คณะภูมิสารสนเทศศาสตร์
Faculty of Geoinformatics Burapha University

Introduction of Machine Learning & Deep Learning for Geoinformatics

Attawut Nardkulpat

Faculty of Geoinformatics

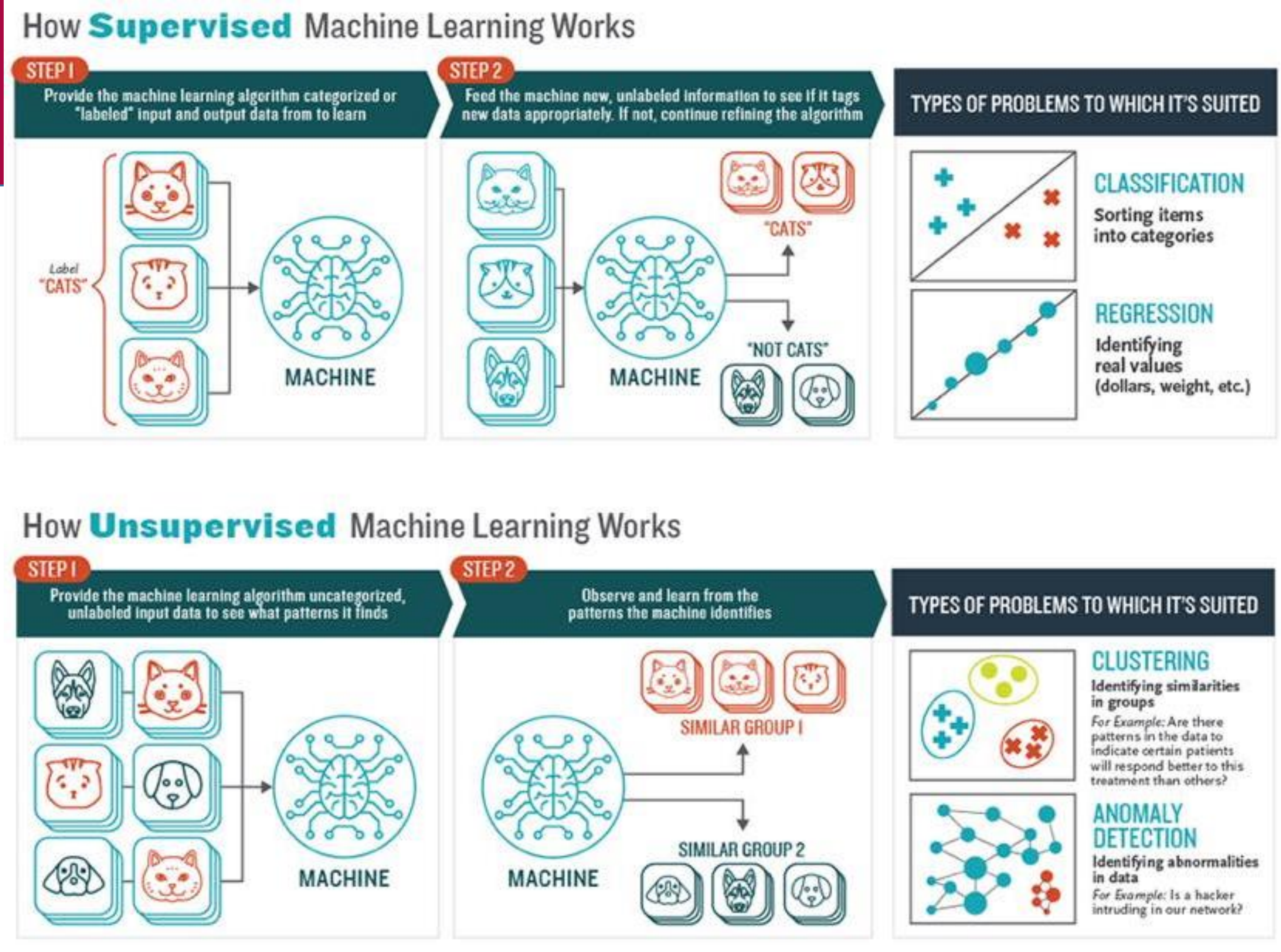
Burapha University

21/09/2020

attawut@buu.ac.th

Outline

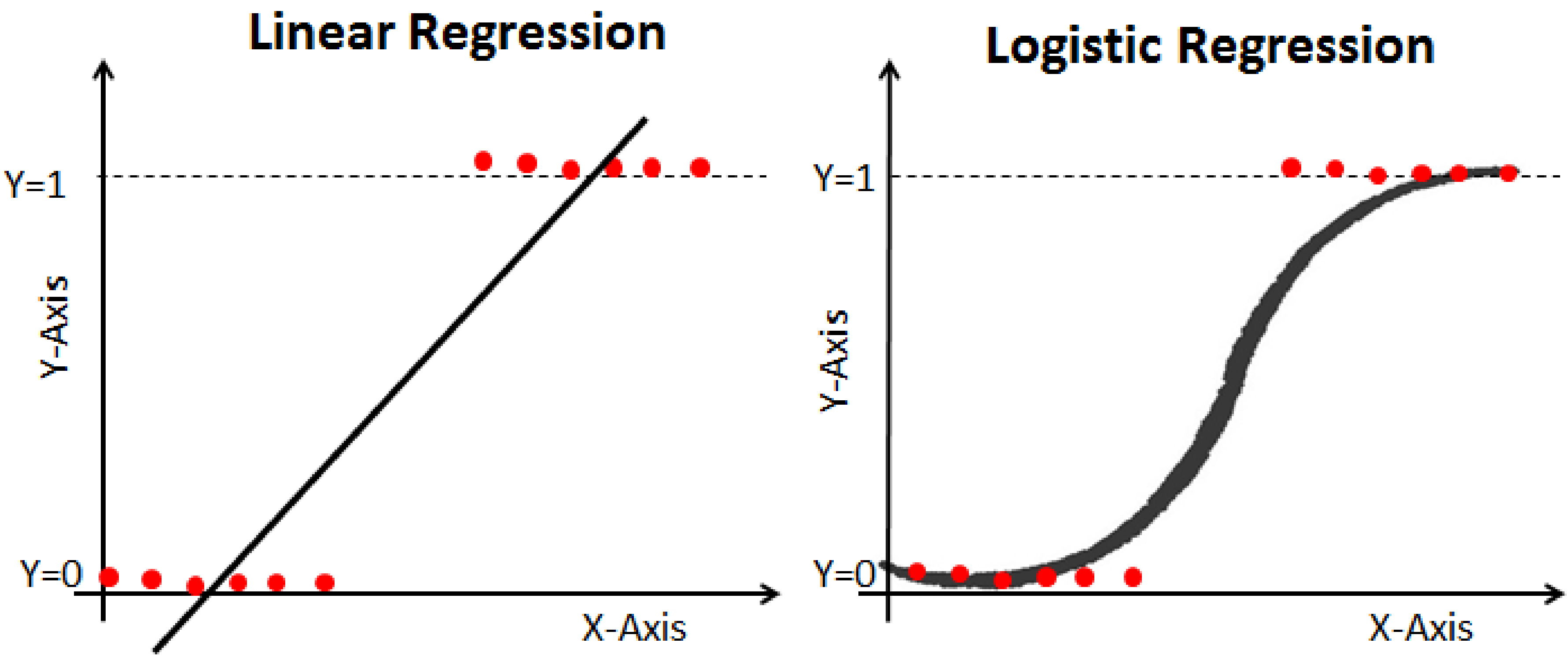
- **Week 1**
  - Overview AI, Machine Learning & Deep Learning
  - Regression
- **Week 2**
  - **Logistic Classification & Image Classification**
- **Week 3**
  - Ensemble Learning & Neural Network
- **Week 4**
  - Deep Learning



