Bioimage Informatics

Daniel Sage & Arne Seitz

Course

Machine Learning

École
polytechnique
fédérale
de Lausanne

BIO-410 □ BIOIMAGE INFORMATICS

Artificial Intelligence

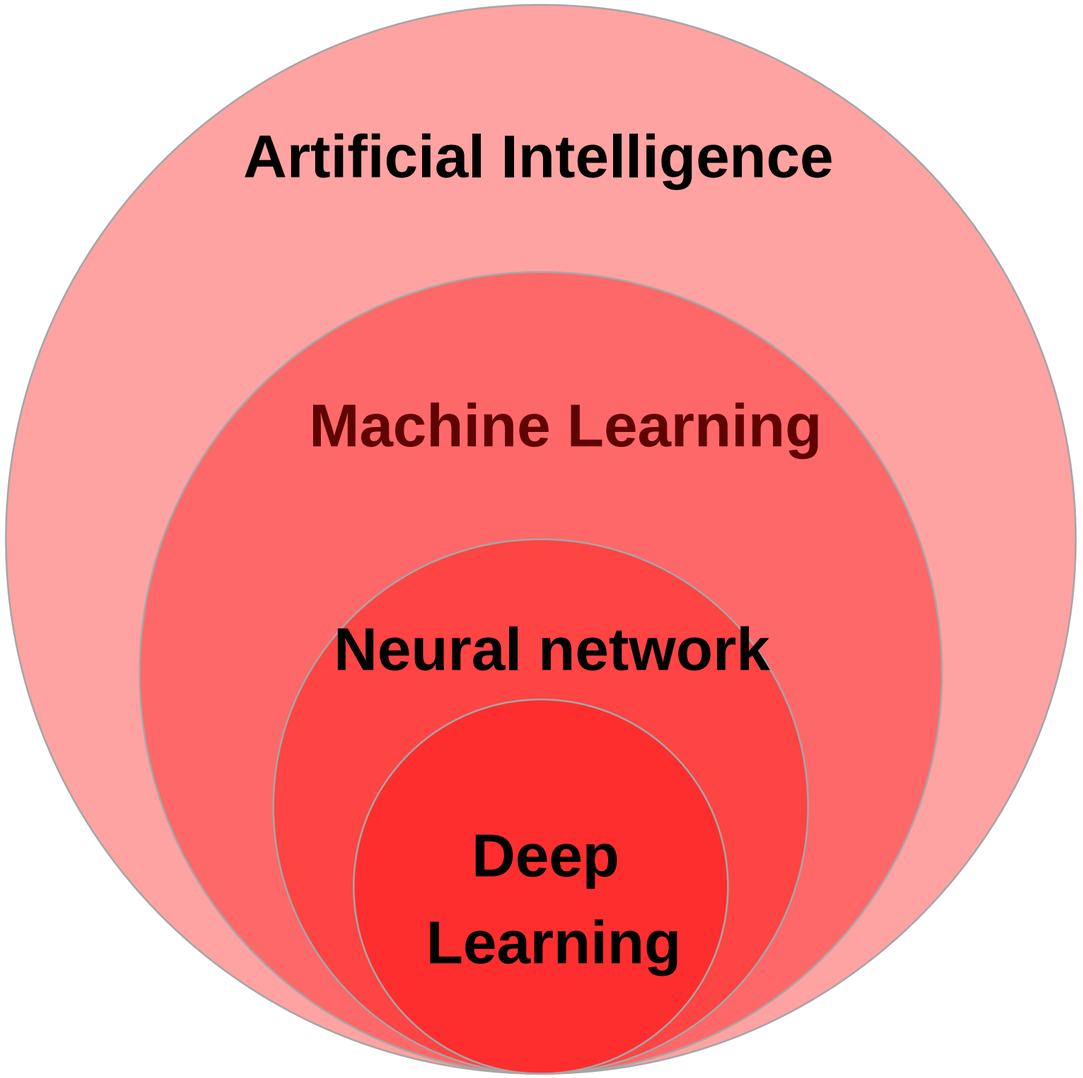
Artificial Intelligence

Machine Learning

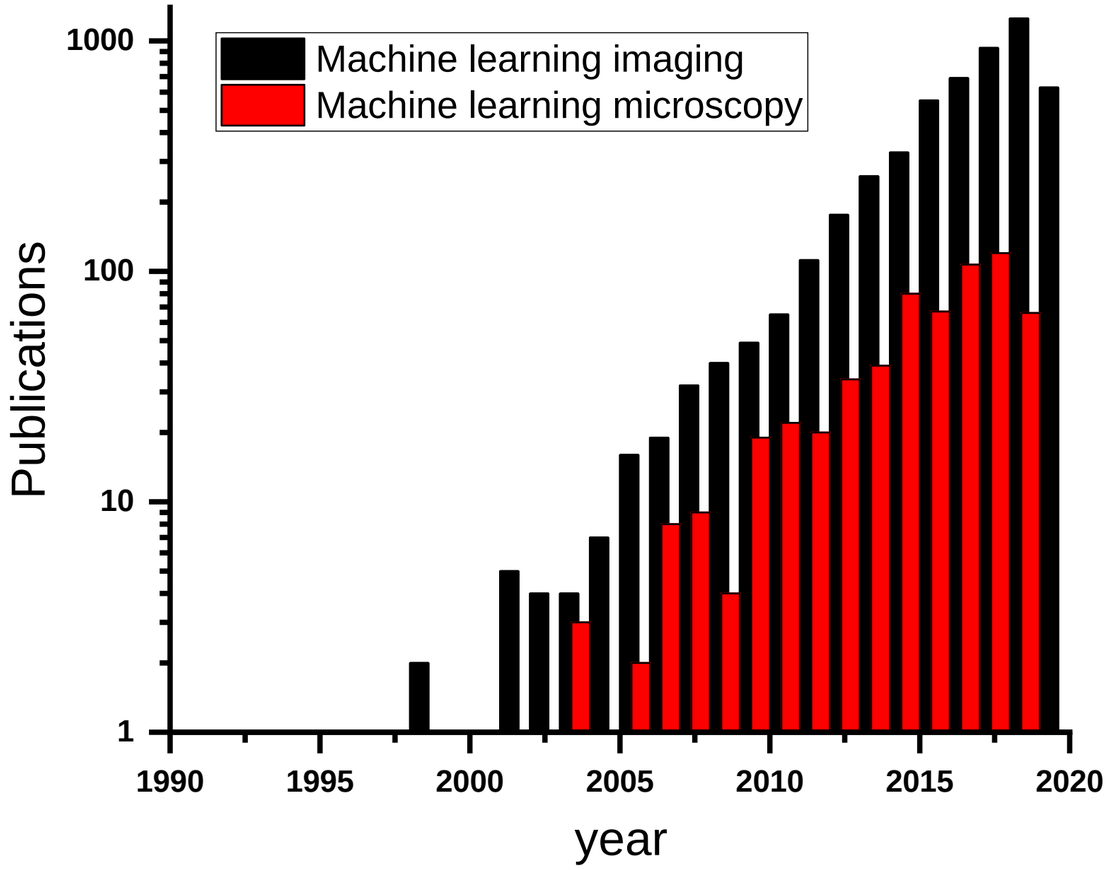
Neural network

Deep Learning

Artificial Intelligence



Machine learning in biosciences



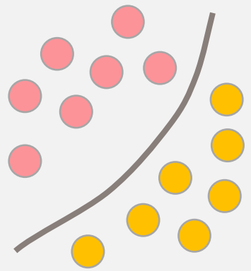
PubMed search: 16.05.2019

Classical Machine Learning

Classical Machine Learning

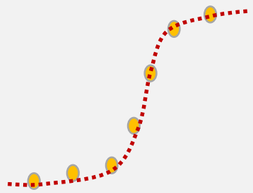
Supervised

Classification



predict a category

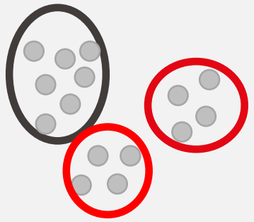
Regression



predict a number

Unsupervised

Clustering

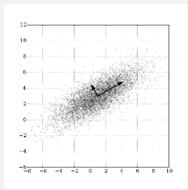


Divide by similarity

Association

Identify sequences

Dimension Reduction



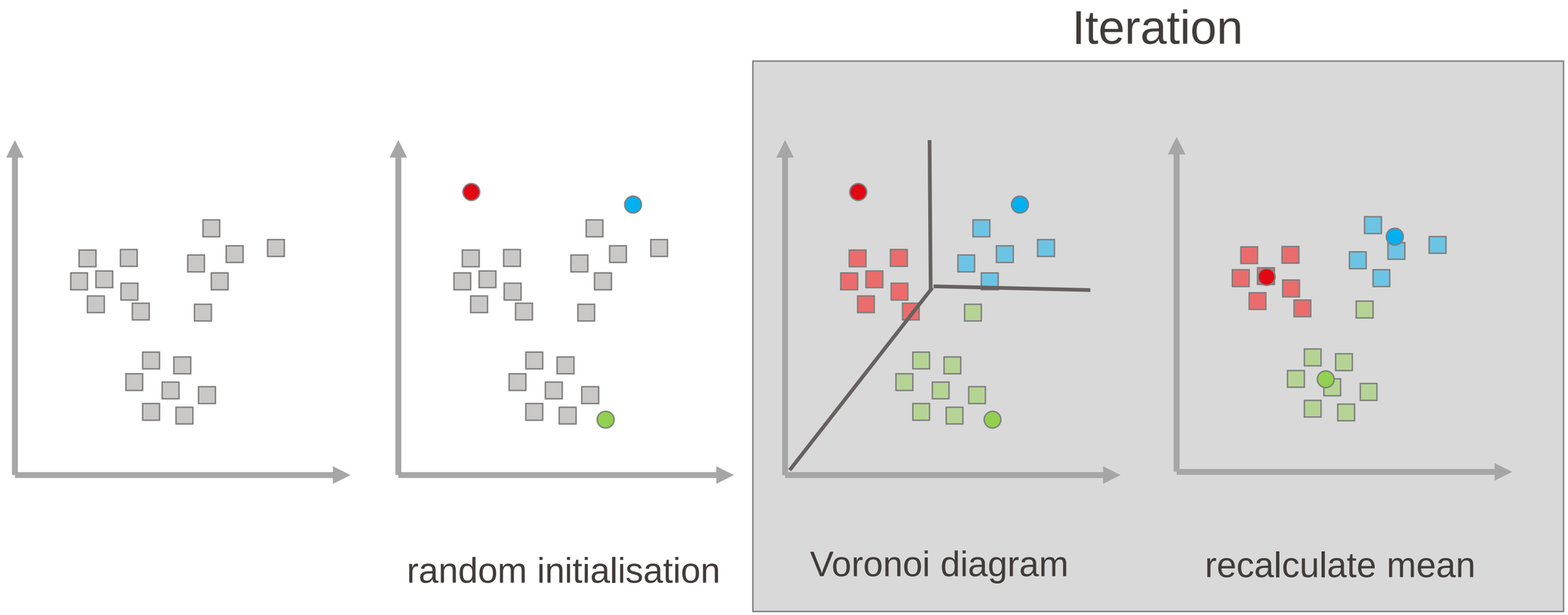
Find hidden dependencies

Clustering

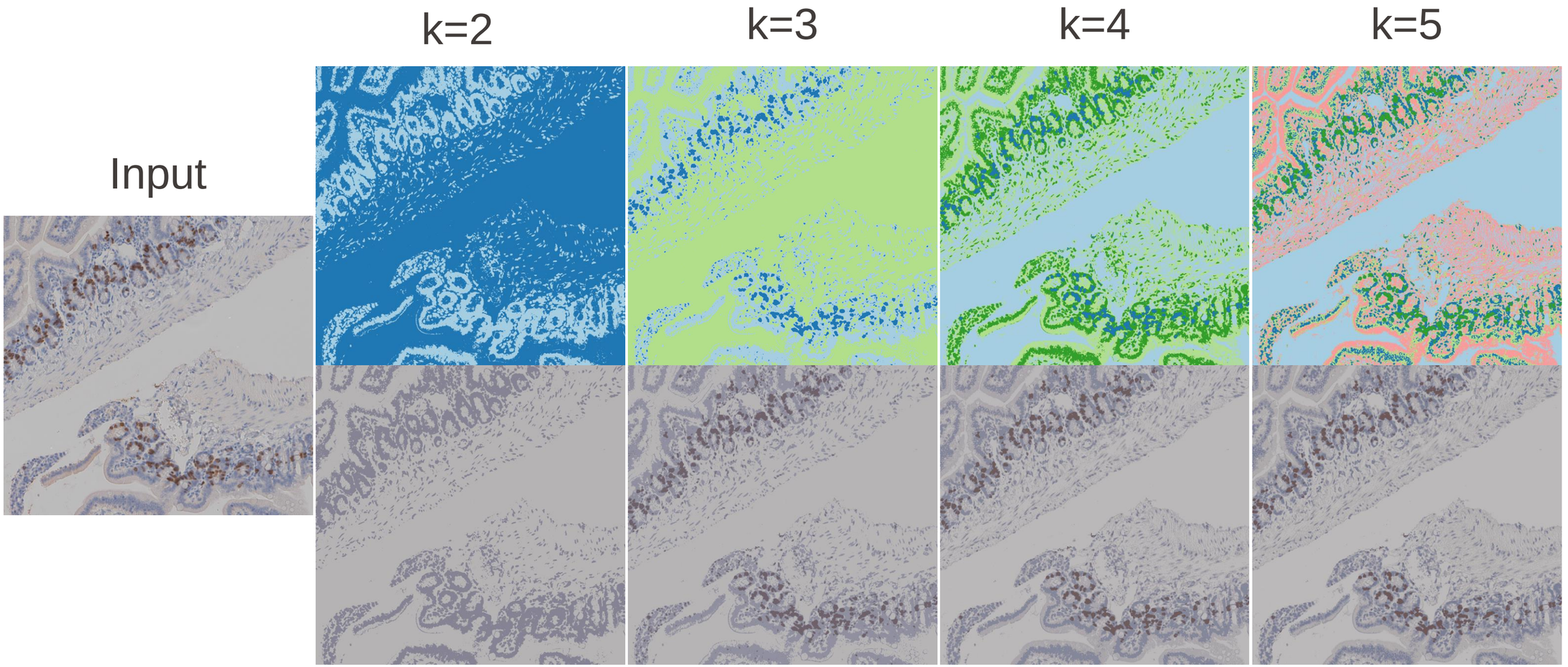
k-means clustering

- Partition **n** observations into **k** clusters
- Each observation belongs to the cluster with the nearest mean (cluster center/centroid)
- Minimizes within custer variances (squared Euclidian distances)
- Computational difficult (NP-hard)

