01 December 2022

UNIVERSITY OF GLOUCESTERSHIRE

CT4031
Maths for Data Science

Week 11 – Introduction to Data Mining

AI – Machine learning

- What is this?
- How can this help us?
- What are the differences?

Data mining

- Data mining is also called knowledge discovery and data mining (KDD)
- Data mining is
  - extraction of useful patterns from data sources, e.g., databases, texts, web, image.
- Patterns must be:
  - valid, novel, potentially useful, understandable

Task

- discuss the following:
  - Application
  - Techniques/Algorithm being used
  - Efficacy
  - Application and benefits
- Try to focus on the following areas:
  - Health technologies
  - Malware analysis
  - Cyber analysis
  - Digital forensics
  - Robotics and computer vision

AI – Machine learning

- Supervised Vs Unsupervised learning

AI – Machine learning

- Supervised Learning
  - produce a data output from the previous experience
  - helps you to optimize performance criteria using experience
  - forecast/prediction models
    - Regression / Classification
- Unsupervised Learning
  - allow the model to work on its own to discover information
  - deals with the unlabeled data
  - can be more unpredictable
  - unknown patterns
    - Clustering / Association rules

Machine learning

- Feature selection
- Model selection
- Model construction
- Training
- Prediction

Machine learning

- **Feature selection** 
- Model selection
- Model construction
- Training
- Prediction

Example of Features - malware detection

- ???

Example of Features - malware detection

- Static ??
- Behaviour??

Example of Features - malware detection

- Static
  - Size
  - Signature
- Behaviour
  - Code injection
  - Network traffic
  - Process creating run entries
  - Process executed from a particular area (%temp%)

Example of Features

- Task
  - Choose an application area for DM and provide example of features that could be used for AI in this area.

Machine learning

- **Feature selection**
- **Model selection** 
- Model construction
- Training
- Prediction

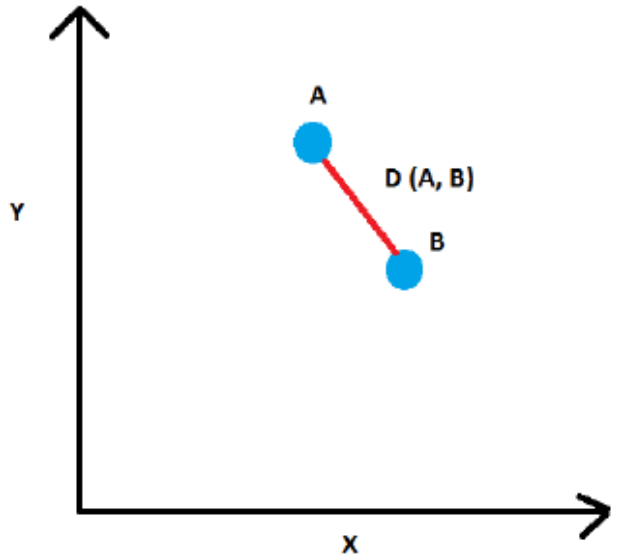
Model selection

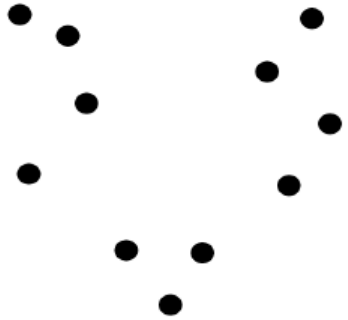
- KNN
- K-Means
- SVM
- Decision Tree
- Random Forest
- Neural Networks
  - Supervised / Unsupervised
- Genetic algorithms
- Etc…

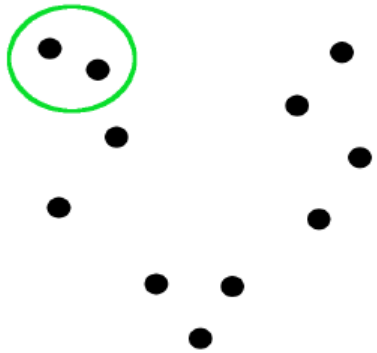
Task – Research and discuss

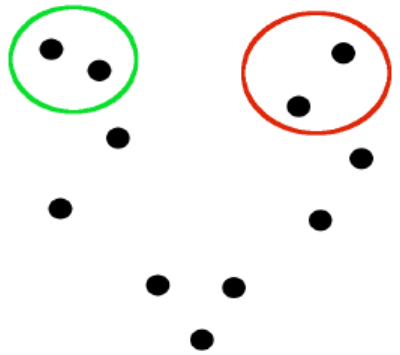
- KNN
- K-Means
- SVM
- Decision Tree
- Random Forest
- Neural Networks
  - Supervised / Unsupervised
- Genetic algorithms
- Etc…