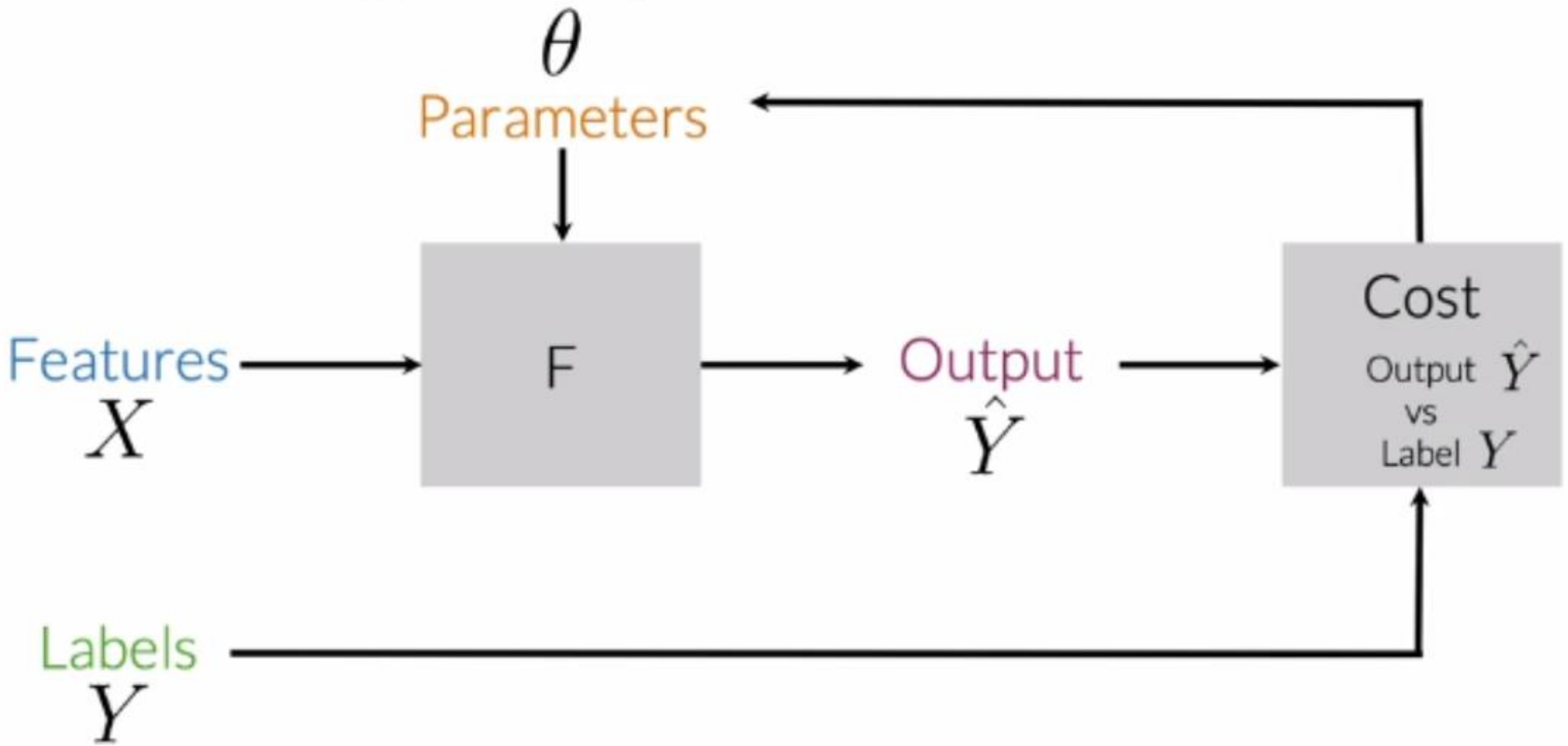
Credit : https://www.boozallen.com/s/insight/blog/how-do-machines-learn.html

Logistic regression

- จะใช้เมื่อคุณมีการจัดหมวดหมู่ปัญหา
- ตัวแปรเป้าหมายของคุณ (หรือตัวแปรที่คุณสนใจในการทำนาย) ประกอบด้วยหมวดหมู่
- หมวดหมู่เหล่านี้อาจเป็นใช่ / ไม่ใช่หรือบางอย่างเช่นตัวเลขระหว่าง 1 ถึง 10 แสดงถึงความพึงพอใจของลูกค้า
- Logistic Regression Model ใช้สมการเพื่อสร้างเส้นโค้งด้วยข้อมูลของคุณ จากนั้นใช้เส้นโค้งนี้เพื่อทำนายผลลัพธ์ของการสังเกตใหม่

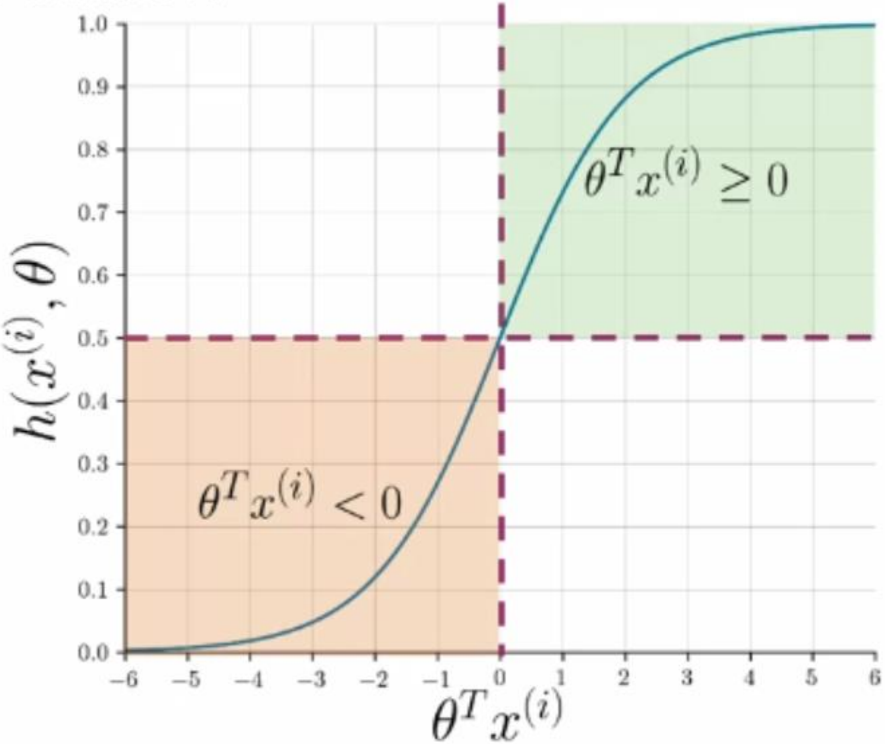

Linear Regression
Y=1
Y=0
Y-Axis
X-Axis
Logistic Regression
Y=1
Y=0
Y-Axis
X-Axis

