



Week 2

Machine Learning and  
Big Data - DATA622

Fall 2023

---

CUNY School of Professional Studies

# Week 1

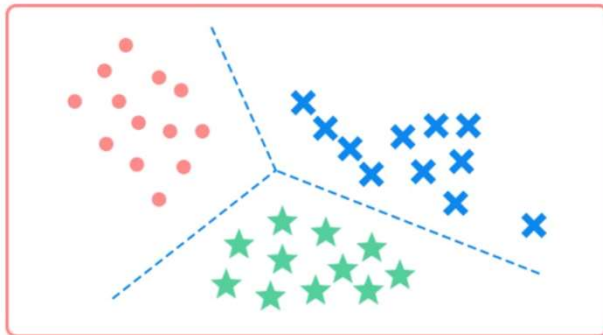
---

1. Discussion Board Week 1 (introduce yourselves) – Due Sunday EOD
2. Reading: Practical Machine Learning, chapters 1 & 2  
(if you prefer Python - let me know)
3. Concepts:
  - a. Supervised and unsupervised learning; machine learning vs. statistical learning  
Watch: <https://youtu.be/5N9V07Elflg> (18 minutes)  
Watch: <https://www.youtube.com/watch?v=Z0v9QMkA3dA> (12 minutes)
  - b. A quick intro to reinforcement learning (20 minutes)  
Read: <https://deepsense.ai/what-is-reinforcement-learning-deepsense-complete-guide>

# Supervised vs Unsupervised

---

Classification

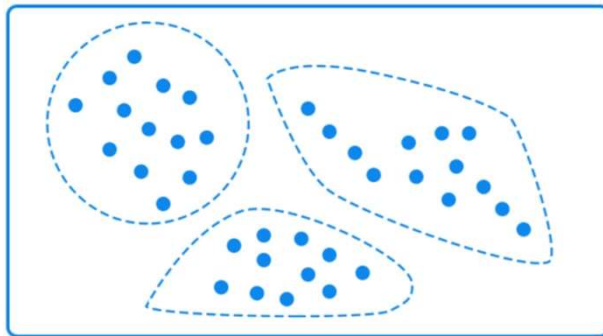


Supervised learning

- **Supervised Learning**

- The computer is presented with **both**
  - inputs (“features” i.e. independent variables) **and**
  - outputs (“labels”) i.e. associated sample outputs, indicating the class of the observation (dependent variable)
- The computer attempts to learn the model that maps inputs (features) to each class (labels)
- It predicts outputs: new data inputs are classified based on the model learned

Clustering



Unsupervised learning

- **Unsupervised Learning**

- The computer is given only inputs (“features” i.e. independent variables)
- The computer attempts to cluster data based on similarity/dissimilarity of inputs
- It cannot predict outputs – as it was never given sample outputs