k-mean clustering

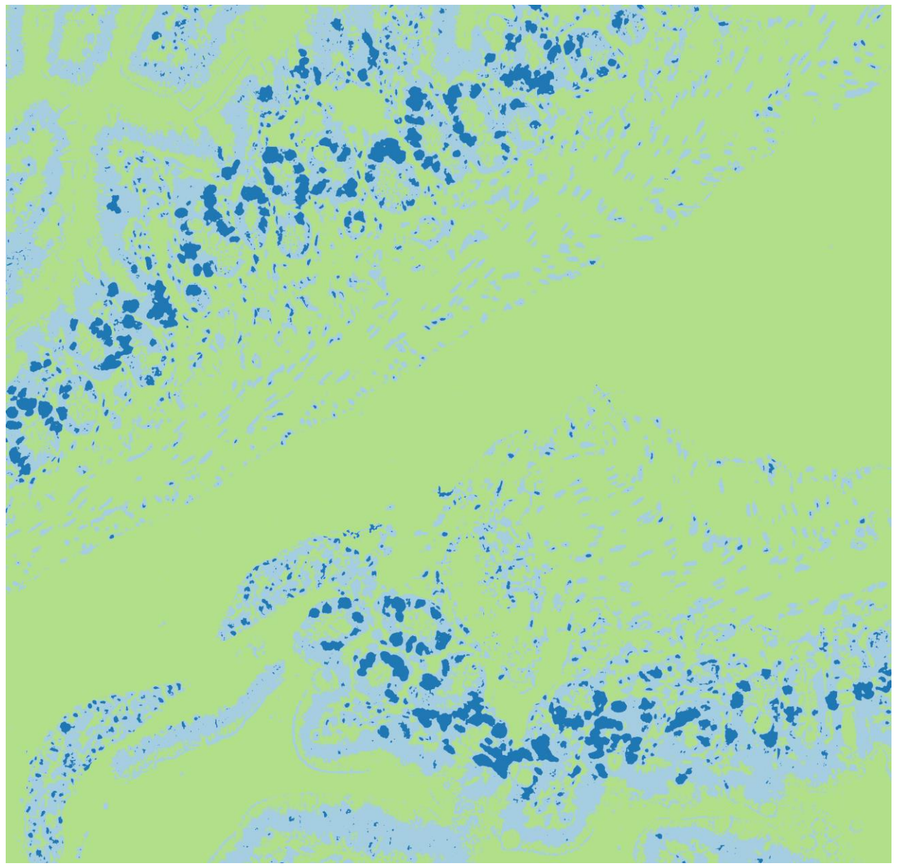


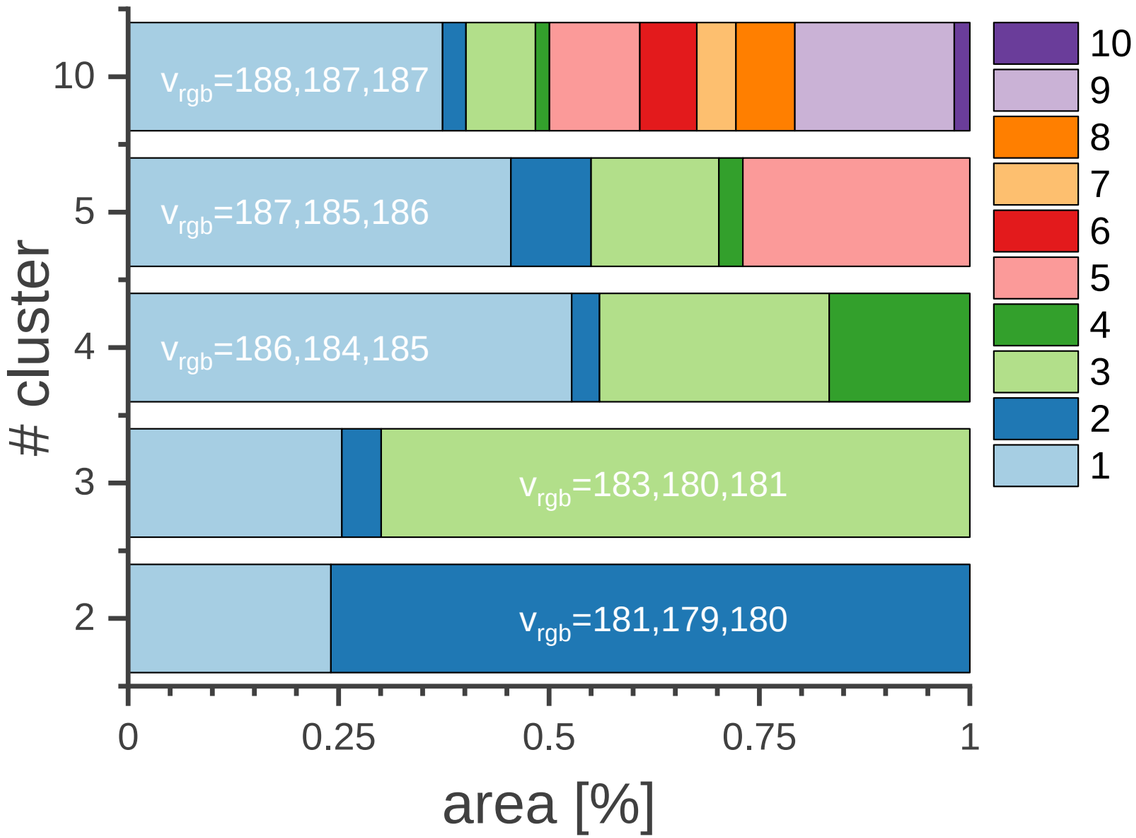
k-means clustering



k-means clustering

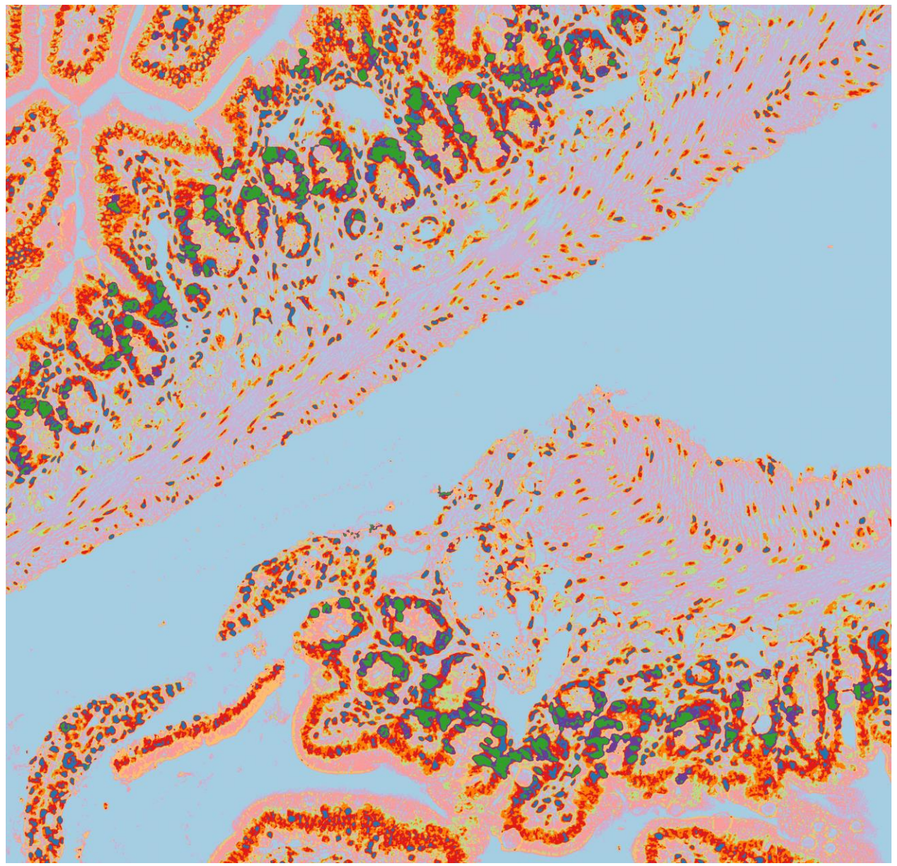
k=3

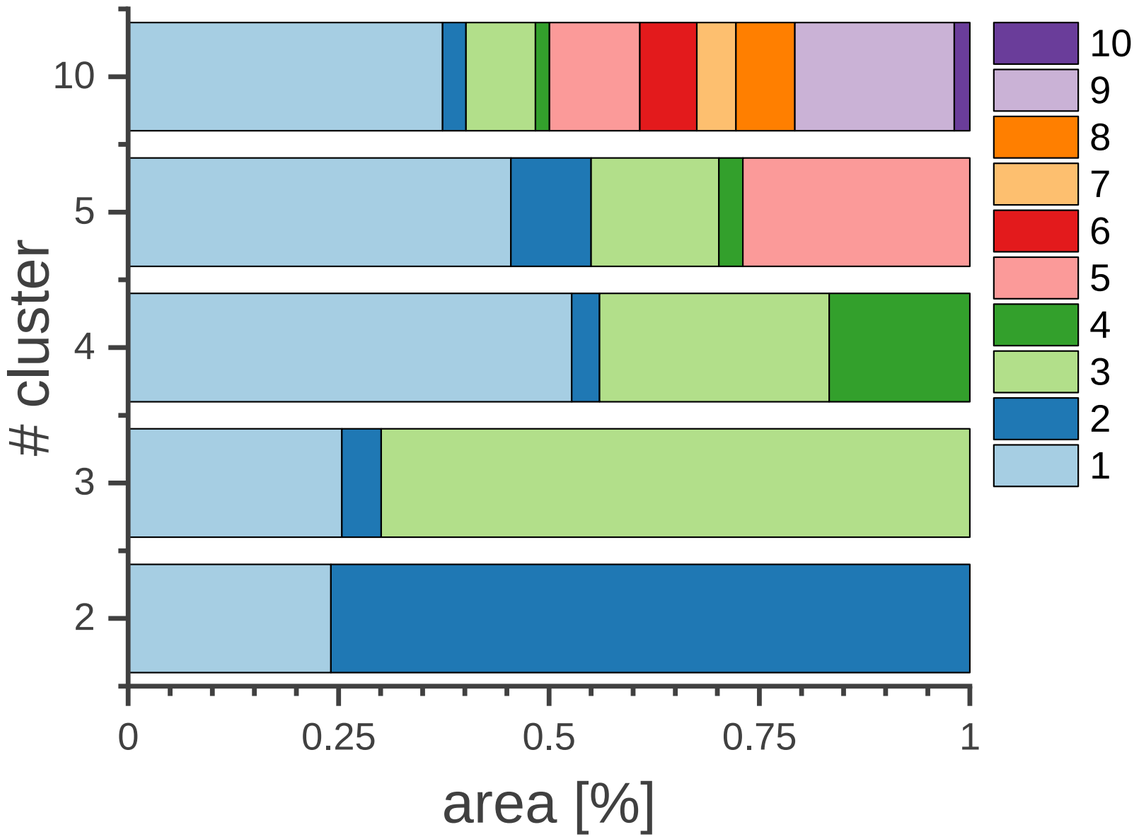




k-means clustering

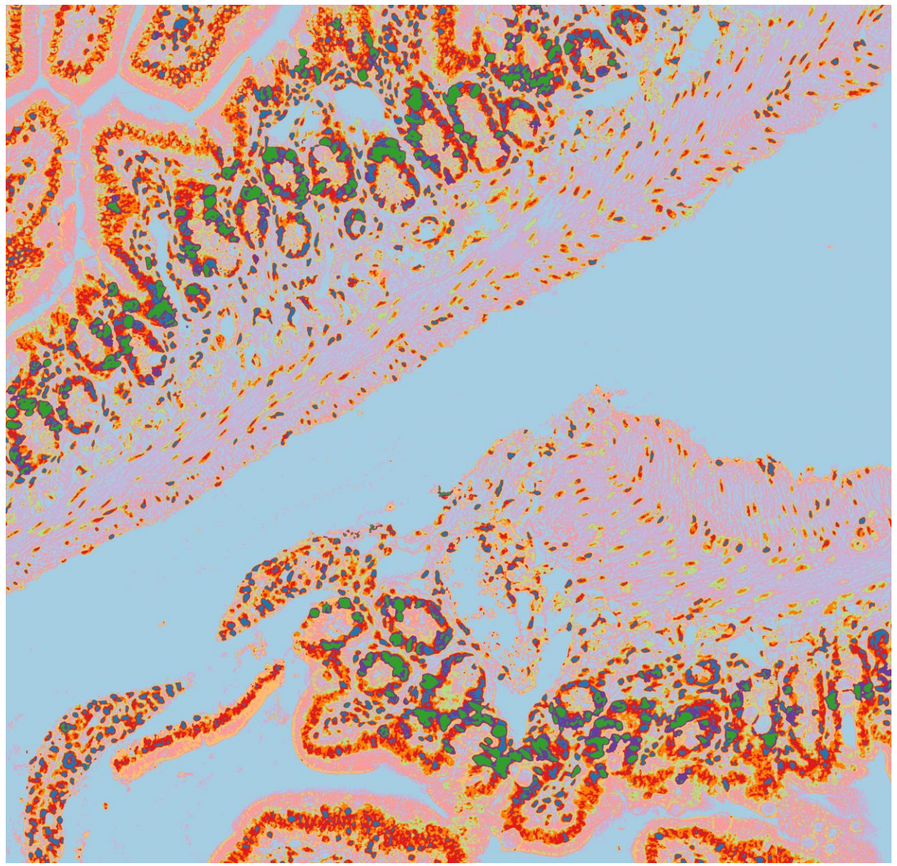
k=10

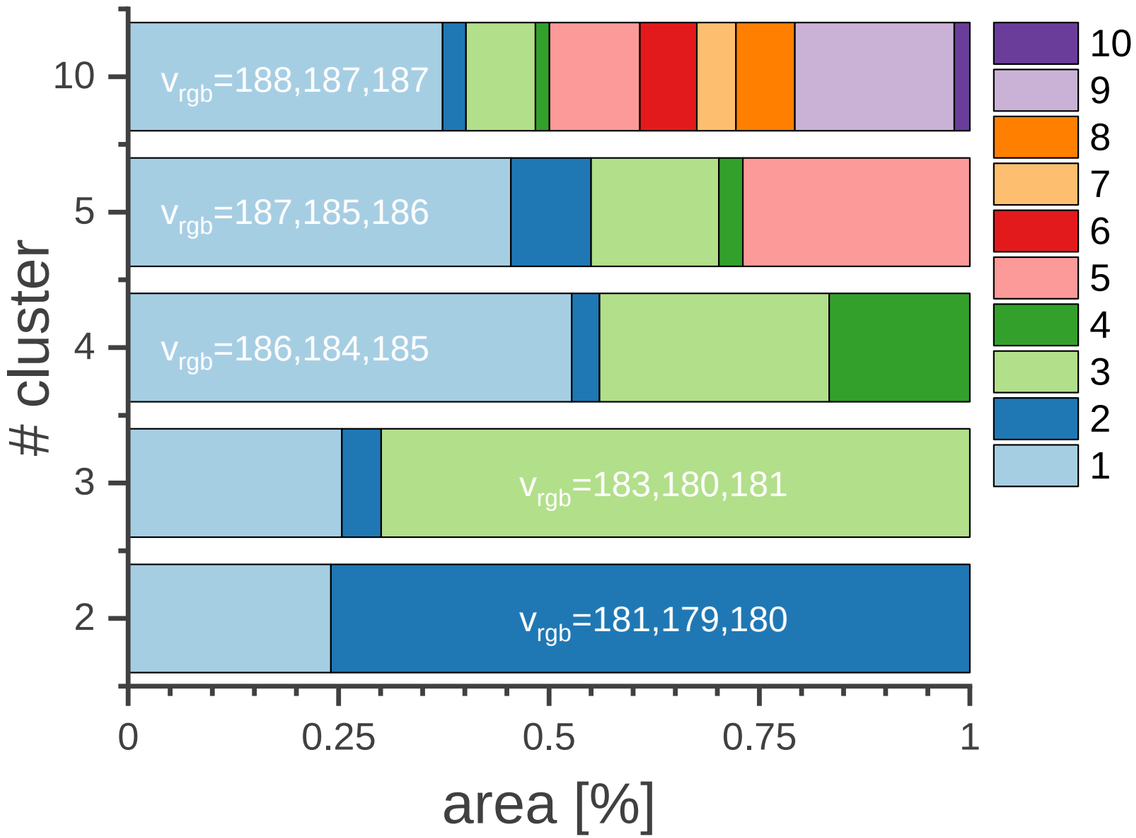




k-means clustering

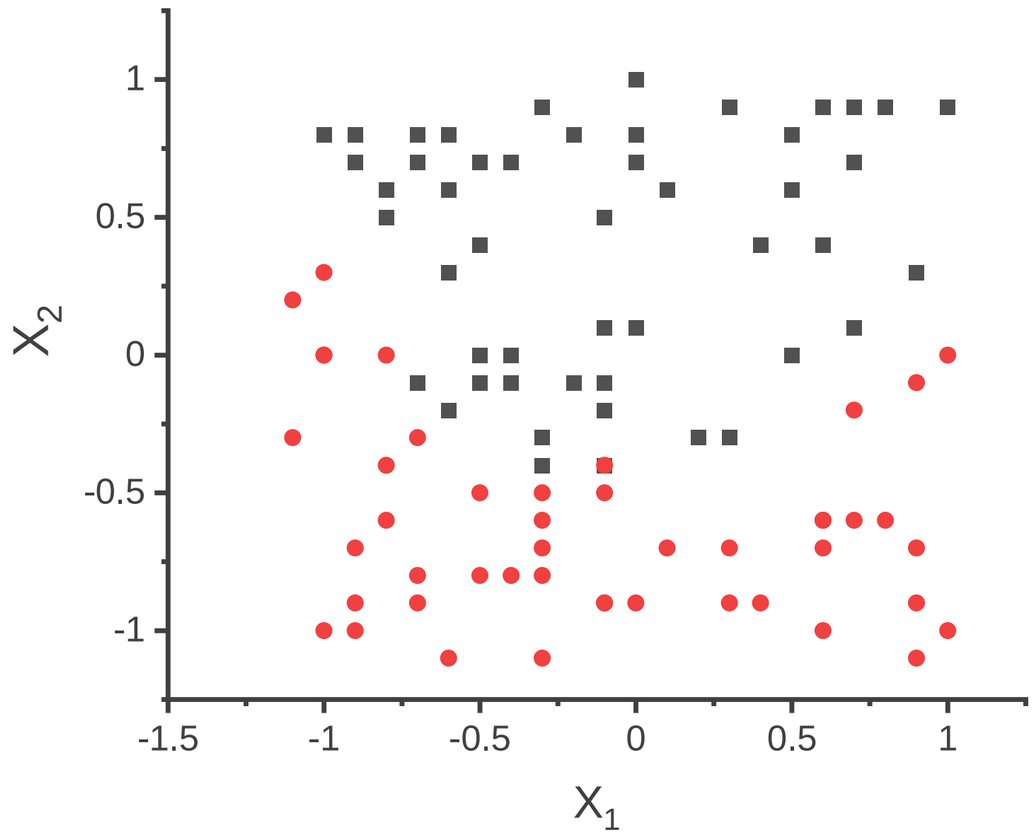
k=10





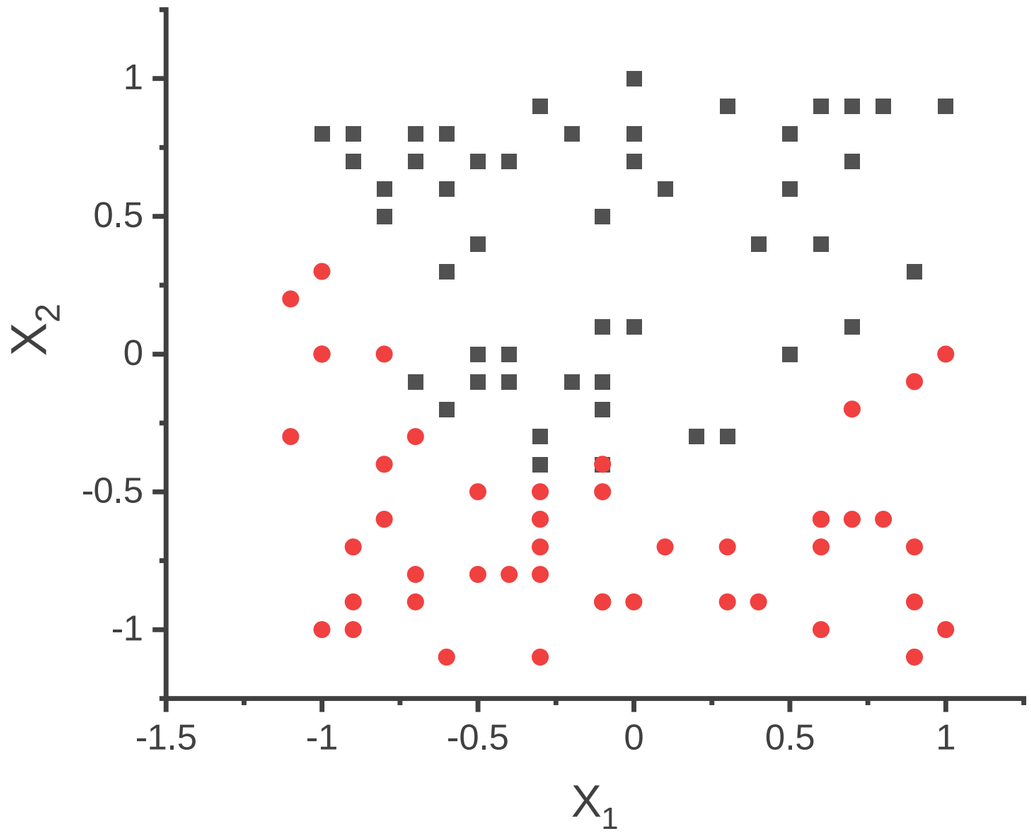
Classification

Classification



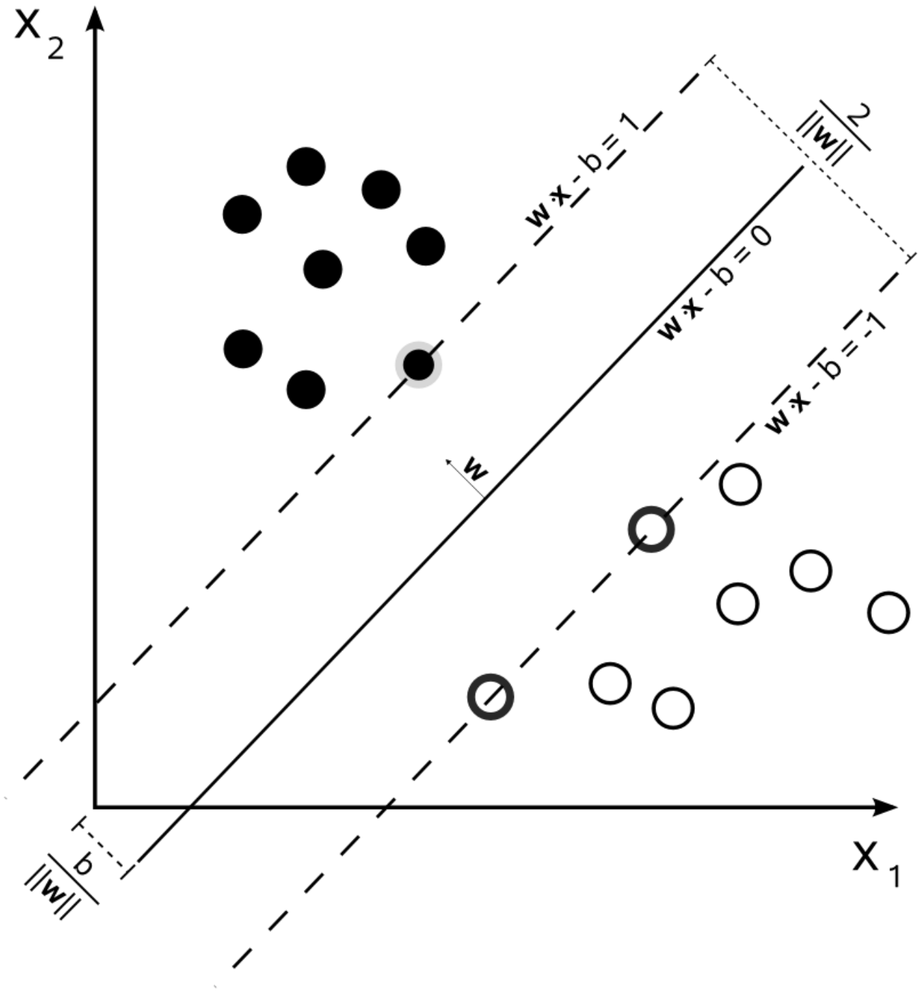
x1	x2	
0.5	-0.9	healthy
0.7	0.4	sick
