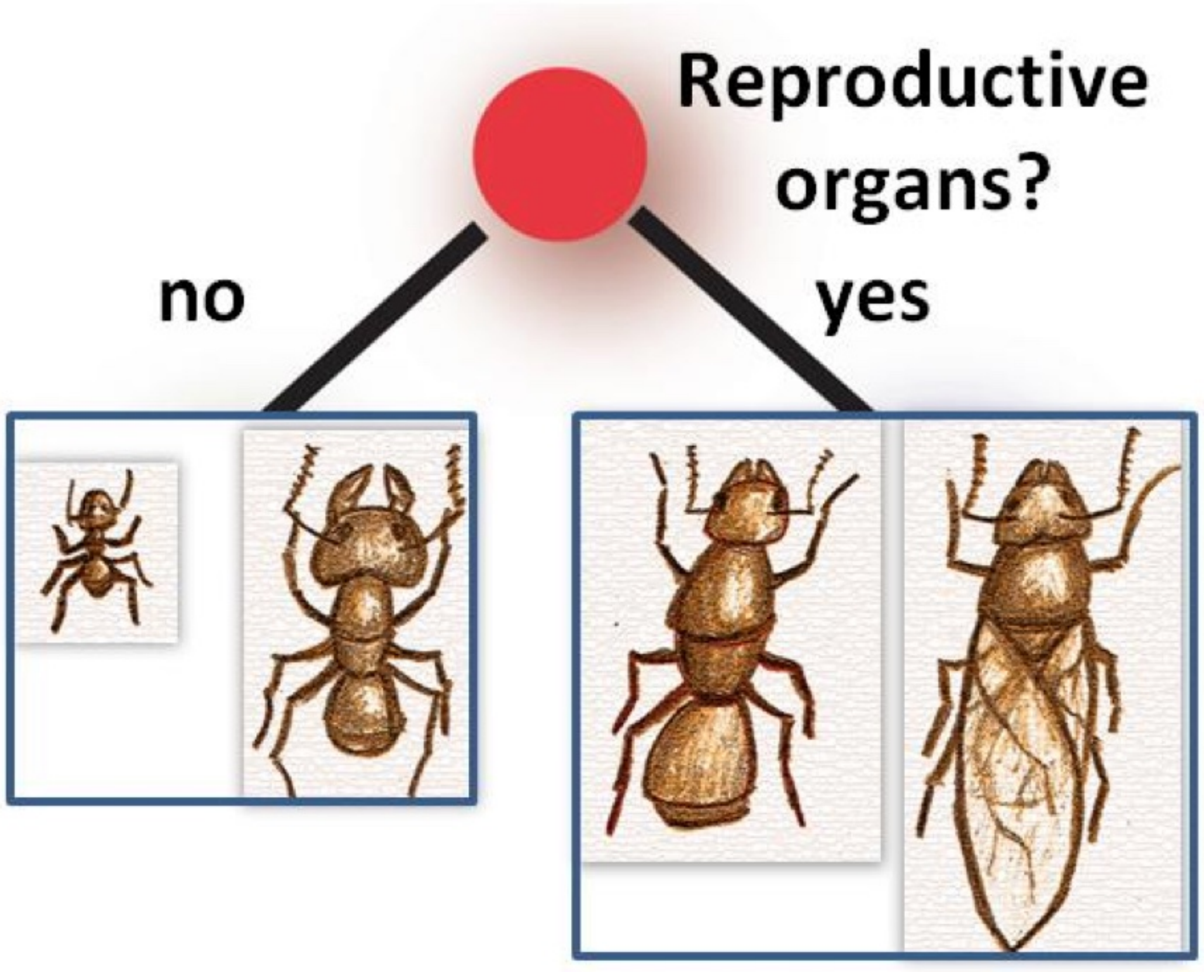
- We will compare two rules: horizontal or **vertical** partition.