# Machine Learning vs Statistical Learning

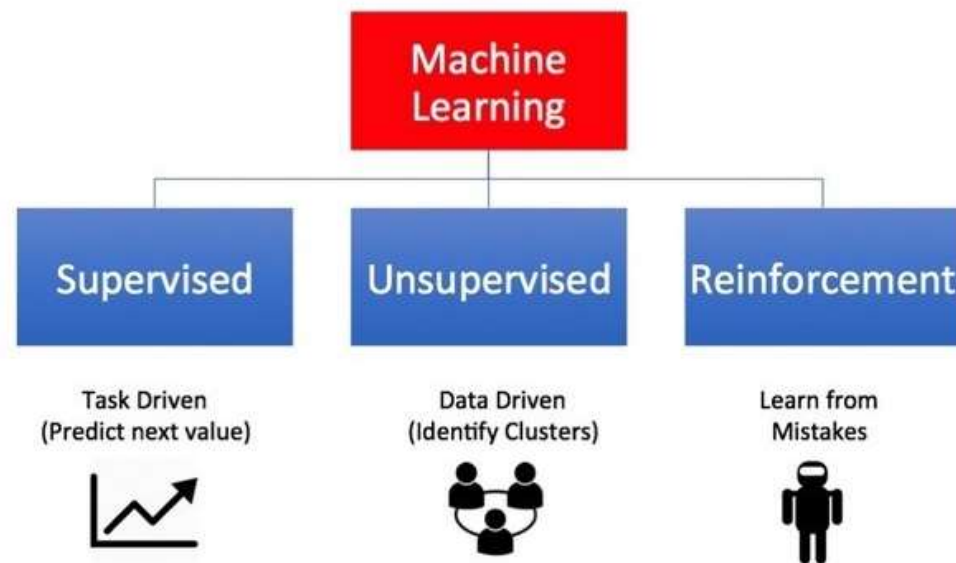
---

MACHINE LEARNING	STATISTICAL LEARNING
Subfield of Artificial Intelligence	Subfield of mathematics
Uses algorithms	Uses equations
Requires minimum human effort; is automated	Requires a lot of human effort
Can learn from large data sets	Deals with smaller data sets
Has strong predictive abilities	Gives a best estimate: you gain some insights into one thing, but it's of little or no help with predictions
Makes predictions	Makes inferences
Learns from data and discovers patterns	Learns from samples, populations, and hypotheses

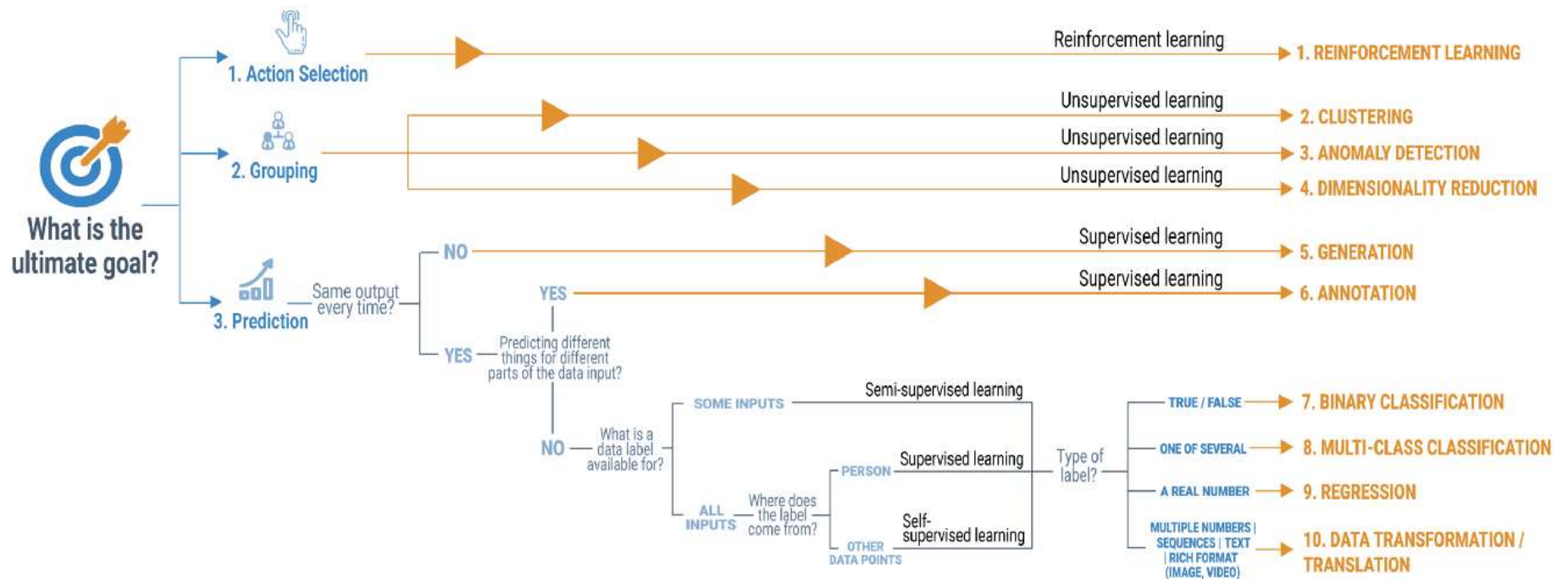
Source: *An introduction to Statistical Learning*