0.5	-1	healthy
0.9	0.1	healthy
0.2	-0.8	healthy
-0.8	-0.7	healthy
-0.1	0.6	sick
0.3	0	sick
0	1	sick
-1	-0.3	healthy
-1	-0.4	healthy
-0.3	-0.2	sick
0.2	-0.5	healthy

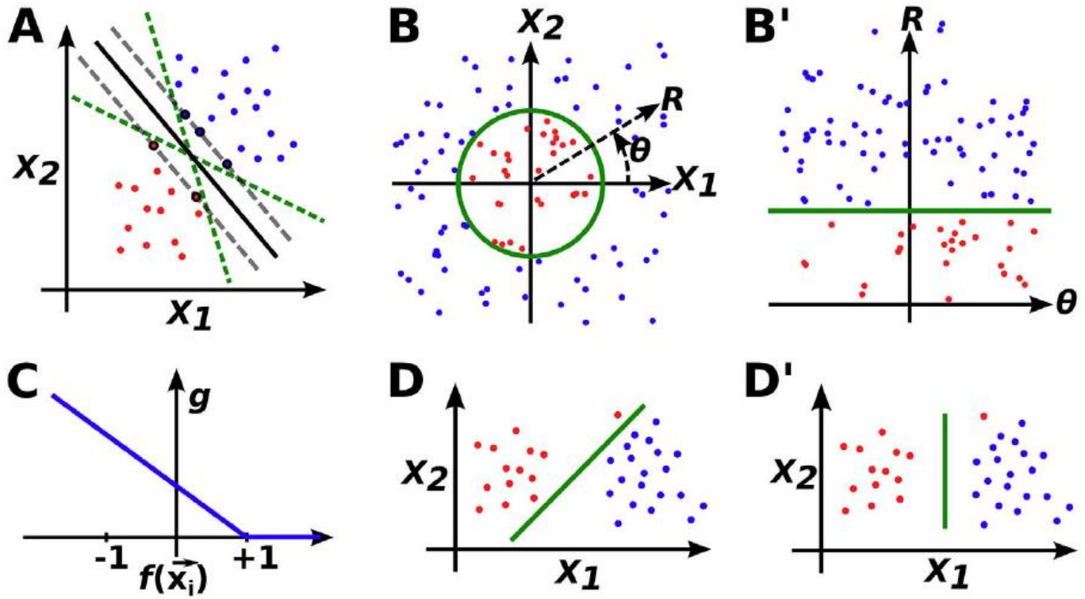
Classification



- Find a function that optimally separates the points from the different classes

Suport Vector Machine (SVM)

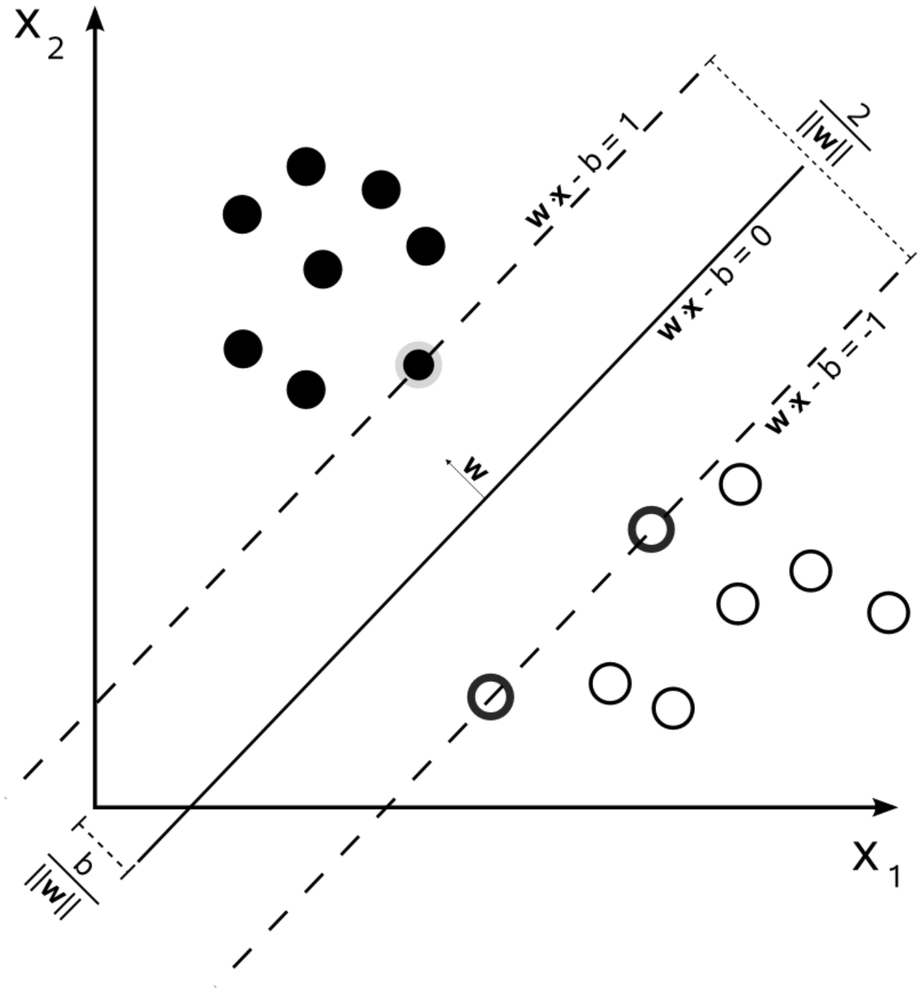




M.F.Z. Wang, R. Fernandez-Gonzalez

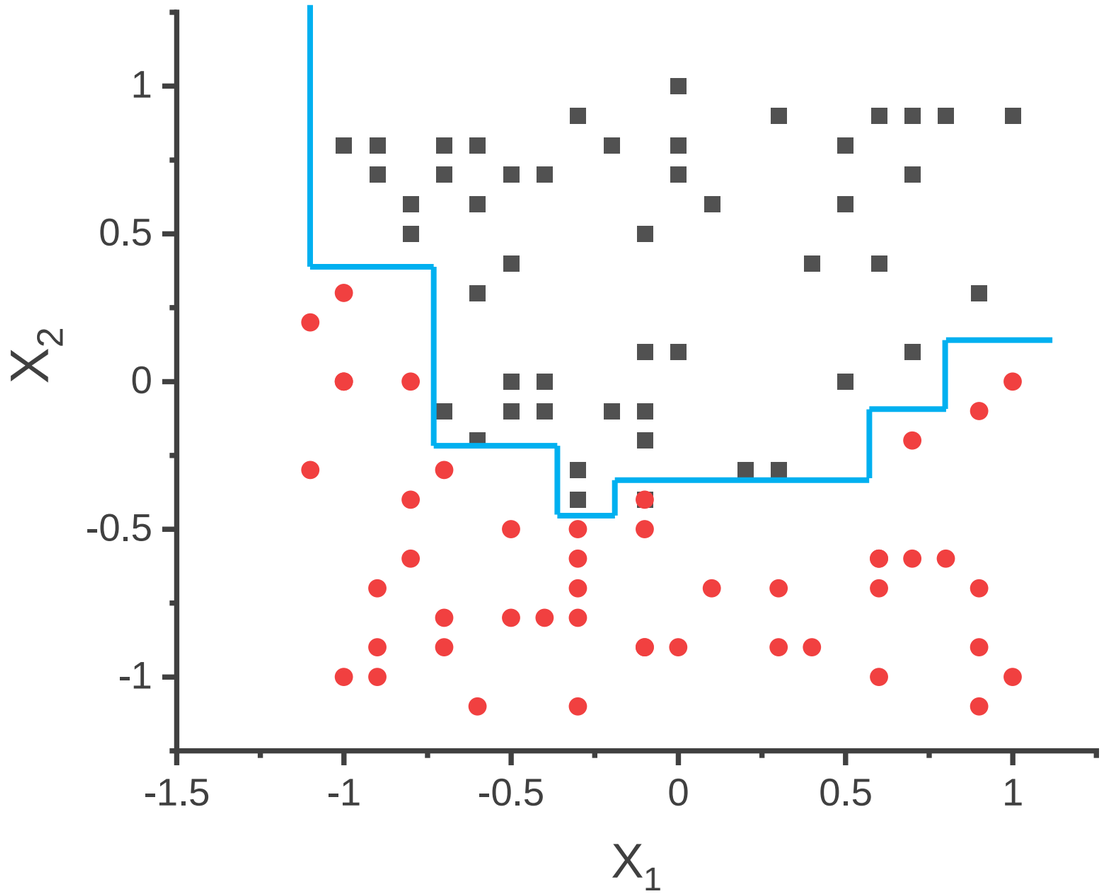
BBA - Proteins and Proteomics 1865 (2017) 1719–1727

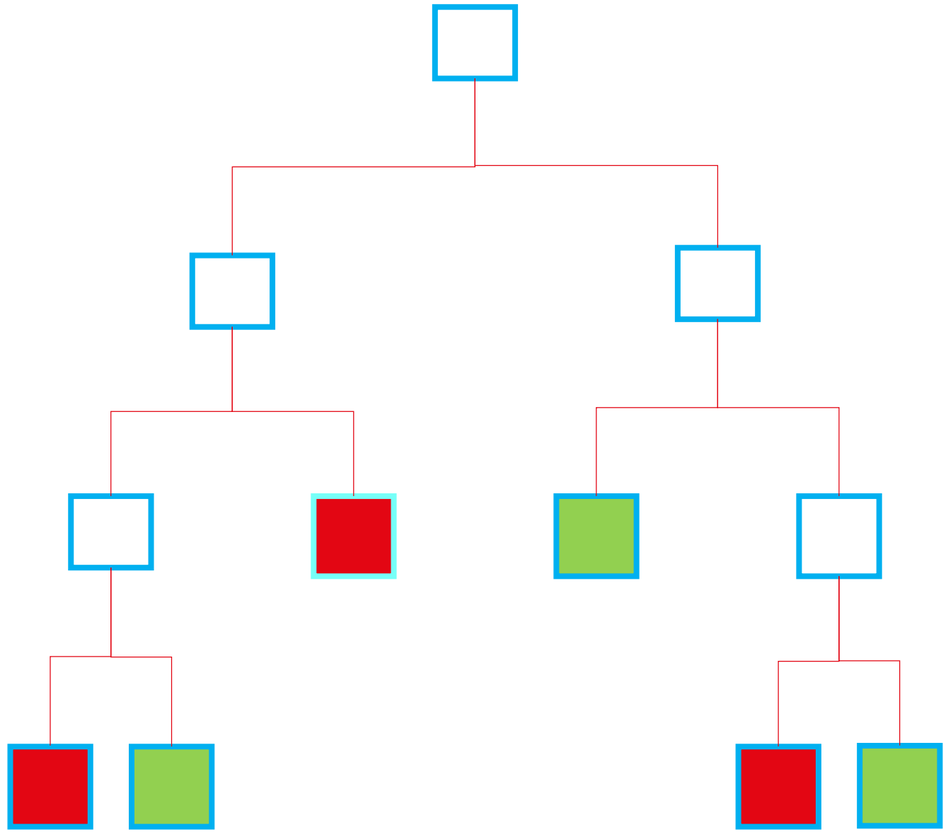
Suport Vector Machine (SVM)