Overview of logistic regression

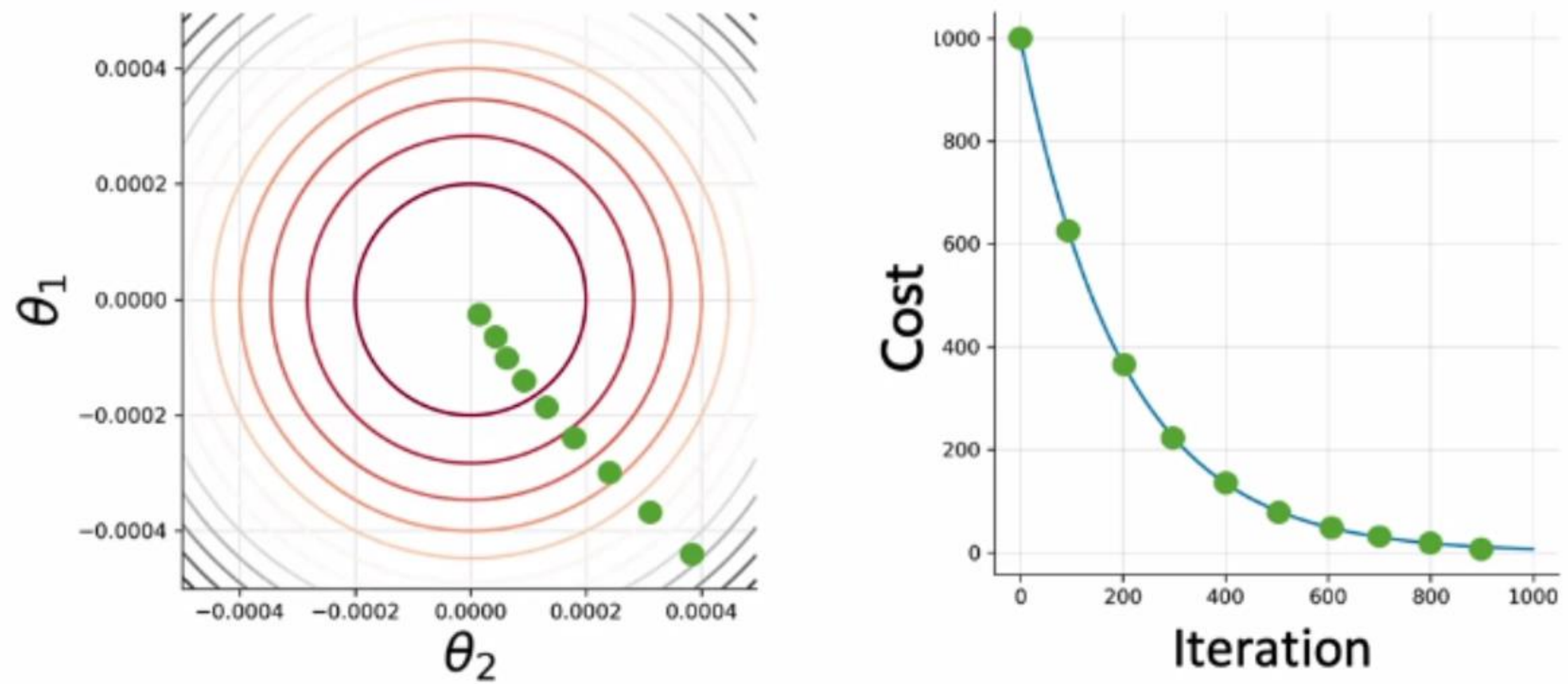


Overview of logistic regression

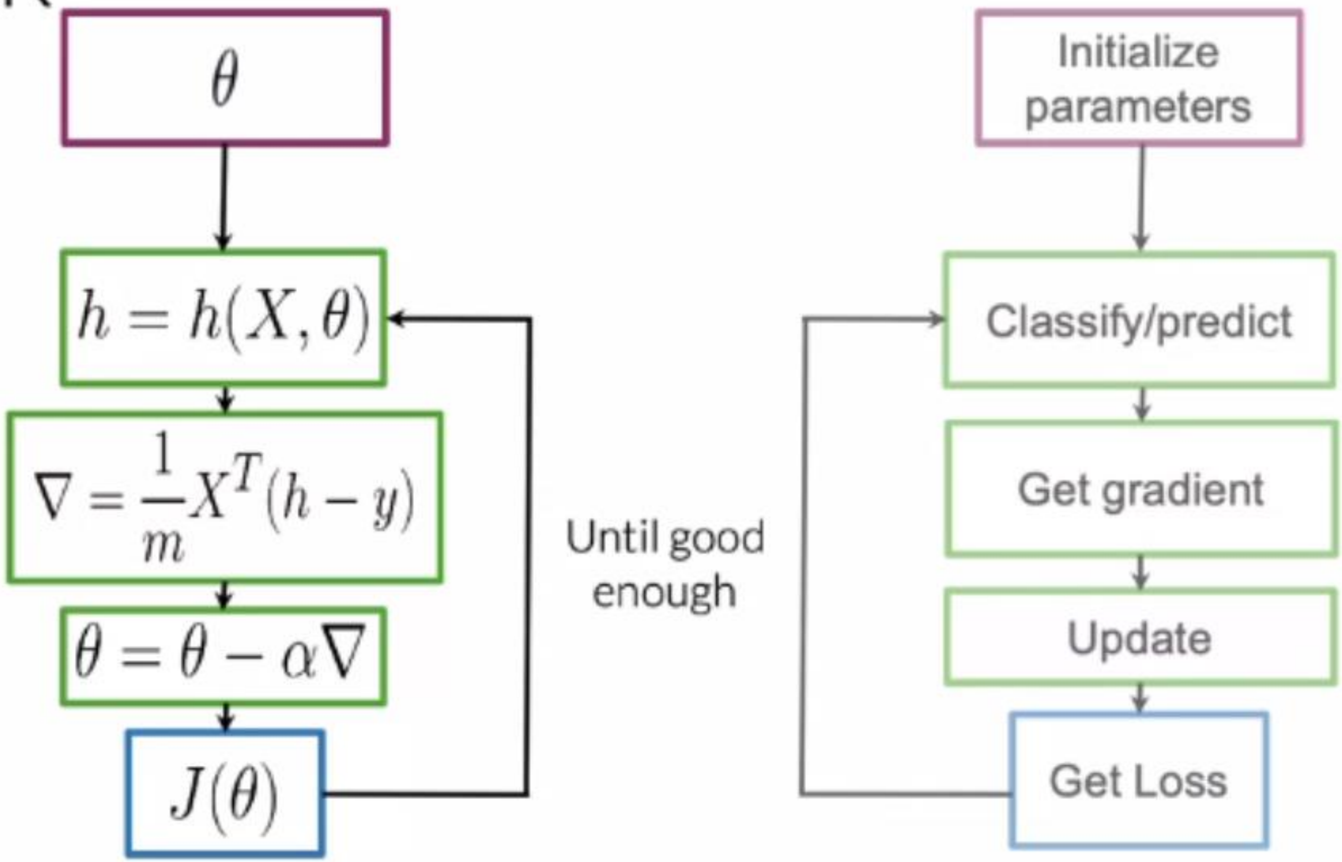
$$h(x^{(i)}, \theta) = \frac{1}{1 + e^{-\theta^T x^{(i)}}}$$