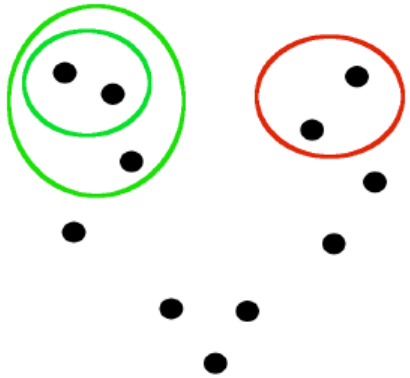
Euclidean distance

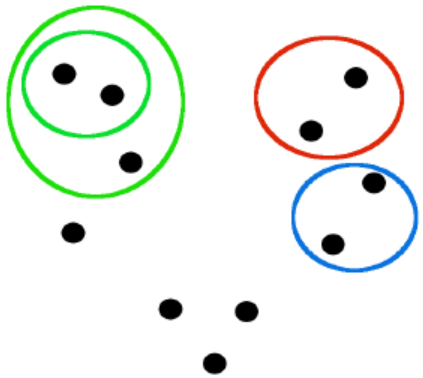


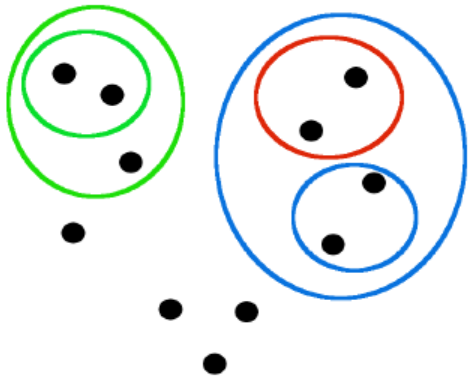


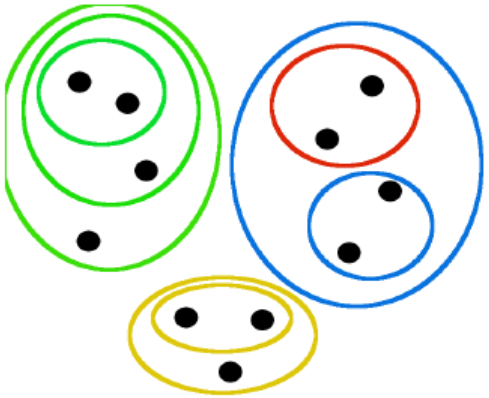




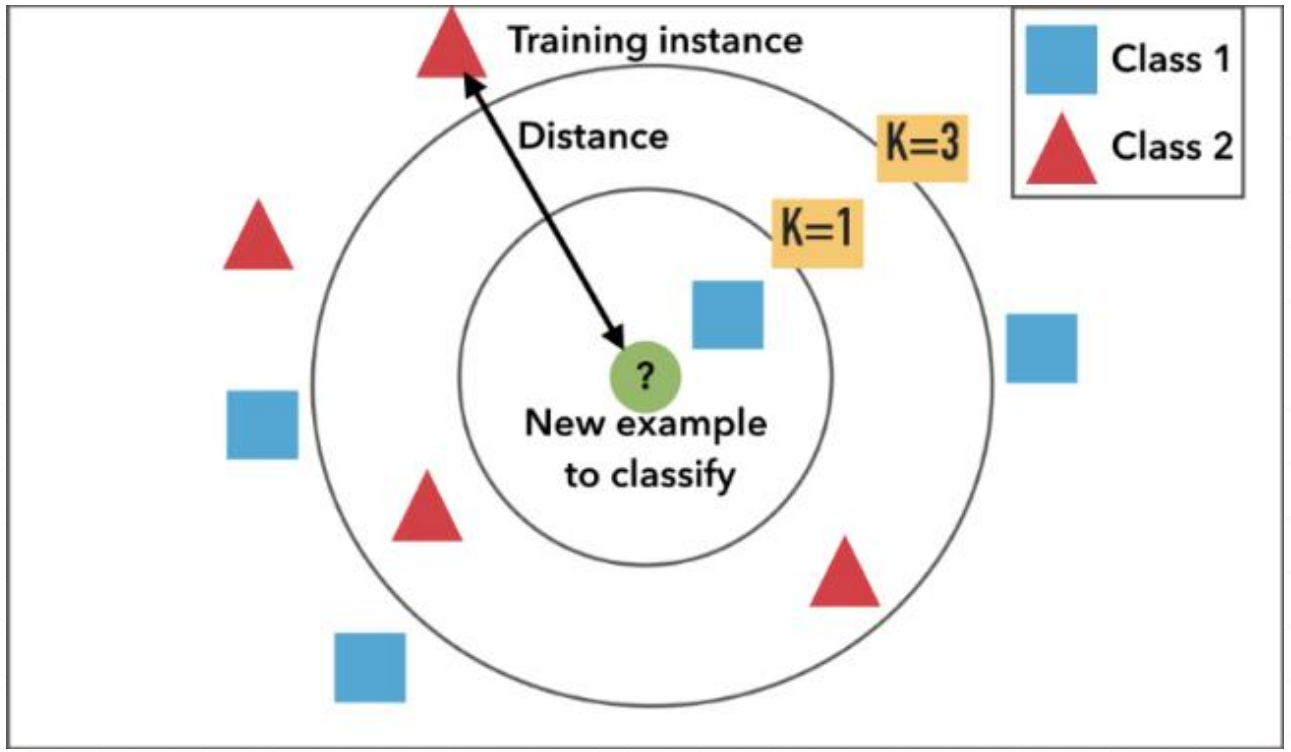








KNN Classification



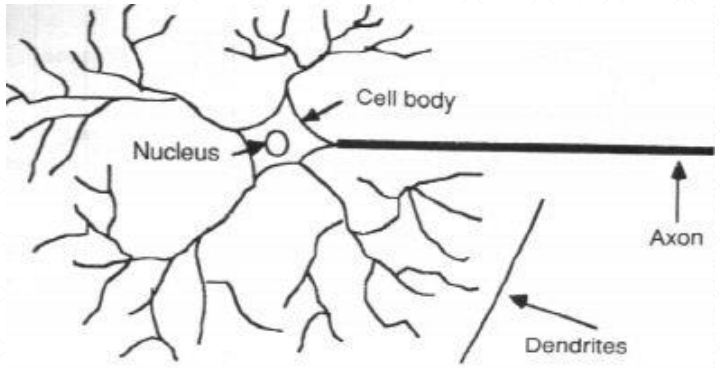
Different types of AI