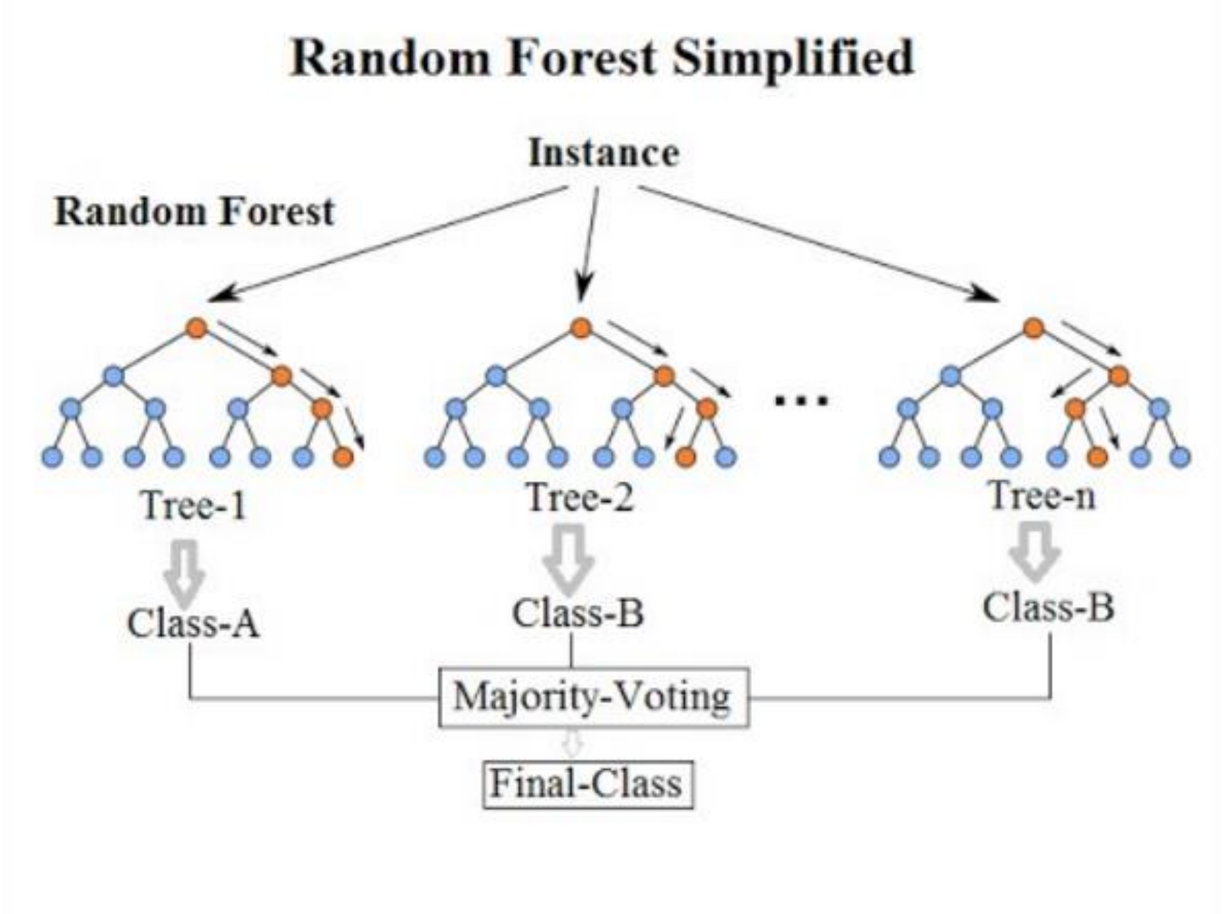
- Hyperplane
- Minimize margin
- Linear classifier
- Best hyperplane = largest separation
- Non-linear implementations exist

Decision tree



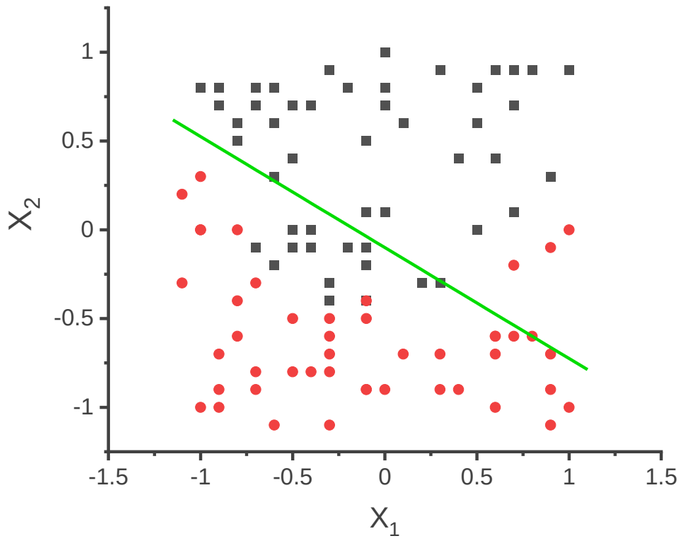


Random Forest

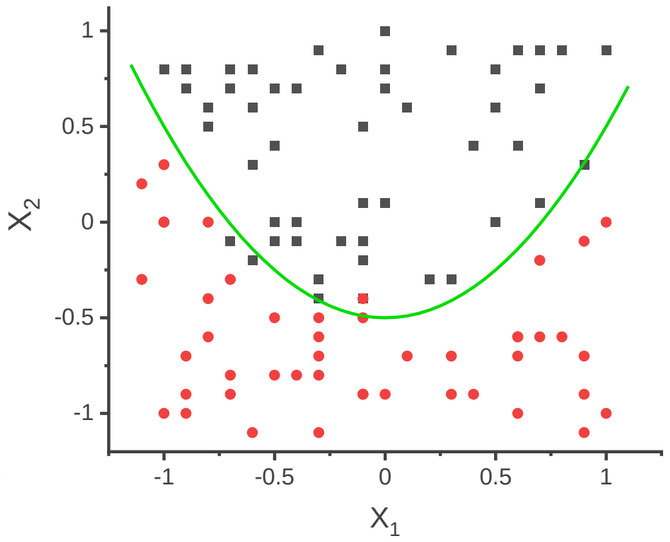


- Ensemble learning technique
- Classification or Regression
- Corrects for overfitting

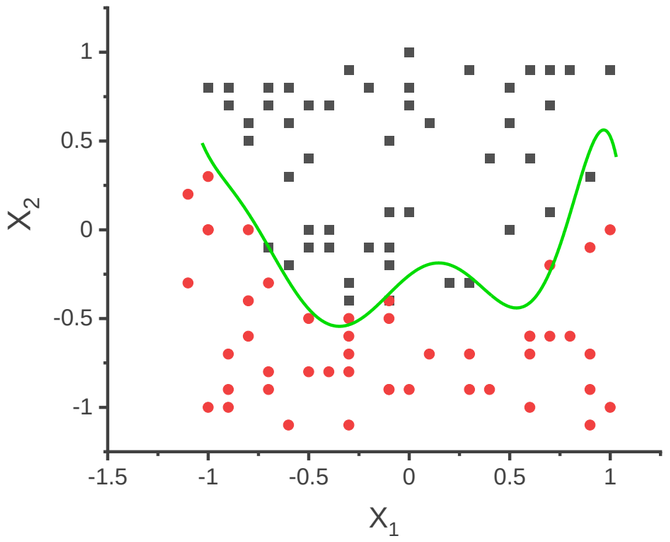
Classification limitations



Underfitting
biases



Optimal

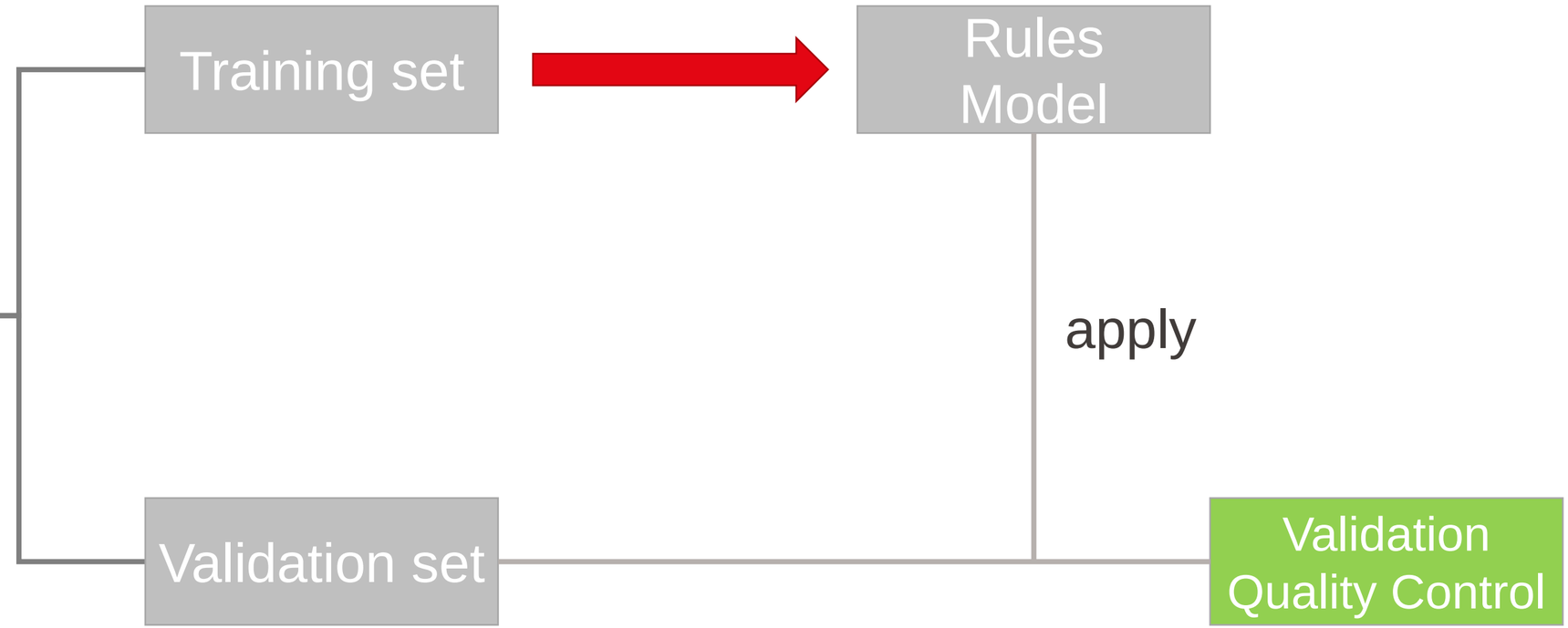


Overfitting
incorporates noise

Supervised learning workflow

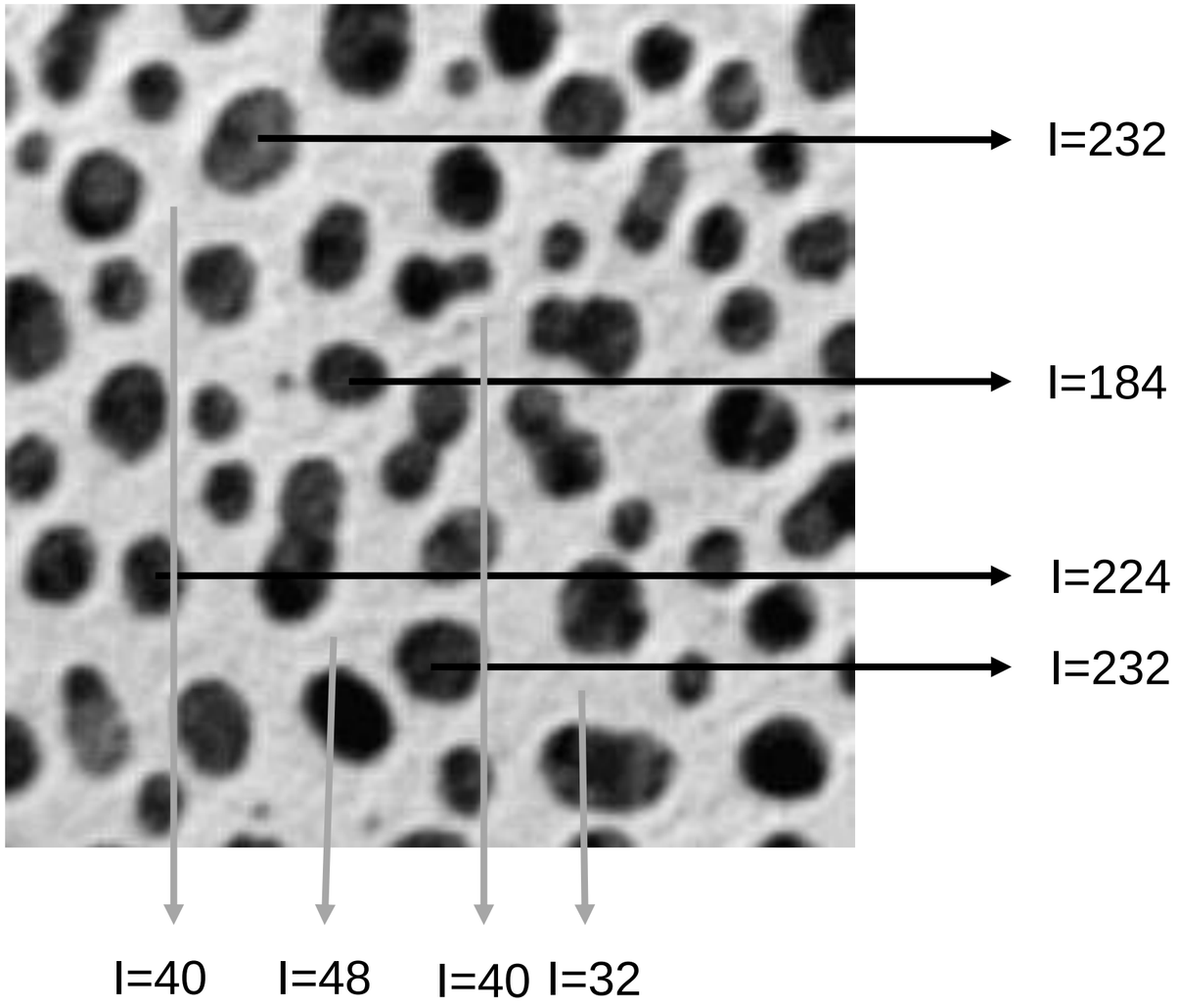
x1	x2	
0.5	-0.9	healthy
0.7	0.4	sick
0.5	-1	healthy
0.9	0.1	healthy
0.2	-0.8	healthy
-0.8	-0.7	healthy
-0.1	0.6	sick
0.3	0	sick
0	1	sick
-1	-0.3	healthy
-1	-0.4	healthy
-0.3	-0.2	sick
0.2	-0.5	healthy