Training LR


0.0004
0.0002
0.0000
−0.0002
−0.0004
θ_1
−0.0004
−0.0002
0.0000
0.0002
0.0004
θ_2
1000
800
600
400
200
0
Cost
0
200
400
600
800
1000
Iteration

Training LR


θ
$h = h(X, \theta)$
$\nabla = \frac{1}{m} X^T (h - y)$
$\theta = \theta - \alpha \nabla$
$J(\theta)$
Until good enough
Initialize parameters
Classify/predict
Get gradient
Update
Get Loss

Testing logistic regression

$$Y_{val} = \begin{bmatrix} 0 \\ 1 \\ 1 \\ 0 \\ 1 \end{bmatrix} \quad pred = \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \\ 1 \end{bmatrix} \quad (Y_{val} == pred) = \begin{bmatrix} 1 \\ 1 \\ 0 \\ 1 \\ 1 \end{bmatrix}$$

Accuracy = 4/5 = 0.8

Naïve Bays

Likelihood Class Prior Probability

$$P(c \mid x) = \frac{P(x \mid c)\,P(c)}{P(x)}$$

Posterior Probability Predictor Prior Probability

$$P(c \mid \mathrm{X}) = P(x_1 \mid c) \times P(x_2 \mid c) \times \cdots \times P(x_n \mid c) \times P(c)$$

Naïve Bayes for Sentiment Analysis

Positive tweets

I am happy because I am learning NLP

I am happy, not sad.

Negative tweets

I am sad, I am not learning NLP

I am sad, not happy

word	Pos	Neg
I	3	3
am	3	3
happy	2	1
because	1	0
learning	1	1
NLP	1	1
sad	1	2
not	1	2
N_{class}	13	12

$P(w_i \mid \text{class})$

word	Pos	Neg
I	3	3
am	3	3
happy	2	1
because	1	0
learning	1	1
NLP	1	1
sad	1	2
not	1	2
Nclass	13	12

$$p(I|Pos) = \frac{3}{13}$$

word	Pos	Neg
I	0.24	

$P(w_i \mid \text{class})$

word	Pos	Neg
I	3	3
am	3	3
happy	2	1
because	1	0
learning	1	1
NLP	1	1
sad	1	2
not	1	2
Nclass	**13**	**12**

$$p(I|Neg) = \frac{3}{12}$$

word	Pos	Neg
I	0.24	0.25

$P(w_i \mid \text{class})$

word	Pos	Neg
I	3	3
am	3	3
happy	2	1
because	1	0
learning	1	1
NLP	1	1
sad	1	2
not	1	2
Nclass	**13**	**12**

word	Pos	Neg
I	0.24	0.25
am	0.24	0.25
happy	0.15	0.08
because	0.08	0.00
learning	0.08	0.08
NLP	0.08	0.08
sad	0.08	0.17
not	0.08	0.17
Sum	**1**	**1**

Naïve Bayes

Tweet: I am happy today; I am learning.

$$\prod_{i=1}^{m} \frac{P(w_i|pos)}{P(w_i|neg)} = \frac{0.14}{0.10} = 1.4 \quad > 1$$

$$\cancel{\frac{0.20}{0.20}} * \cancel{\frac{0.20}{0.20}} * \boxed{\frac{0.14}{0.10}} * \cancel{\frac{0.20}{0.20}} * \cancel{\frac{0.20}{0.20}} * \cancel{\frac{0.10}{0.10}}$$

word	Pos	Neg
I	0.20	0.20
am	0.20	0.20
happy	0.14	0.10
because	0.10	0.05
learning	0.10	0.10
NLP	0.10	0.10
sad	0.10	0.15
not	0.10	0.15

What means classification?

- Overall objective → (automatically) categorize all pixels in an image in certain (i.e. predefined) classes or themes
- Thematic classification allocates pixels to classes based on functions of the spectral (or backscatter) properties

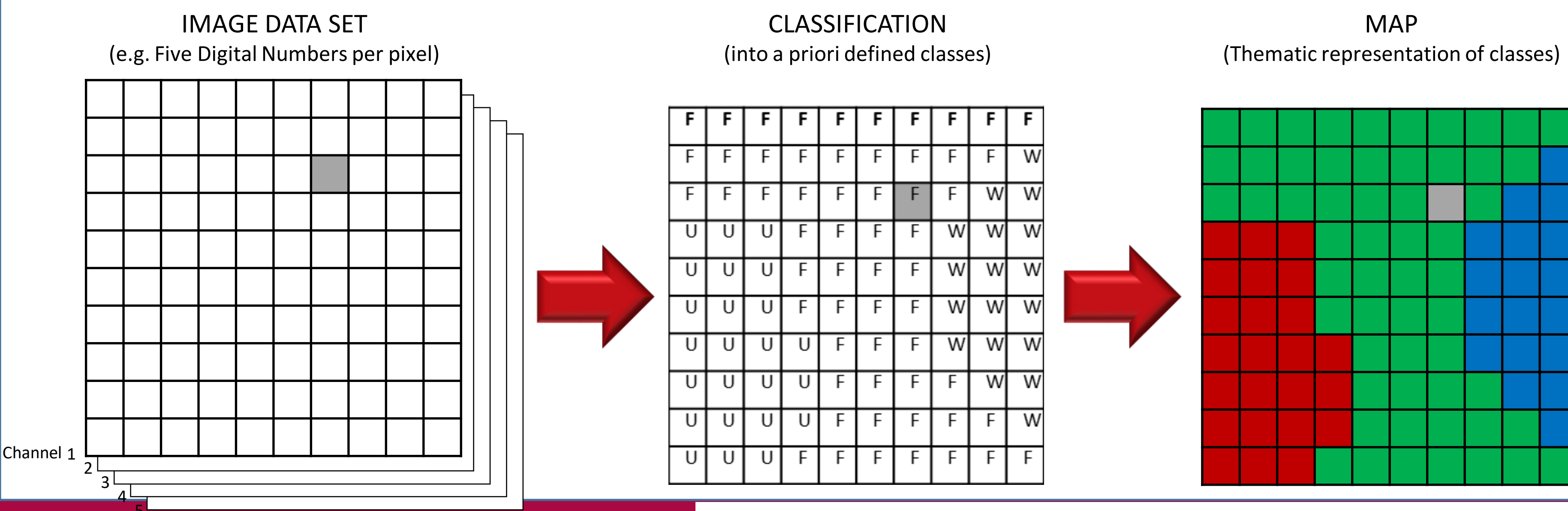
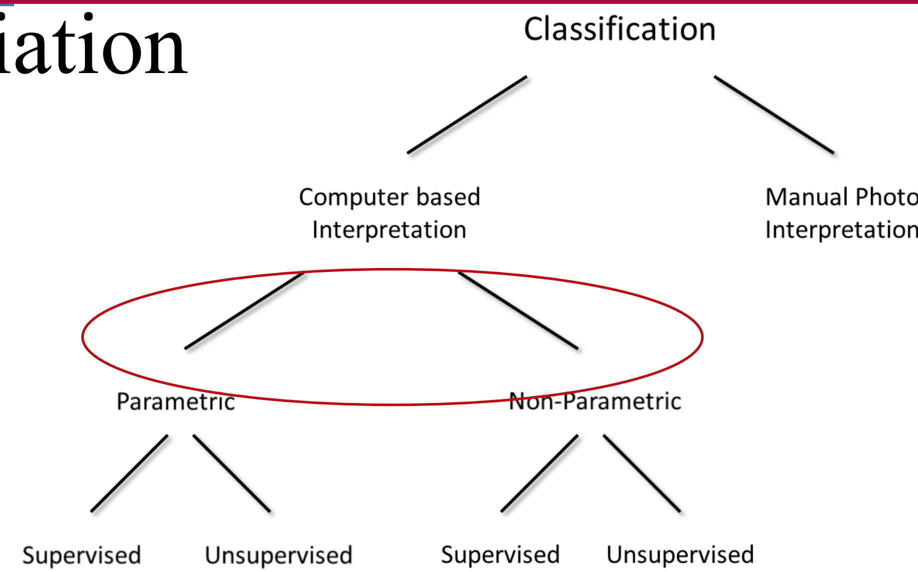