# Types of Machine Learning

---



# Types of Machine Learning



## Week 2

---

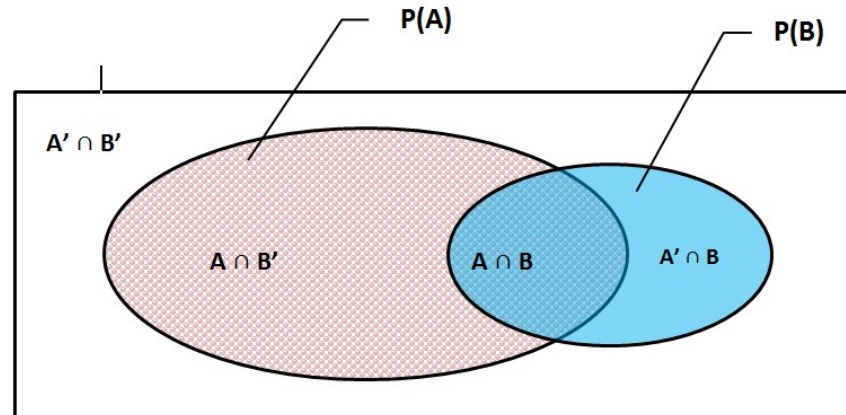
1. Discussion Board Week 2
2. Reading (Practical Machine Learning with R):
  - a) Linear Regression: chapter 4
  - b) Logistic regression: chapter 5
3. Concepts:
  - a) Helpful lecture slides (these slides go into the topics we will be covering next week. Up to slide #28 covers this week's topic):  
<https://web.stanford.edu/~hastie/MOOC-Slides/classification.pdf>
  - b) Multinomial logistic regression:  
Please see: <https://stats.idre.ucla.edu/r/dae/multinomial-logistic-regression/>
4. Helpful videos:
  - a) Statistical learning and models (the whole playlist):  
<https://www.youtube.com/watch?v=p9n2w236B48&list=PL5-da3qGB5IDvuFPNoSqheihPOQNJpzyy&index=1>

# Bayes Theorem

---

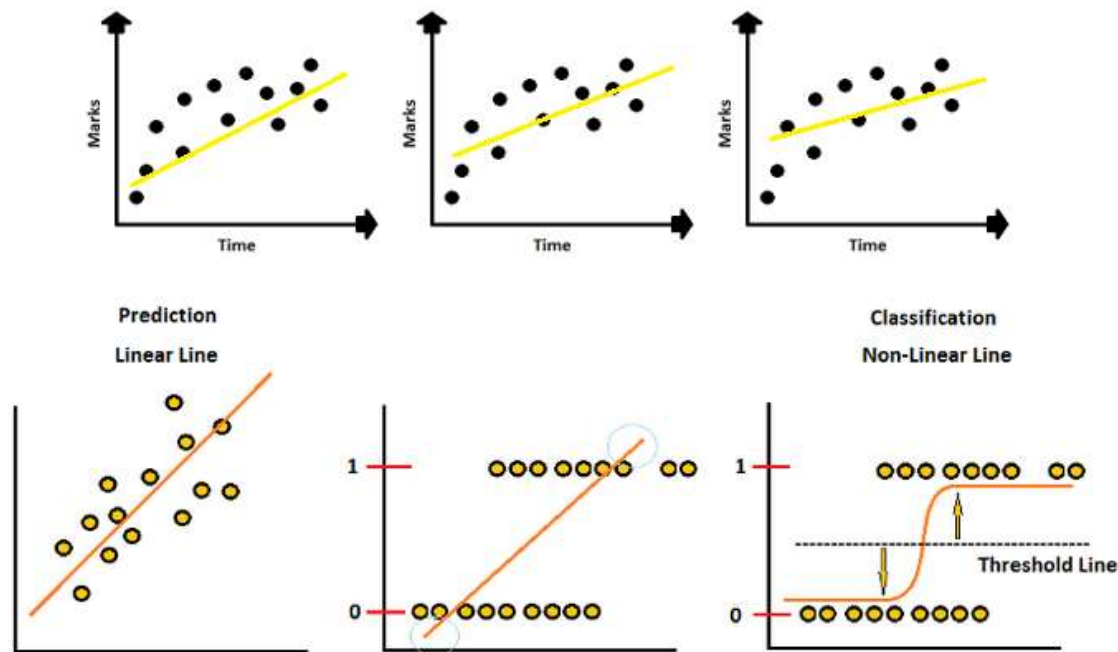
$$P(A|B) = \frac{P(B|A) P(A)}{P(B)}$$

$P(A)$  = The probability of A occurring       $P(B)$  = The probability of B occurring  
 $P(A|B)$  = The probability of A given B       $P(B|A)$  = The probability of B given A  
 $P(A \cap B)$  = The probability of both A and B occurring





# Linear vs Logistics Regression



# Visualizing data