Annotated data

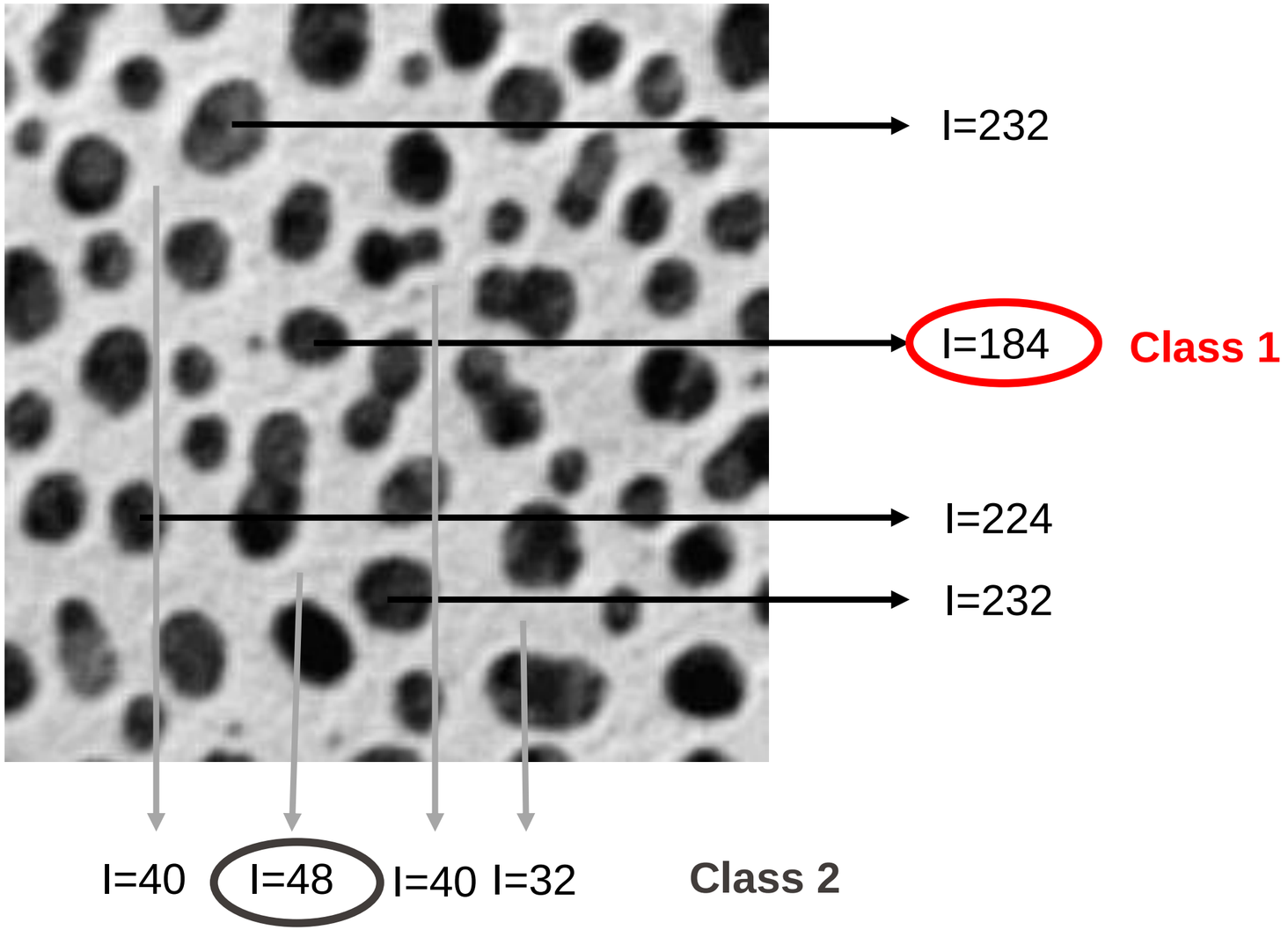


Pixel Classification

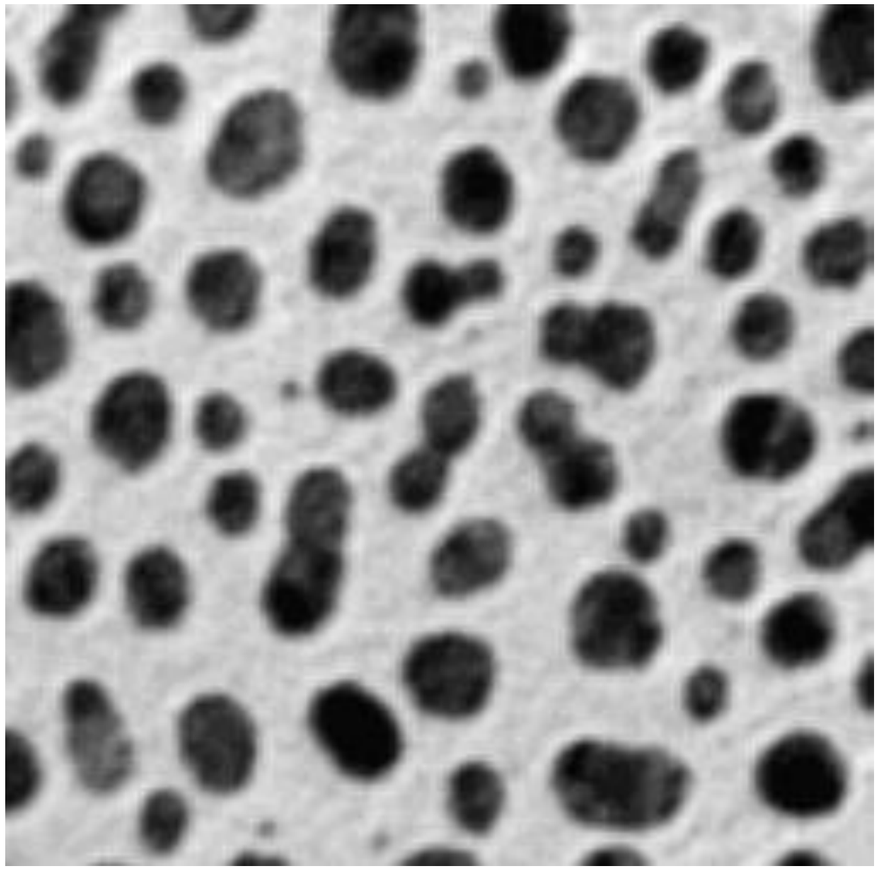
Pixel classification



Pixel classification



Pixel classification



Class 1	Class 2
232	40
184	48
224	40
232	32

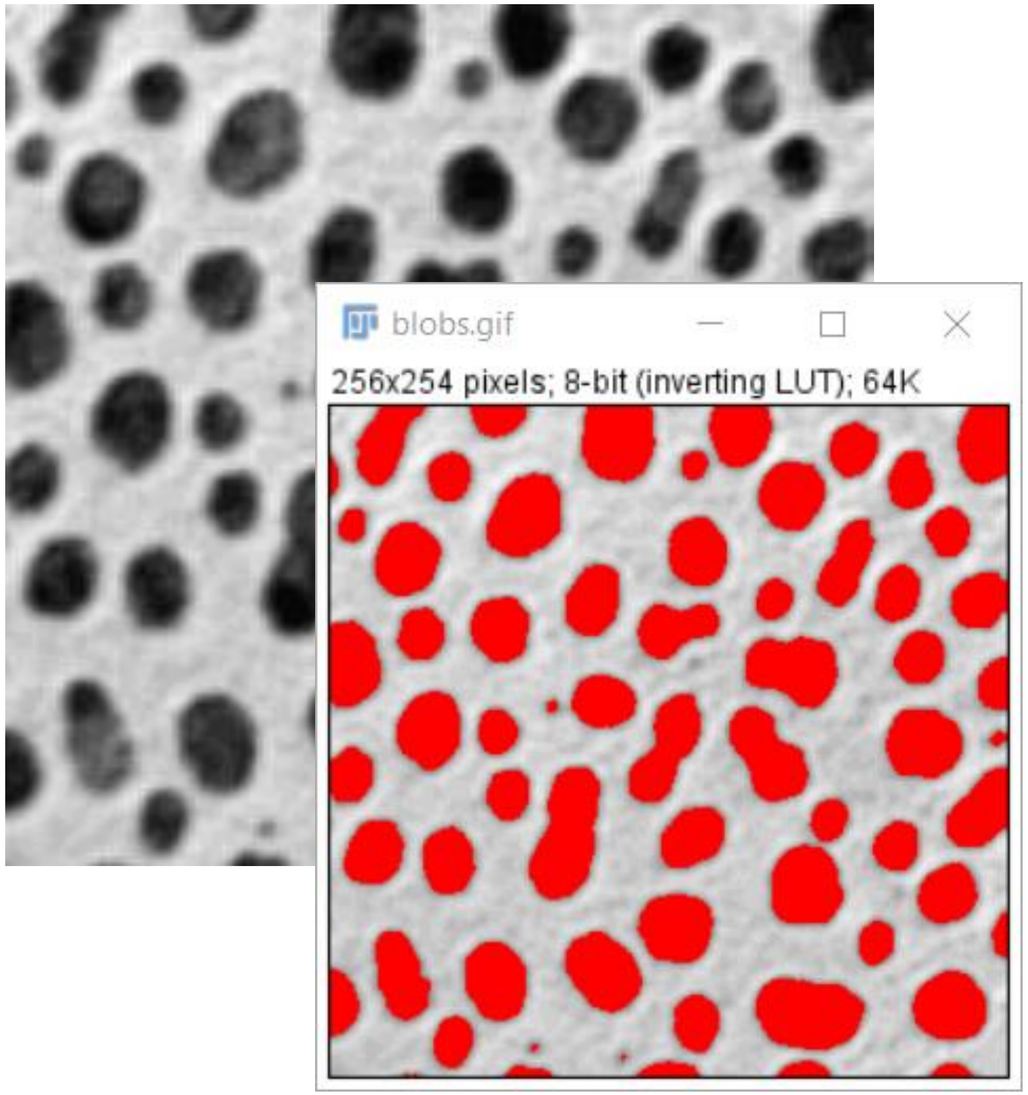
$Min_{C1}=184$

$Max_{C2}=48$

$$S = Max_{C2} + \frac{Min_{C1} - Max_{C2}}{2}$$

$$S = 116$$

Pixel classification



Class 1	Class 2
232	40
184	48
224	40
232	32

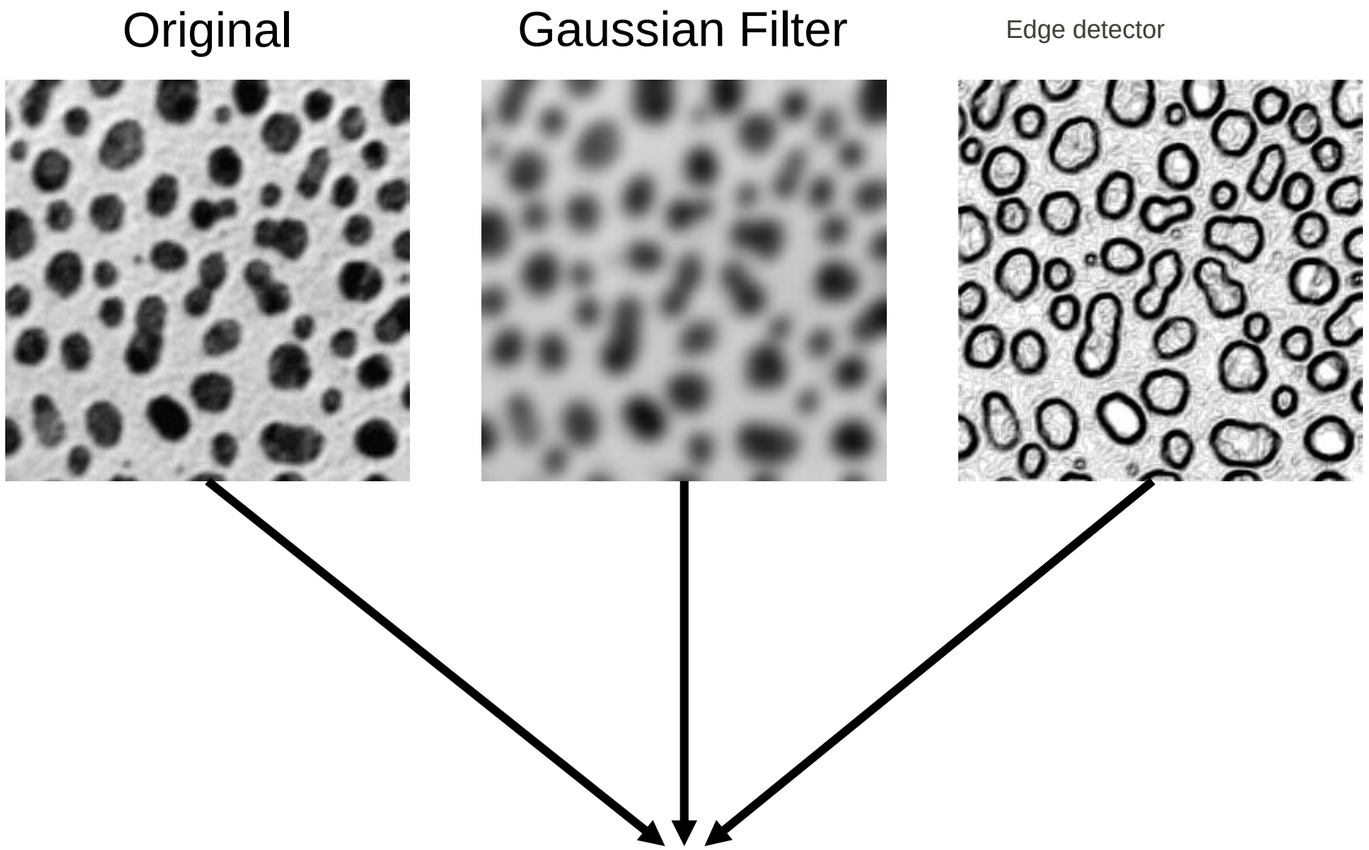
Min_{C1}=184

Max_{C2}=48

$$S = Max_{C2} + \frac{Min_{C1} - Max_{C2}}{2}$$

$$S = 116$$

Pixel classification



3 Values per pixel

=

3 Features/pixel

Pixel Classification

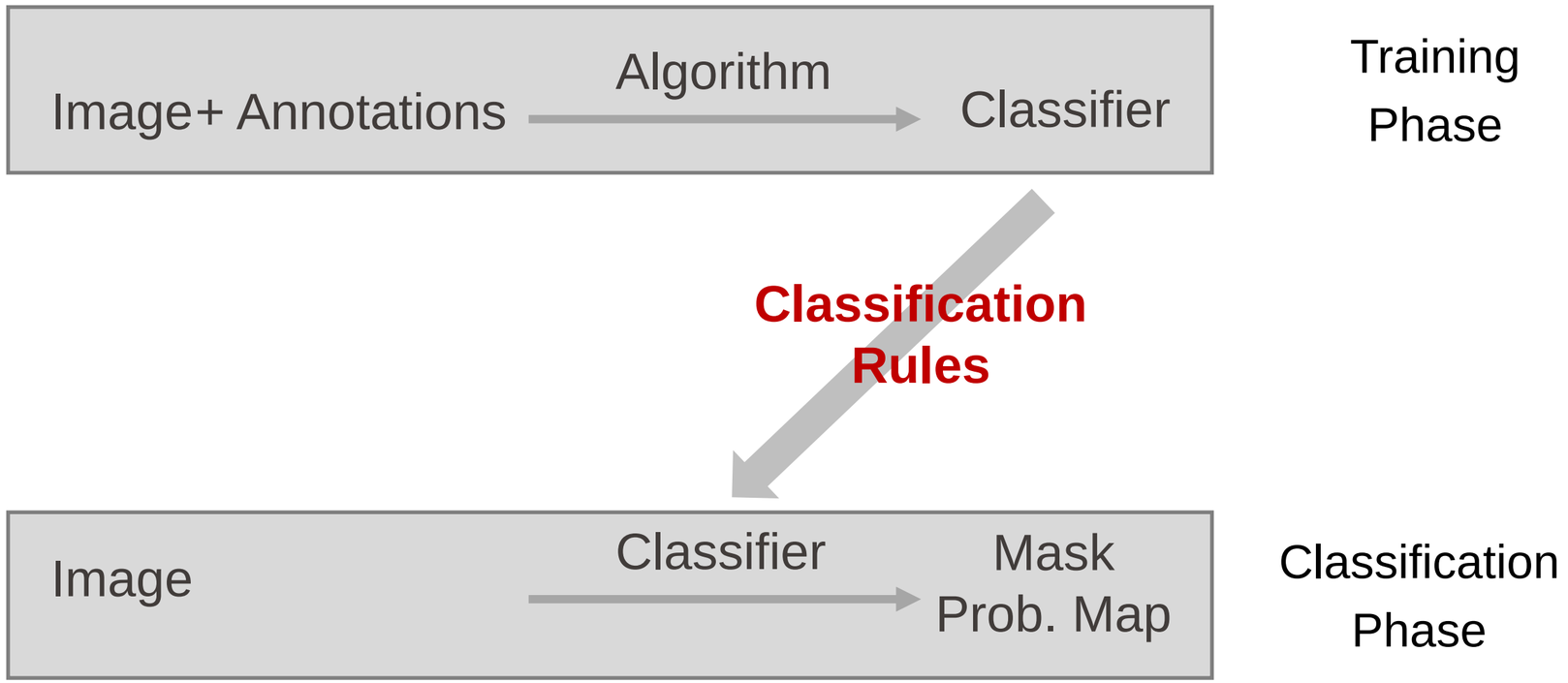
Training Phase

Pixel Classification

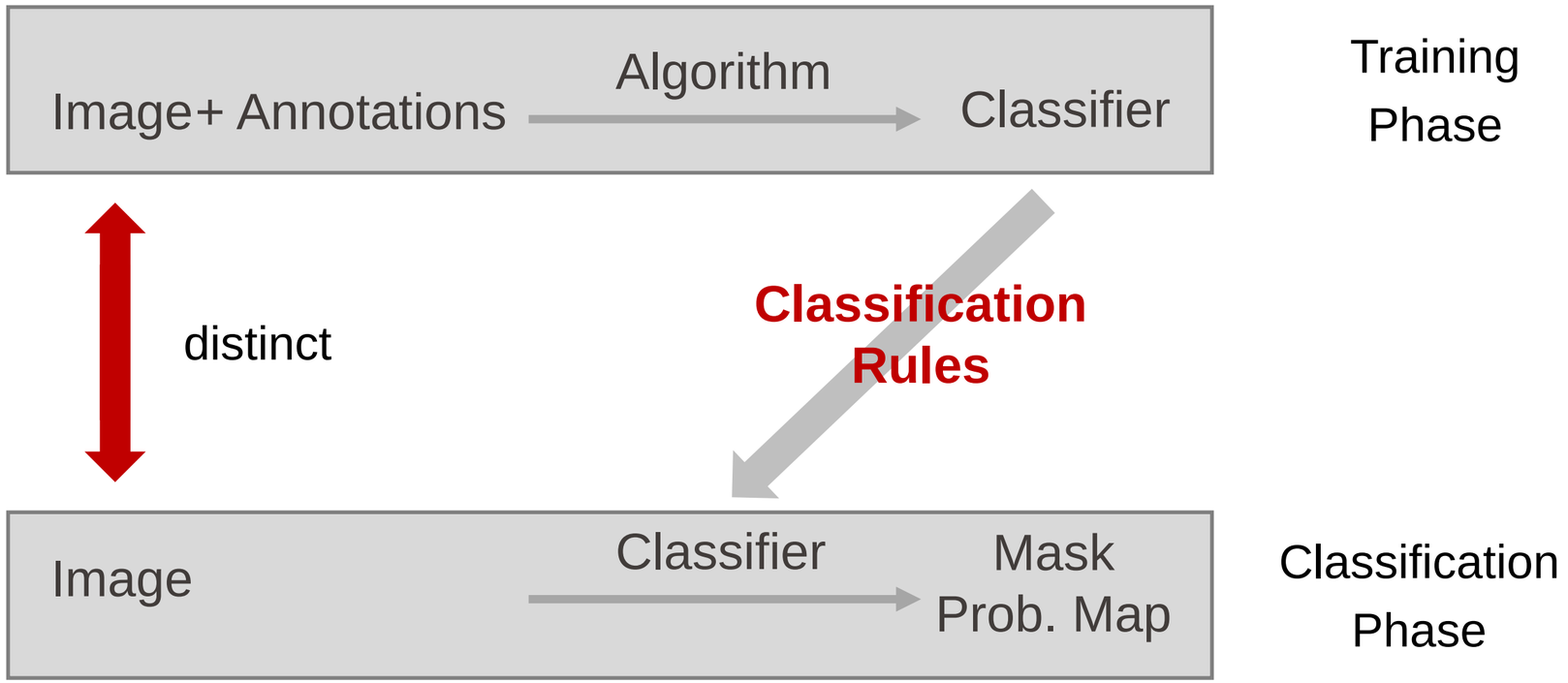
Training Phase

Classification Rules

Pixel Classification

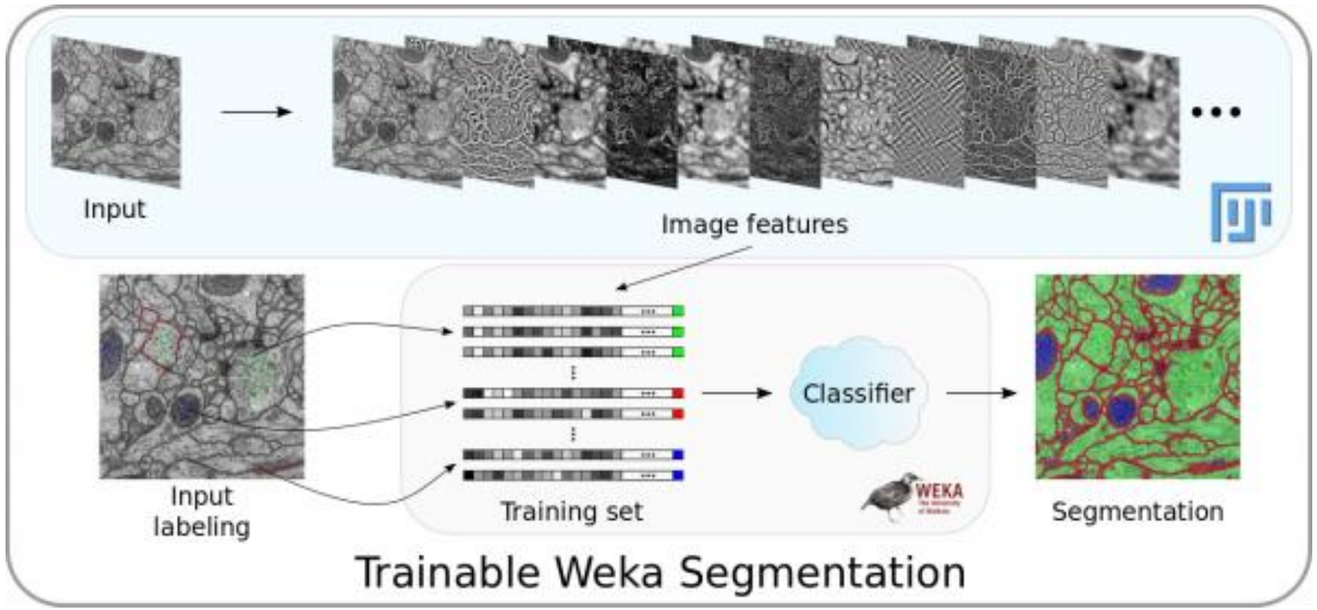


Pixel Classification

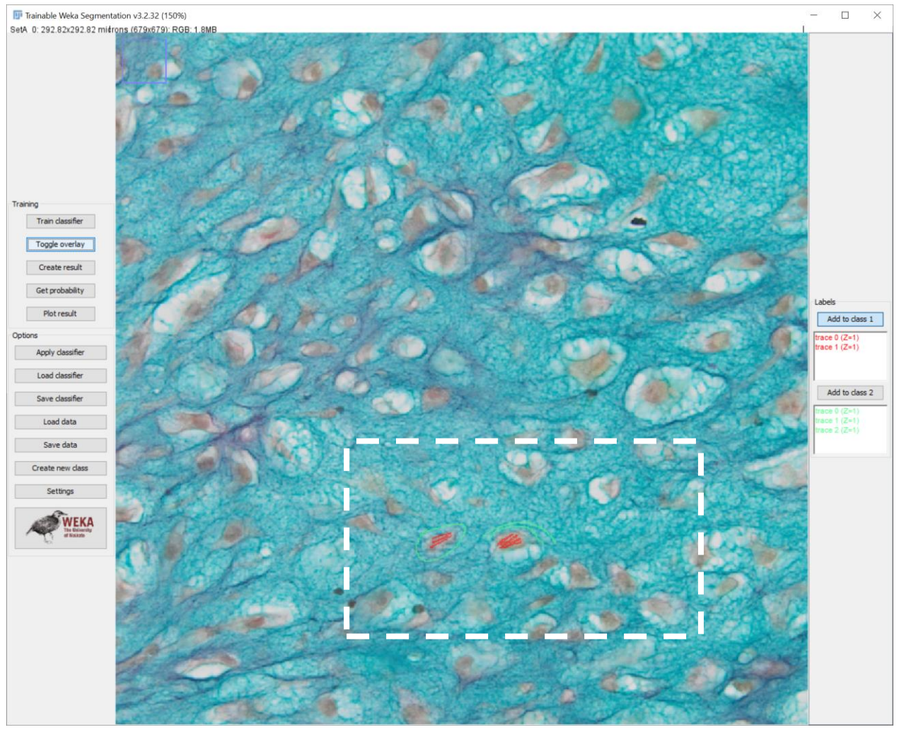


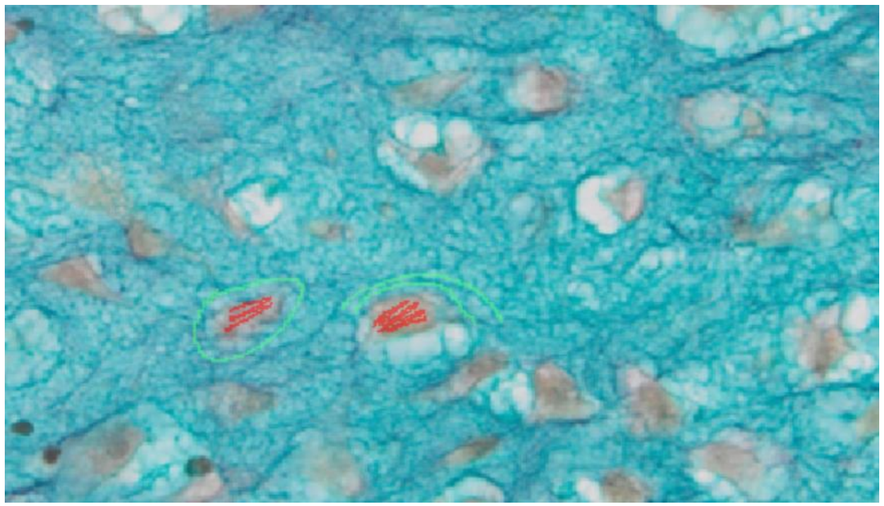
Trainable Weka Segmentation

- image.net documentation page

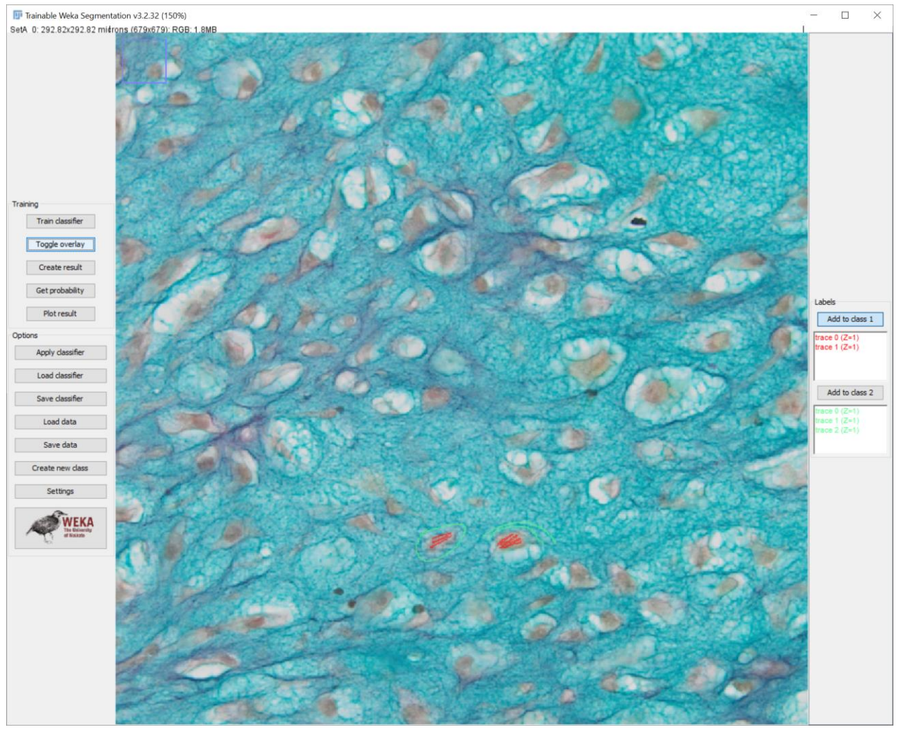


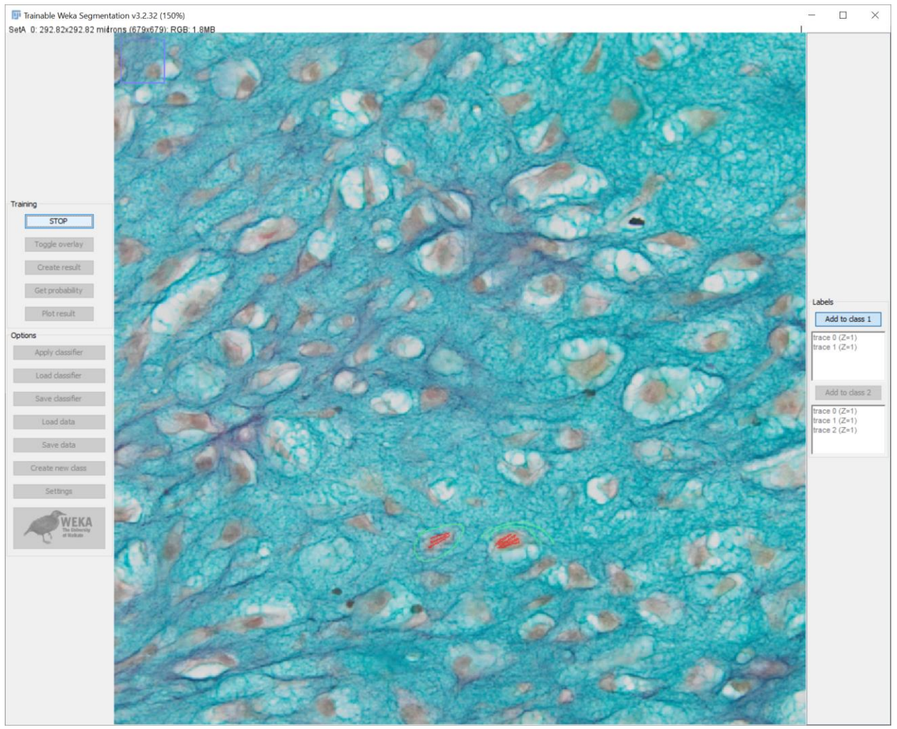
Weka : Annotations



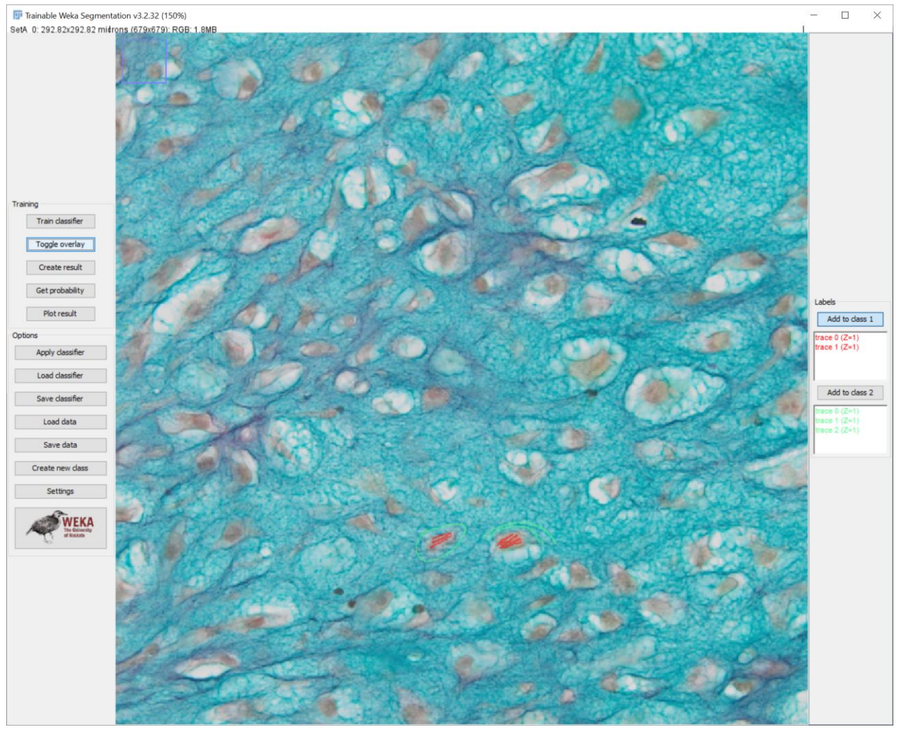


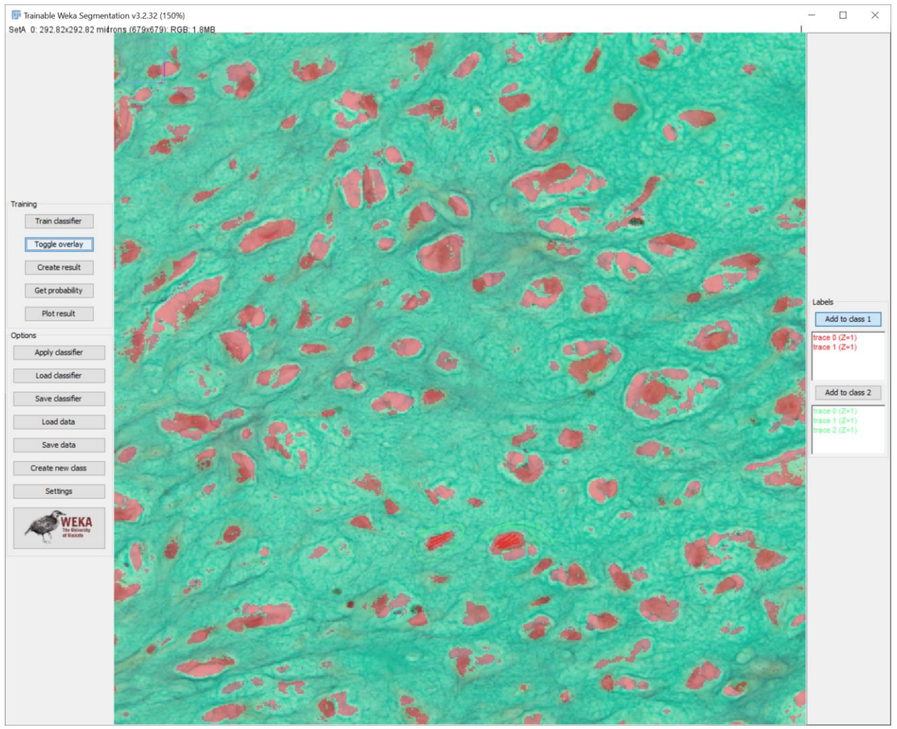
Weka : Training



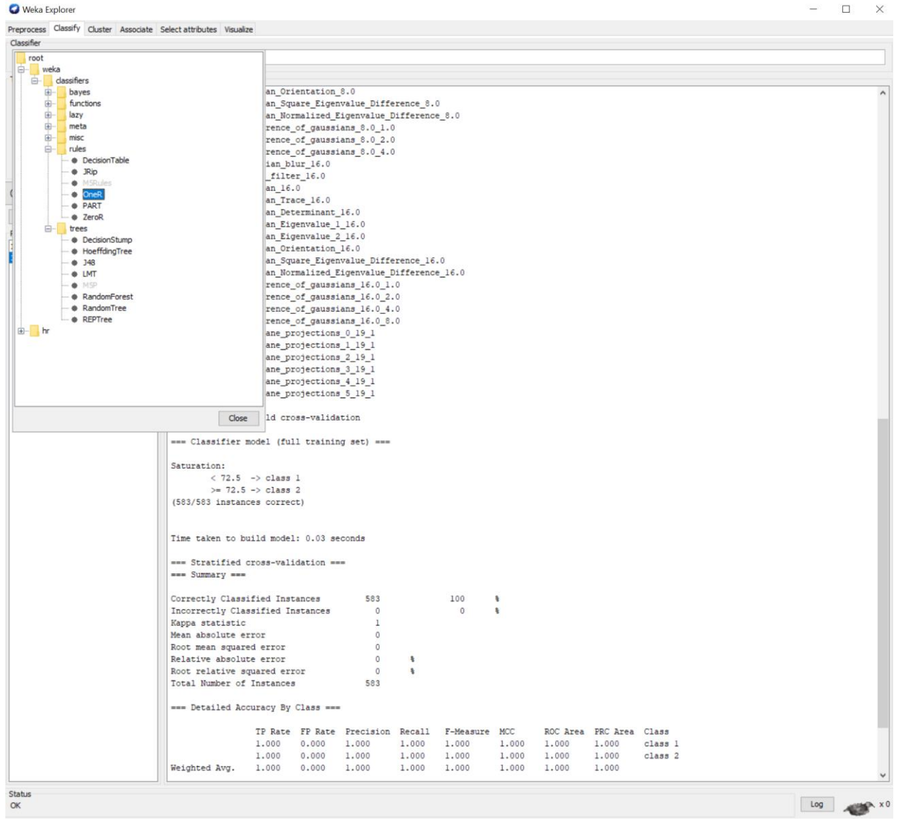


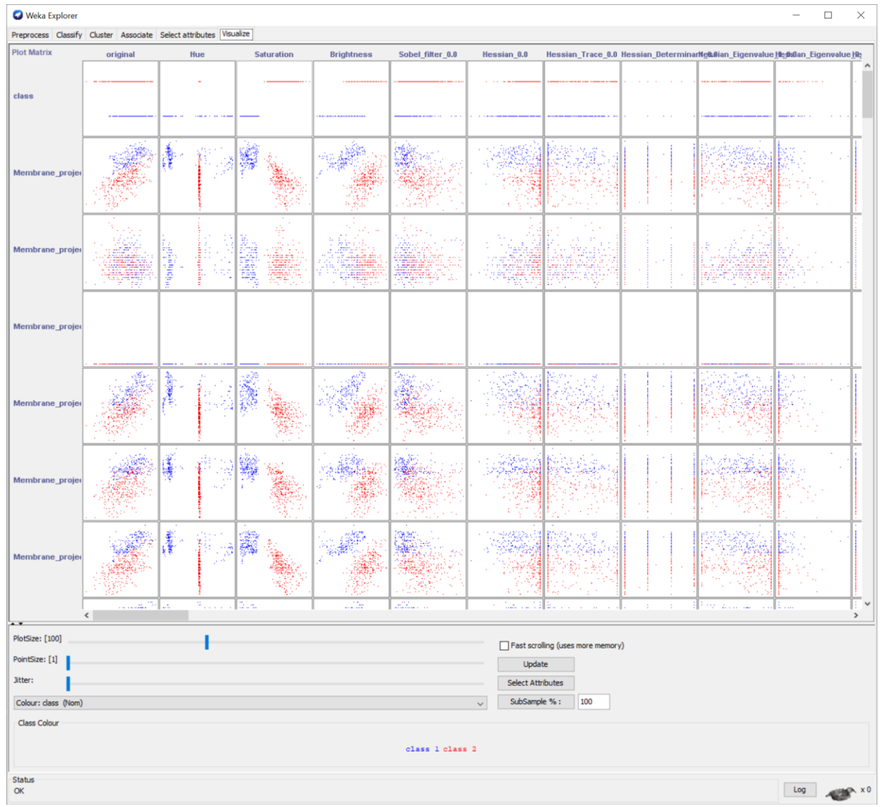
Weka : Prediction





Weka Nice Educational Features

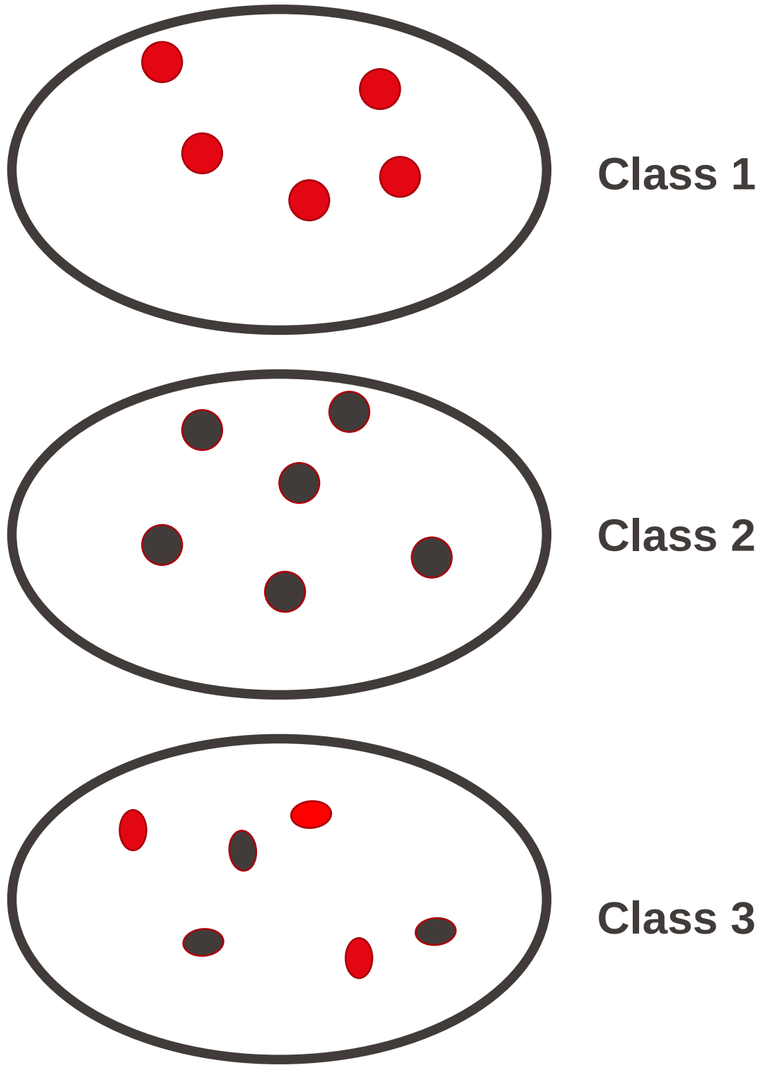


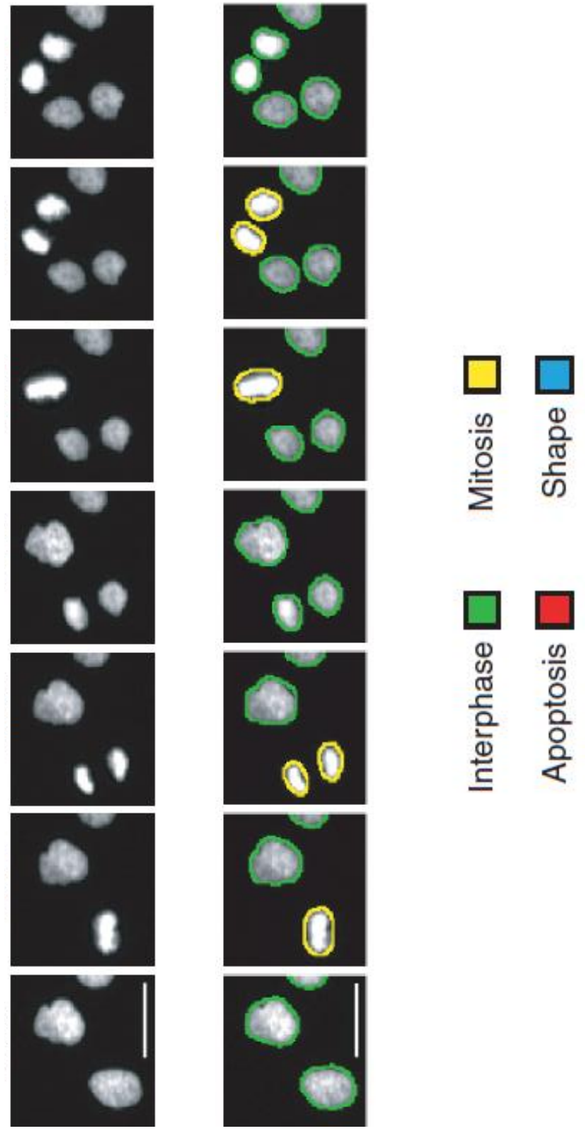


Object Classification

Object classification