Fig. 1: Schematic Classification Workflow (after Lillesand et al. (2008))

Method overview

- Classification
  - Computer based Interpretation
    - Supervised
      - Parametric
      - Non-Parametric
    - Unsupervised
      - Supervised
      - Unsupervised
  - Manual Photo Interpretation

Algorithm based differentiation



Parametric Classifiers

- Implying a specific statistical distribution
- → Generally gaussian distribution
- Calculation statistical measurement

  (e.g. Standard deviation or Covariance

Non-Parametric Classifiers

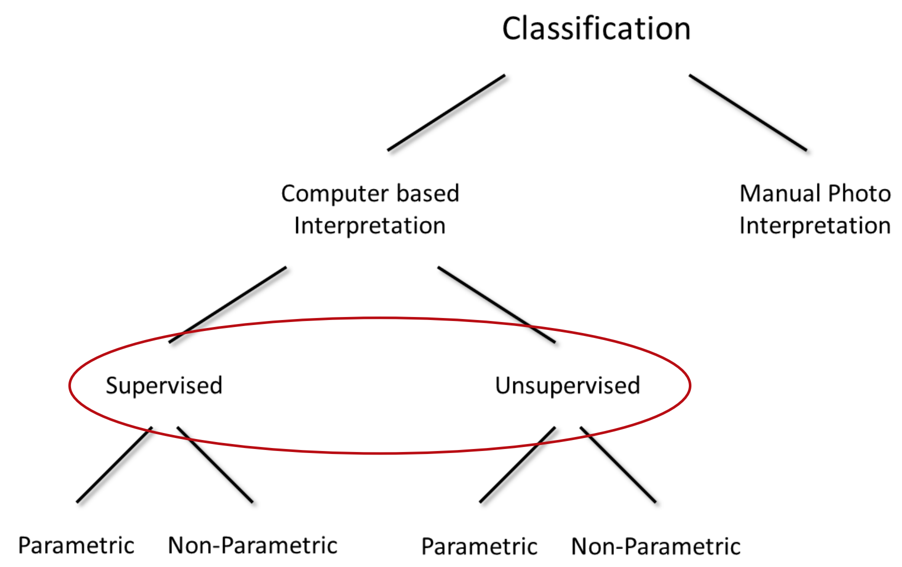
- No assumtion on the statistical distribution of the data
- Robust due to ability to describe numerous statistical distributions other than gaussian distribution

SAR dats is usually not gaussian distributed!

(see PPT ‚The histogram‘ in #1201)

➔ Non-parametric classifiers are more appropriate in Radar remote sensing

Differentiation by training concept



Unsupervised Classifiers

- No Training stage
- Purely based on the statistical distribution of the input data

Supervised classifiers

- Employing manual traing of the data set to distinguish the desired classes

Differentiation by training concept

Unsupervised vs. Supervised

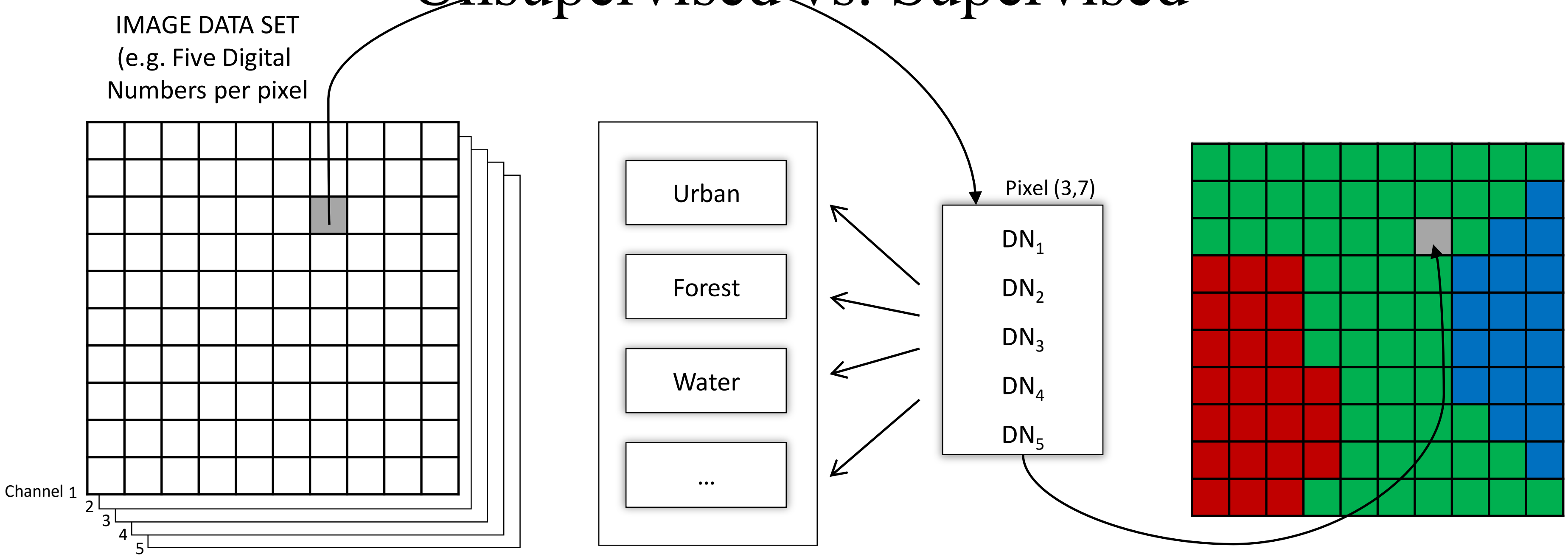


Fig. 2: Basic steps of classification
(after Lillesand et al. (2008))