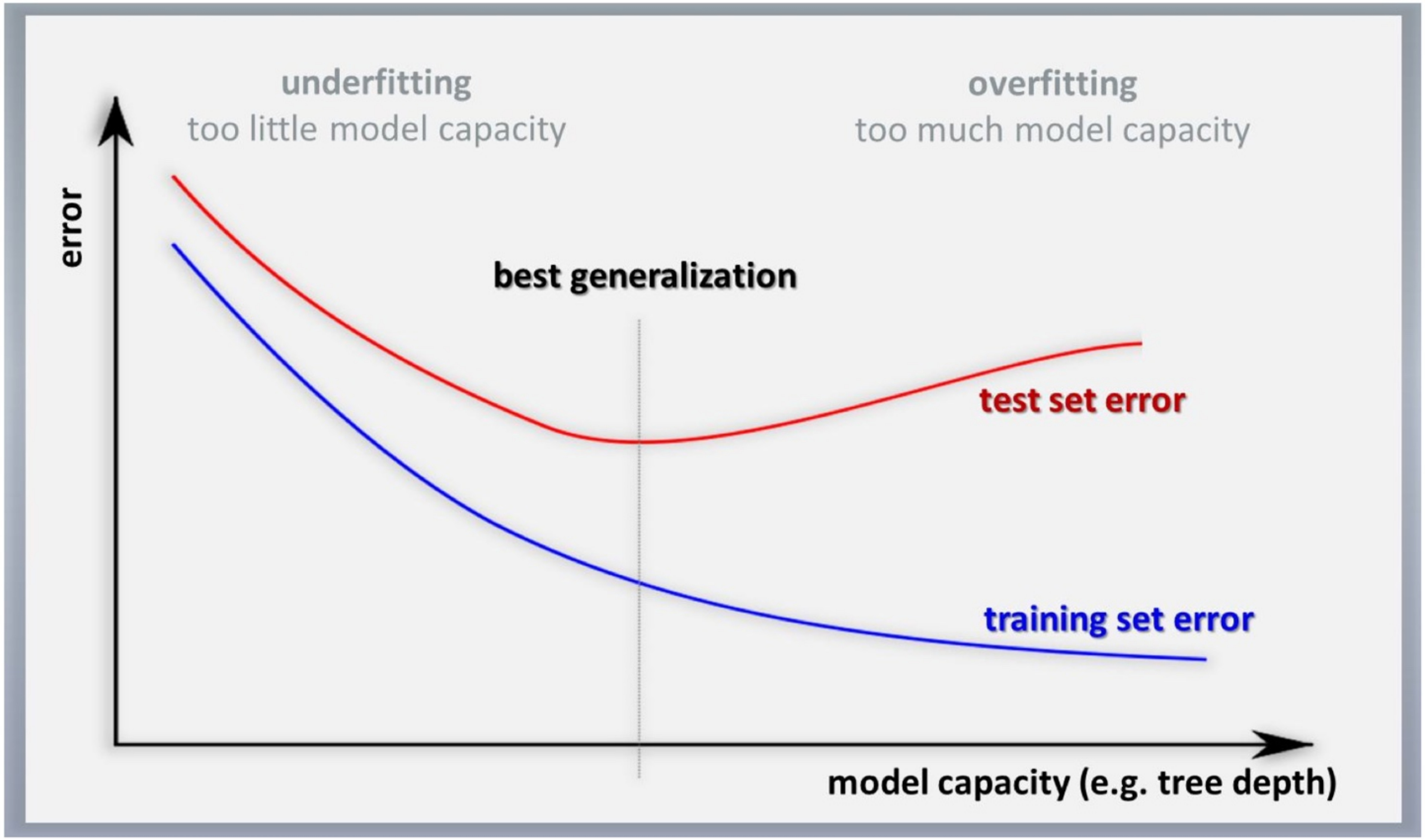
- Horizontal or vertical?



- But maybe, “Reproductive organs” was a too good or too bad question to start with or we did too many questions (overfitting)

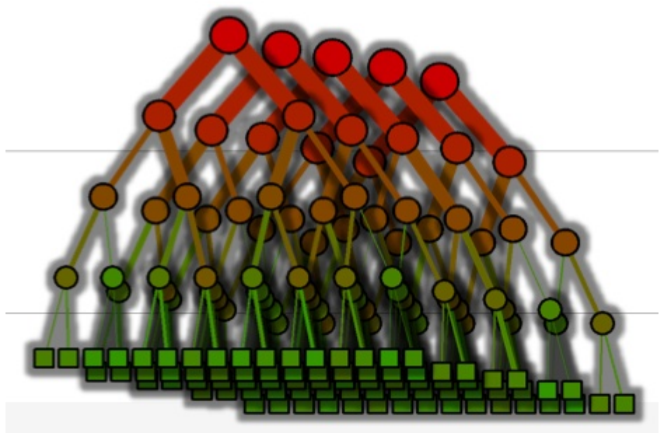


- A single **decision tree** is sensitive to overfitting.



- **Idea**: to replace the tree by a forest.
- In a forest each tree is slightly different.
- The uncorrelated tree set improves generalization properties.

- Main parameters:
  - Trees depth
  - Number of trees
  - Select input features



- **DEMO: Use weka to train for k-mean threshold a random forest to segment nuclei + parasites [FIJI plugin]**