$$O_i = \begin{bmatrix} f_1 \\ f_2 \\ f_3 \\ \vdots \\ f_n \end{bmatrix}$$





Pixels Classification

- Images + Pixel Annotations

Pixels + Objects Classification

- Images + Pixel Annotations

Probabilities + Thresholding
+ Connected Component

- Objects

Pixels + Objects Classification

- Images + Pixel Annotations

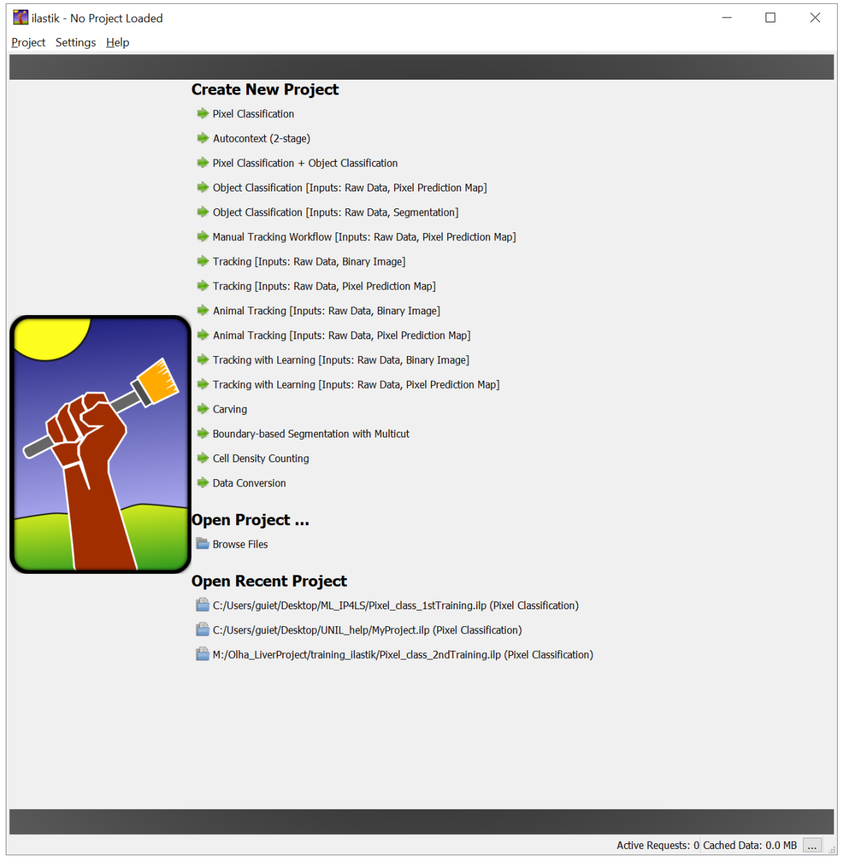
Probabilities + Thresholding
+ Connected Component

- Objects + Objects Annotations

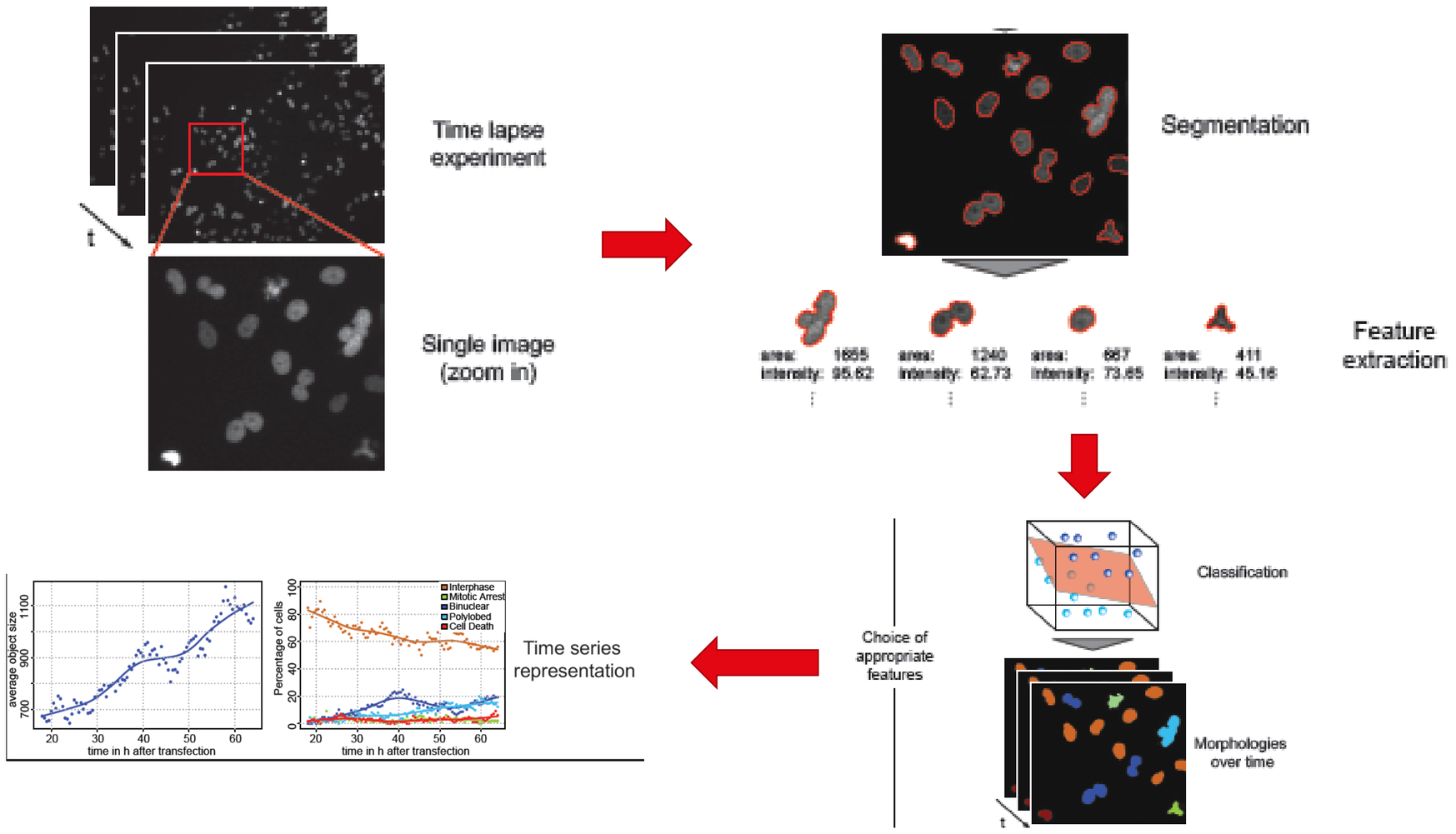
- Classified Objects

ilastik


ilastik
the interactive learning and segmentation toolkit
Next workshop! The next ilastik workshop will take place on December 4th, 2017 (9:30-15:00), in Heidelberg, in the Mathematikon building A, 2nd Floor, Room 103: PC Pool. Sign up at this google form and preview the workshop program if you'd like to attend. As usual, we will include a session to work on your own data and a "show and tell" hour. The workshop is organized within the ilastik team activities in the SFB1129, but we'd be happy to also admit outside people for the remaining places.
ilastik is a simple, user-friendly tool for interactive image classification, segmentation and analysis. It is built as a modular software framework, which currently has workflows for automated (supervised) pixel- and object-level classification, automated and semi-automated object tracking, semi-automated segmentation and object counting without detection. Most analysis operations are performed lazily, which enables targeted interactive processing of data subvolumes, followed by complete volume analysis in offline batch mode. Using it requires no experience in image processing.