Minimum-Distance-to-Mean

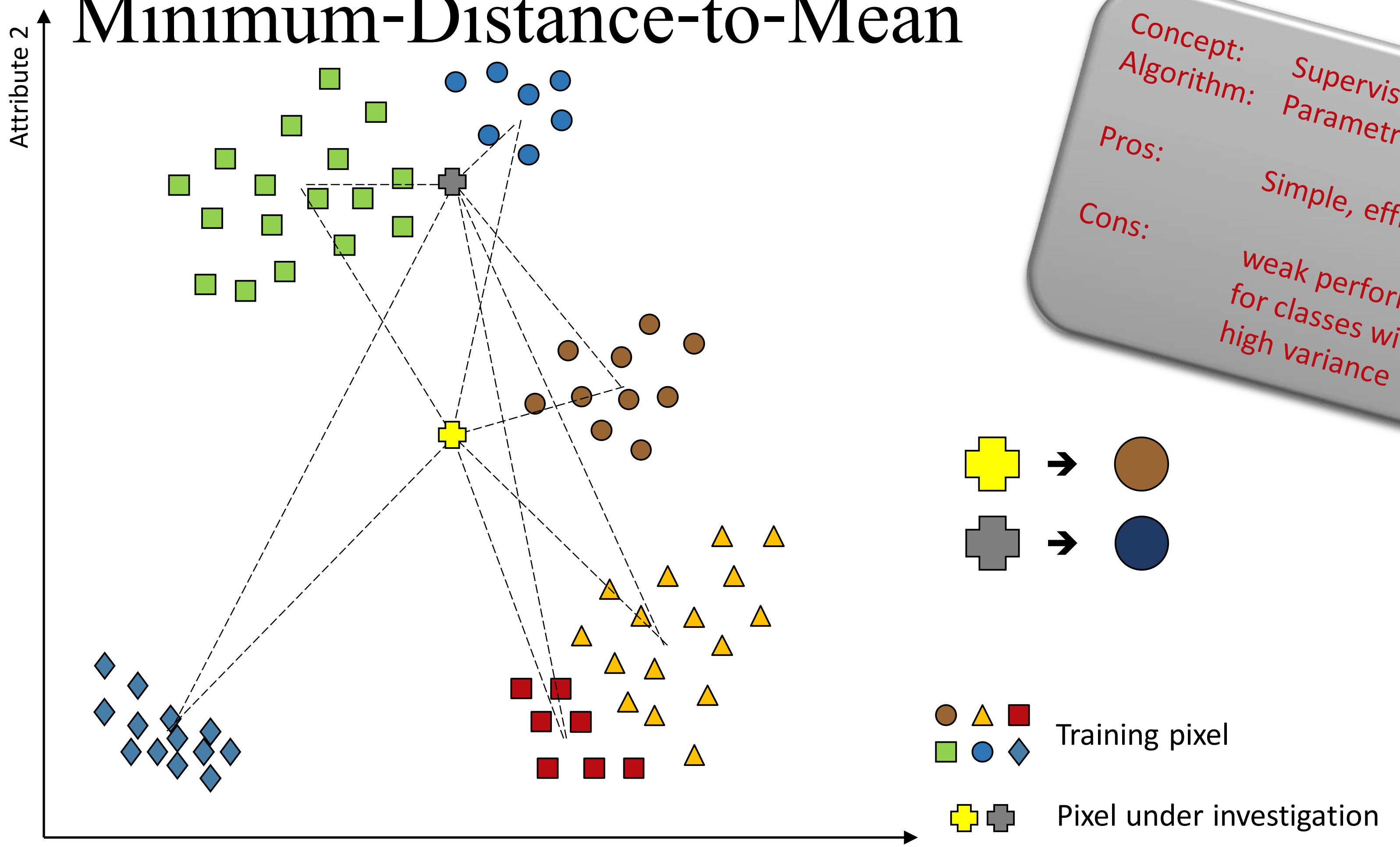


Fig. 3: Minimum-Distance-to-Mean Classifier

Concept: Supervised
Algorithm: Parametric
Pros: Simple, efficient
Cons: weak performance for classes with high variance

K-Means Clustering

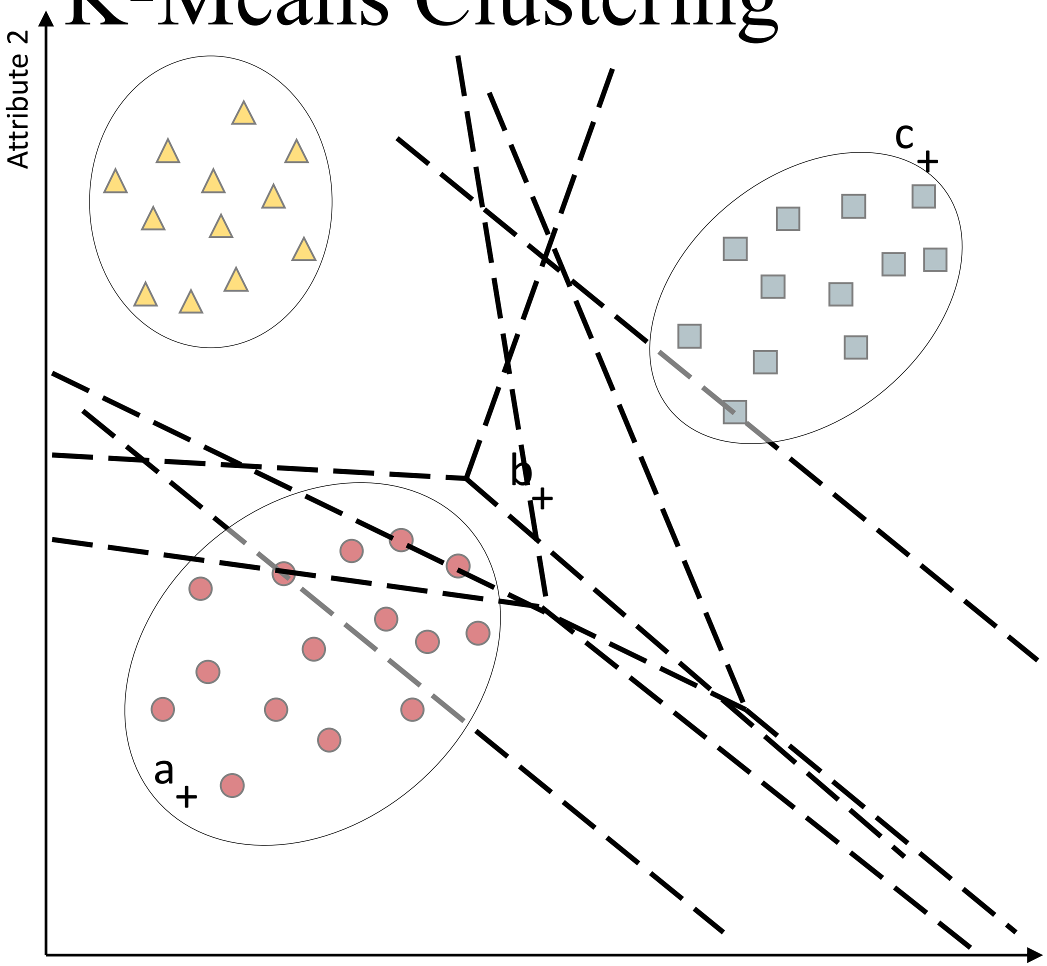


Fig. 4: K-Means Classifier

Concept: Unsupervised
Algorithm: Parametric
Pros: No interaction or a priori tuning necessary
Cons: Result depends on initial cluster centers
Empty classes possible

Maximum Likelihood

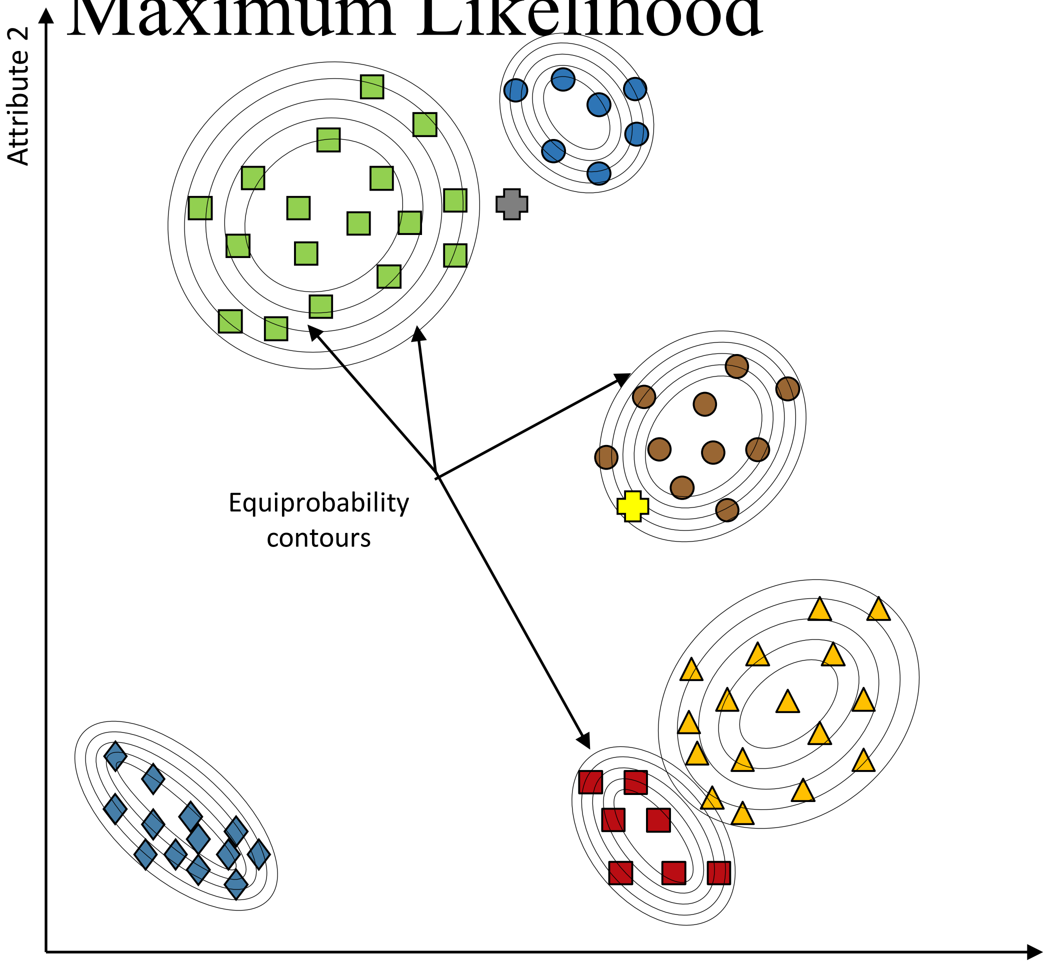
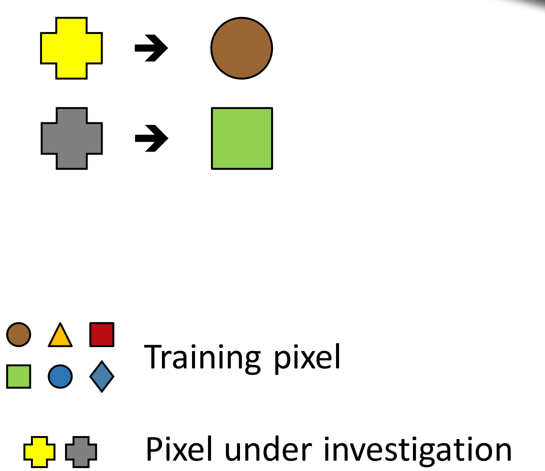


Fig. 5: Maximum Likelihood Classifier

Concept: Supervised
Algorithm: Parametric
Pros: based on multiple statisitcal parameters → robust
Cons: Complex Processing