Pixel Classification
Semantic segmentation using predicted class probabilities

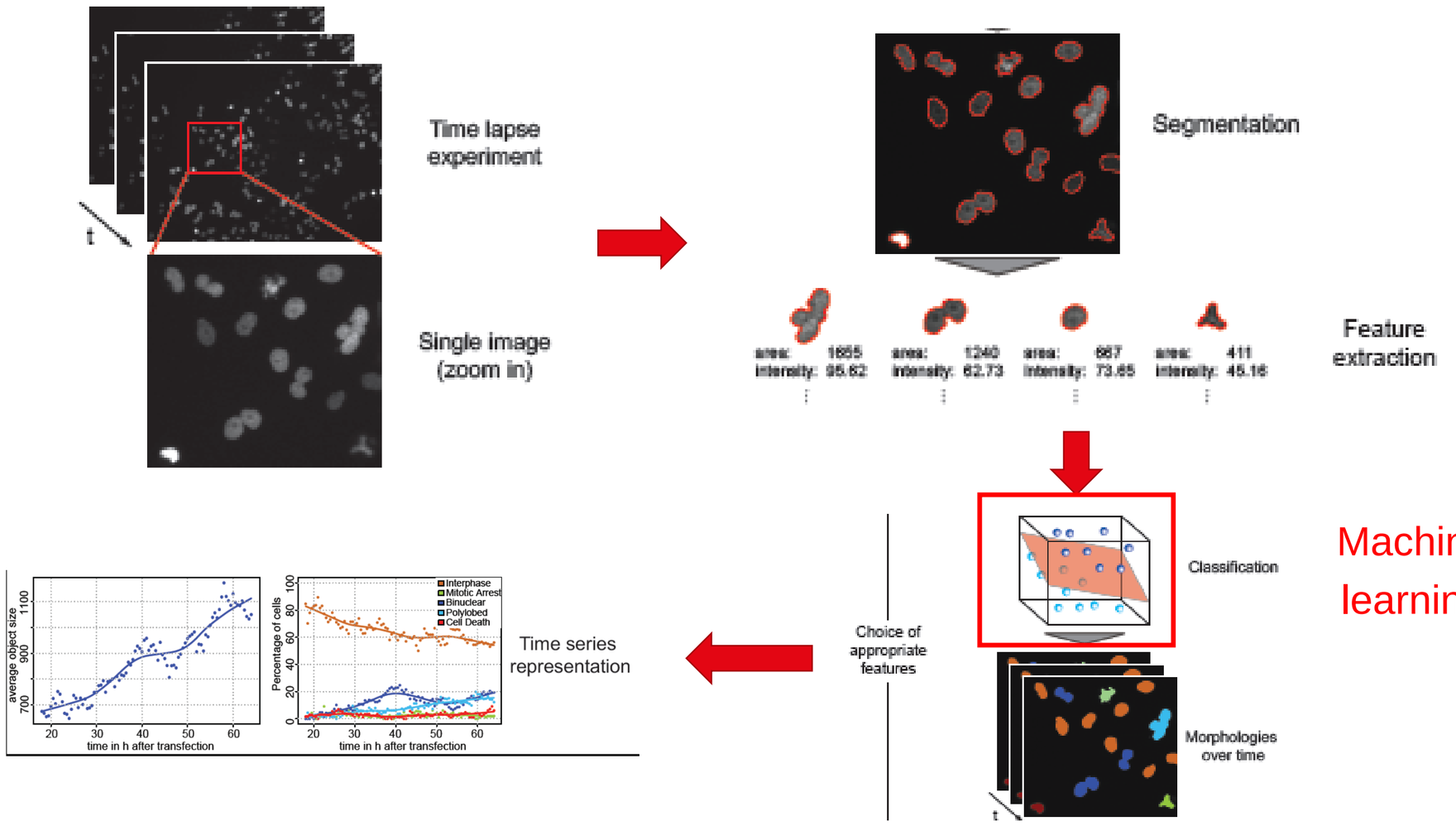

ilastik - No Project Loaded
Project Settings Help
Create New Project
Pixel Classification
Autocontext (2-stage)
Pixel Classification + Object Classification
Object Classification [Inputs: Raw Data, Pixel Prediction Map]
Object Classification [Inputs: Raw Data, Segmentation]
Manual Tracking Workflow [Inputs: Raw Data, Pixel Prediction Map]
Tracking [Inputs: Raw Data, Binary Image]
Tracking [Inputs: Raw Data, Pixel Prediction Map]
Animal Tracking [Inputs: Raw Data, Binary Image]
Animal Tracking [Inputs: Raw Data, Pixel Prediction Map]
Tracking with Learning [Inputs: Raw Data, Binary Image]
Tracking with Learning [Inputs: Raw Data, Pixel Prediction Map]
Carving
Boundary-based Segmentation with Multicut
Cell Density Counting
Data Conversion
Open Project ...
Browse Files
Open Recent Project
C:/Users/guiet/Desktop/ML_IP4LS/Pixel_class_1stTraining.ilp (Pixel Classification)
C:/Users/guiet/Desktop/UNIL_help/MyProject.ilp (Pixel Classification)
M:/Olha_LiverProject/training_ilastik/Pixel_class_2ndTraining.ilp (Pixel Classification)
Active Requests: 0 Cached Data: 0.0 MB

Automated Image Analysis



Neumann, B., M. Held, et al. (2006). "High-throughput RNAi screening by time-lapse imaging of live human cells." Nat Methods **3**(5): 385-390.