(K-)Nearest Neighbor

Concept: Supervised
Algorithm: Non-Parametric

Pros: Simple implementation

Cons: Slow for many image bands
strong influence of „k“ on result

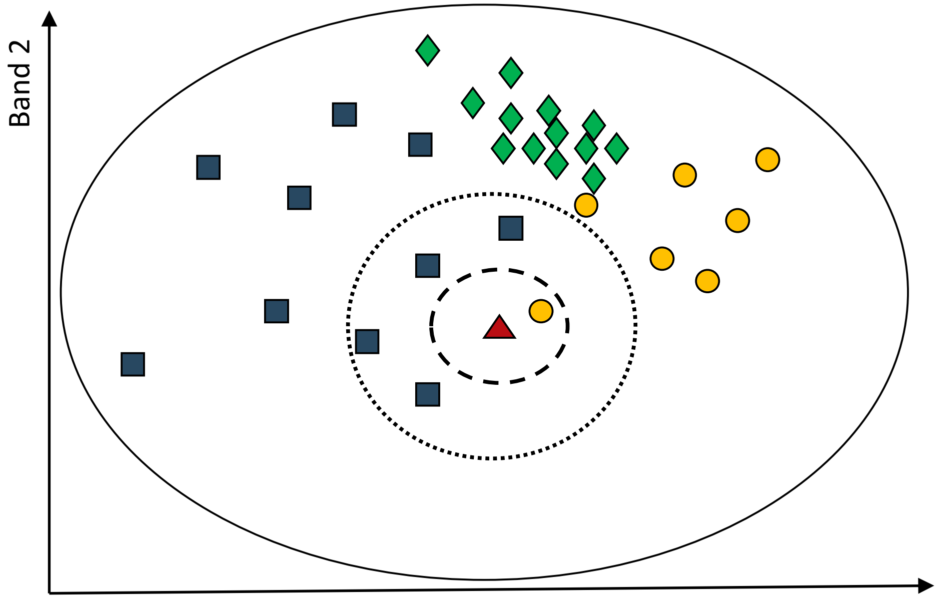


Fig. 6: K-Means Classifier

- - - - *k = 1; 1x* ● → ●

········ *k = 5; 4x* ■, *1x* ● → ■

——— *k = 30; 7x* ●, *10x* ■ ,*13x* ◆ → ◆

■ Class A
● Class B
◆ Class C
▲ Data point under investigation

Decision Trees

Concept: Supervised
Algorithm: Non-Parametric

Pros:
- Simple Structure
- Fast
- Combination of data possible

Cons:
- Complex design phase

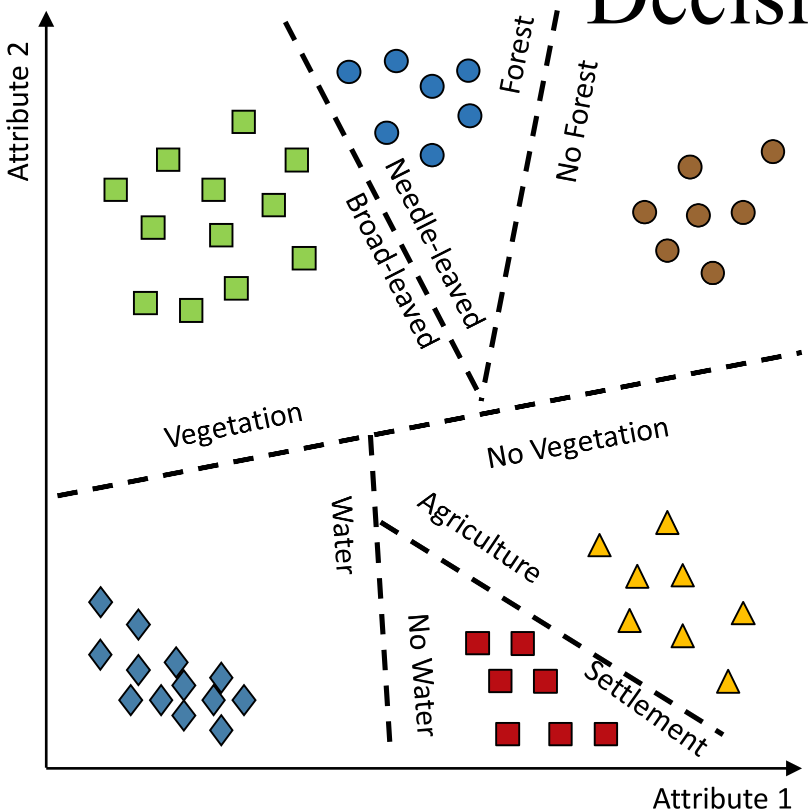
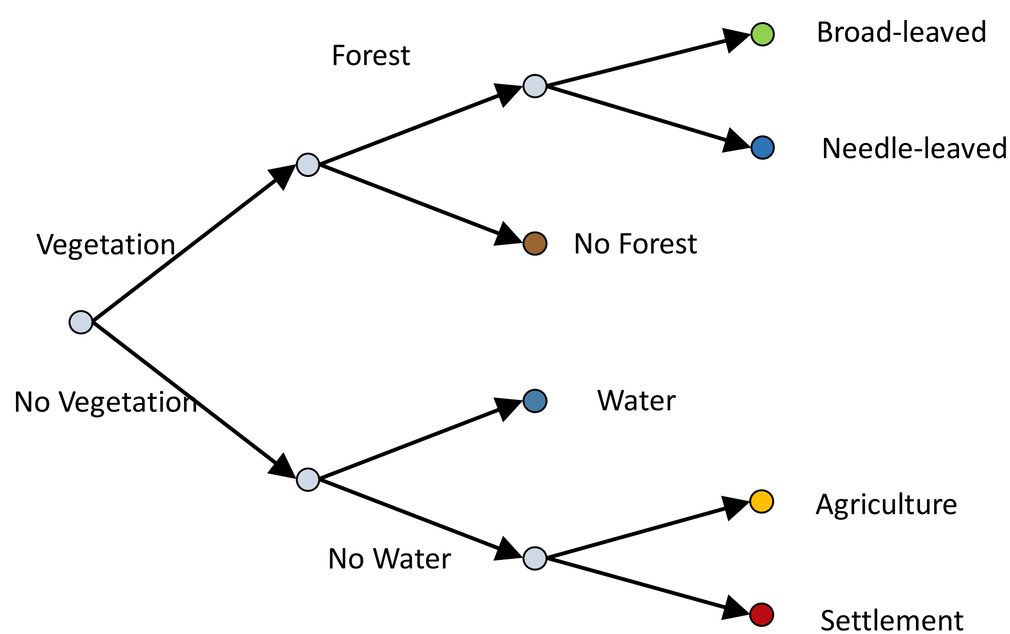


Fig. 7: Decision Tree Classifier



Support Vector Machines (SVM)

Concept: Supervised
Algorithm: Non-Parametric
Pros:
- Useful for high dimensional data
- White Box Approach
- very flexible

Cons:
- long processing time
- needs multiple iterations
- can stay unresolved

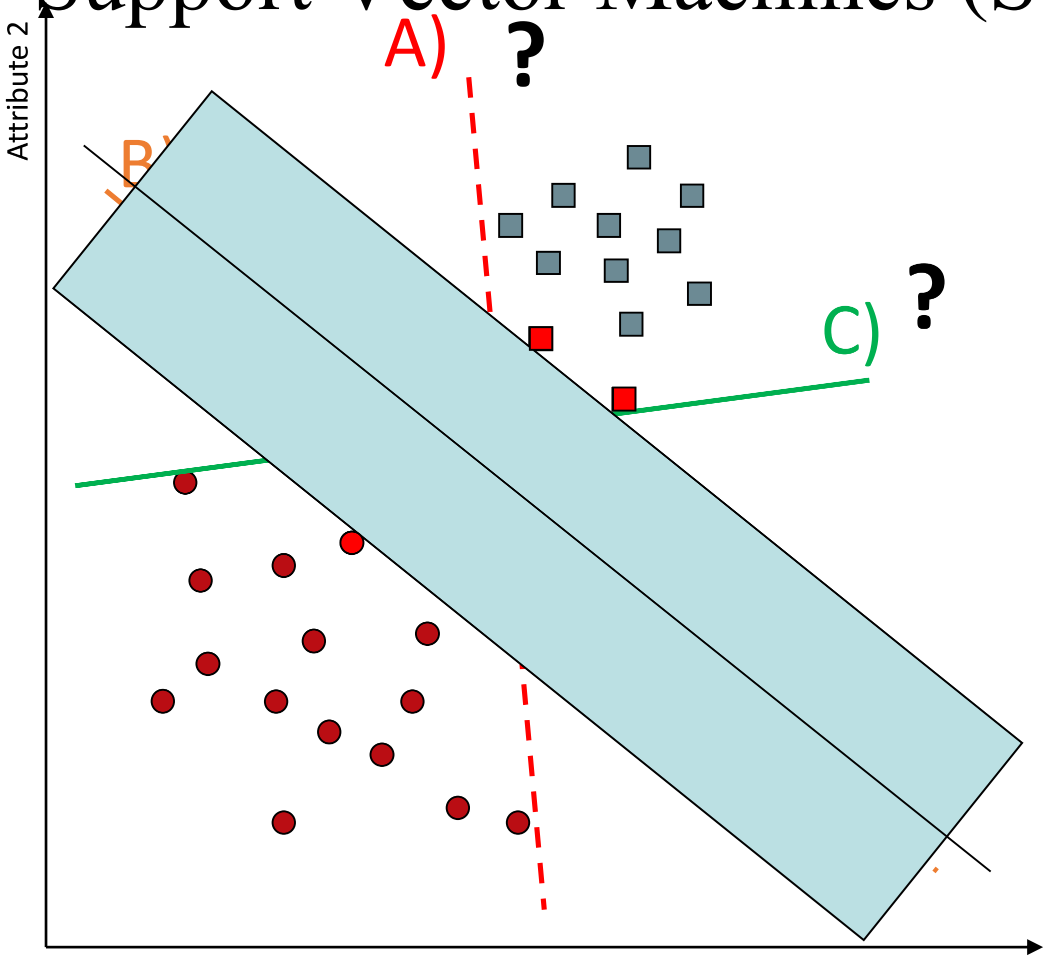


Fig. 9: SVM Classifier

Which line divides the classes best?

- Support Vectors
- Training Pixel Class **A**
- Training Pixel Class **B**