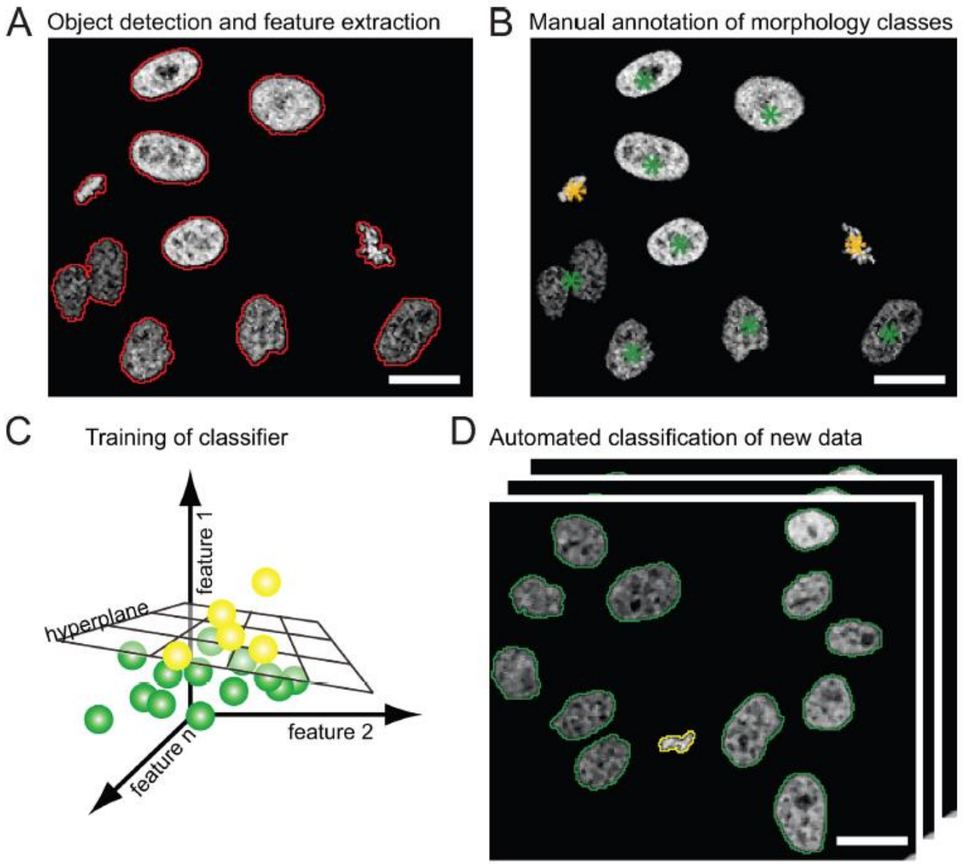
Automated Image Analysis



Neumann, B., M. Held, et al. (2006). "High-throughput RNAi screening by time-lapse imaging of live human cells." Nat Methods **3**(5): 385-390.

Supervised learning

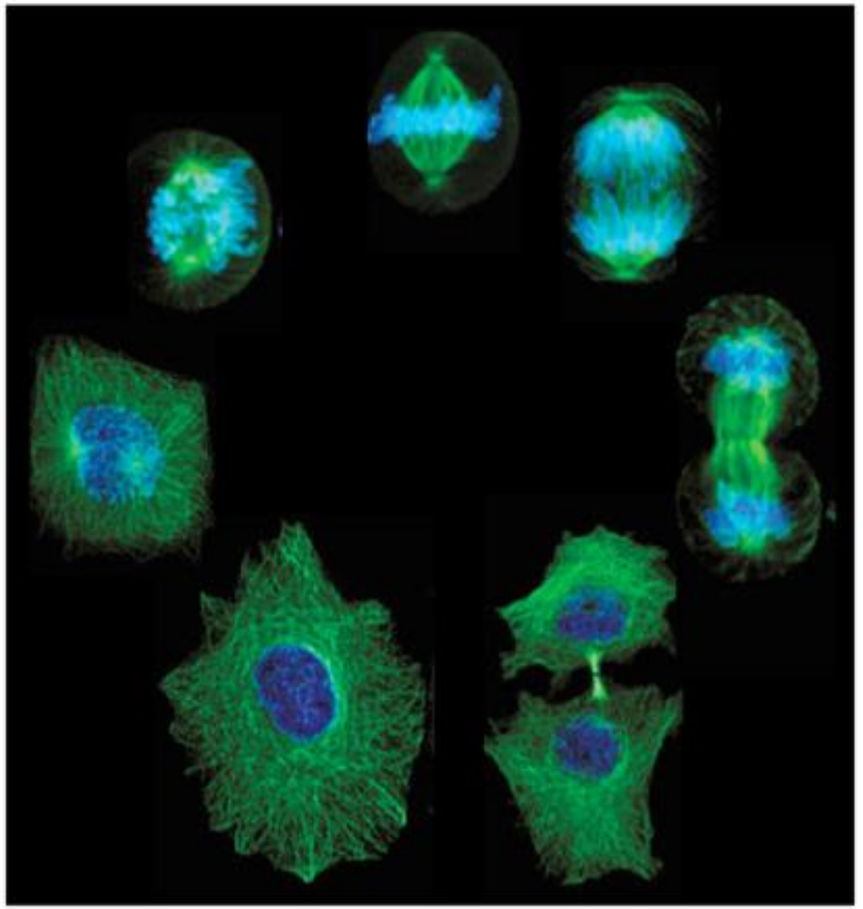
- Object segmentation
- Feature extraction
- Manual Annotation
- Training Classifier
- Classification of new data

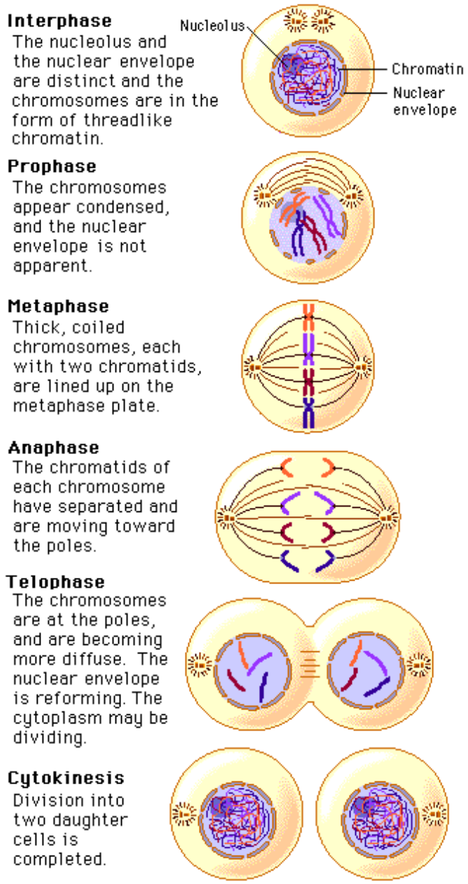


Images: C. Conrad & D.W. Gerlich, 2010, Journal of Cell Biology

Supervised learning & Object tracking

- Cell Cycle
  a priori knowledge
- Interphase
- Prophase
- Metaphase
- Anaphase
- Cytokinesis



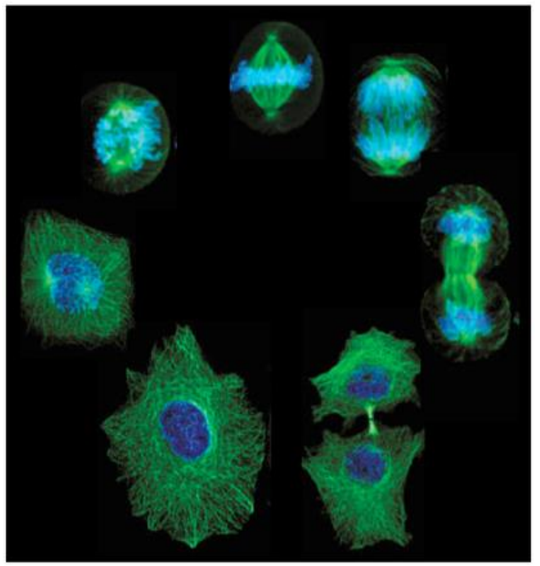


Images from: http://biomafia-biology.blogspot.com/2008/01/cell-growth-and-division-definitions.html

Supervised learning/Object tracking

- Biological a priori knowledge can help to improve:
- Object classification
- Tracking
- => Important tool for automated image/data analysis

	t=1	t=2	t=3	t=4	t=5	t=6	t=7
Track 1	Inter	Pro	Meta	Ana	Cyto	Inter	Pro
Track 2	Meta	Ana	Cyto	Inter	Pro	Meta	Ana
Track 3	Ana	Cyto	Inter	Pro	Meta	Ana	Cyto
Track 4	Cyto	Inter	Pro	Meta	Meta	Meta	Meta
Track 5	Inter	Pro	Meta	Cyto	Ana	Meta	Ana



Performance Supervised learning

Actual class \ Predicted class	Positive	Negative	
Positive	True Positive (TP)	False Negative (FN)	Sensitivity $\frac{TP}{TP+FN}$
Negative	False Positive (FP)	True Negative (TN)	Specificity $\frac{TN}{TN+FP}$
	Precision $\frac{TP}{TP+FP}$	Negative Predictive Value $\frac{TN}{TN+FN}$	Accuracy $\frac{TP+TN}{TP+FP+FN+TN}$

Performance Supervised learning

TP	50	50	50	50	50	50
TN	40	40	40	40	40	40
FP	5	4	3	2	1	0
FN	6	6	6	6	6	6
Precission	0.91	0.93	0.94	0.96	0.98	1
Neg. Pred.	0.87	0.87	0.87	0.87	0.87	0.87
Sensitivity	0.89	0.89	0.89	0.89	0.89	0.89
Specificity	0.89	0.91	0.93	0.95	0.98	1
Accuracy	0.89	0.9	0.91	0.92	0.93	0.94

Performance Supervised learning

Predicted class

Actual class	A	B	C
A	**16**	**0**	**0**
B	**0**	**17**	**1**
C	**0**	**0**	**11**

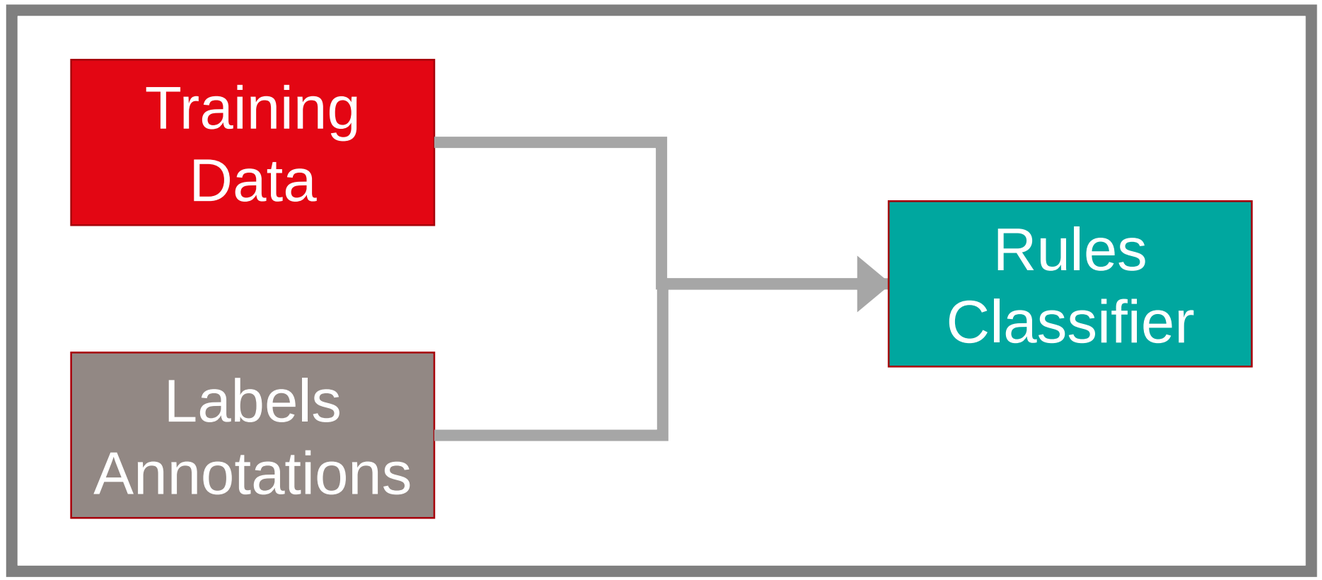
TP=16

FN=0+0=0

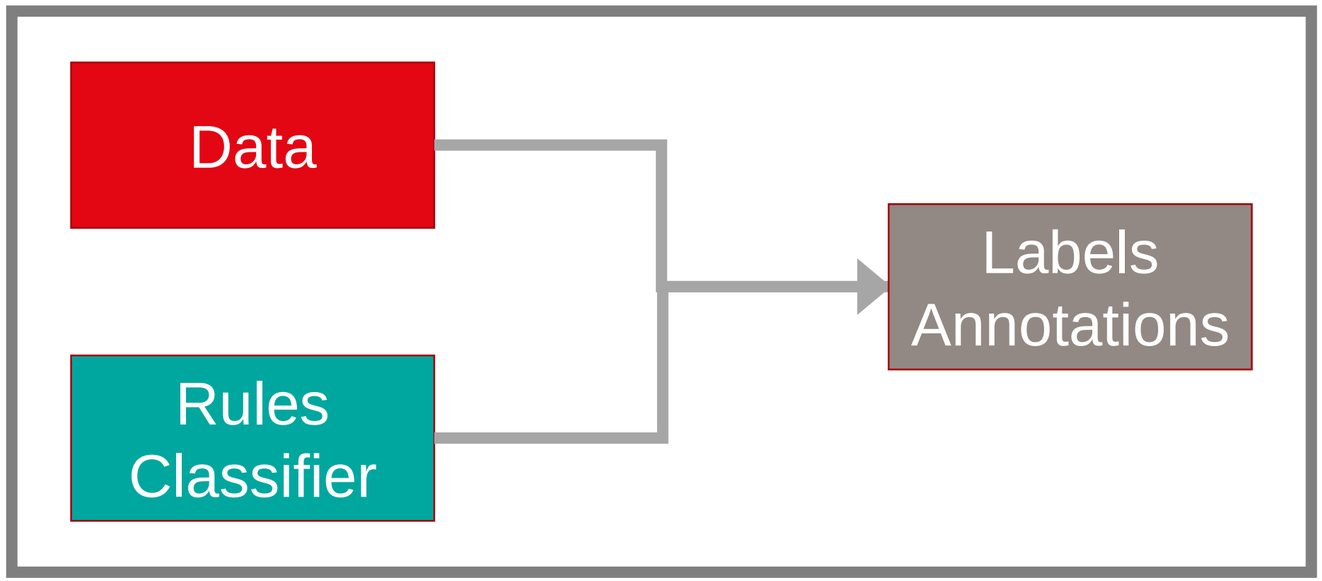
FP=0+0=0

TN=17+1+0+11=29

Summary



Training phase



Classification phase