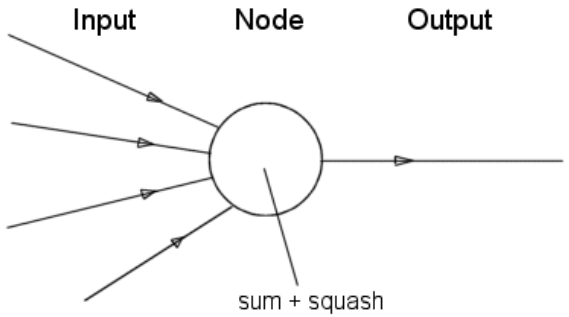
- Symbolic AI Vs.
- Non-Symbolic AI
- Neural Networks
- Deep Learning
- Multimodal Abstraction

Task - Research

- Symbolic AI Vs.
- Non-Symbolic AI
- Neural Networks
- Deep Learning
- Multimodal Abstraction

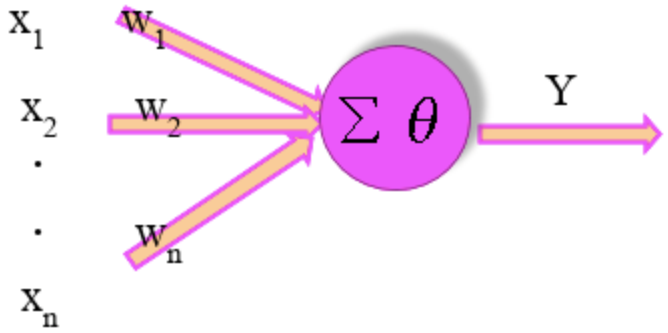
Neural networks



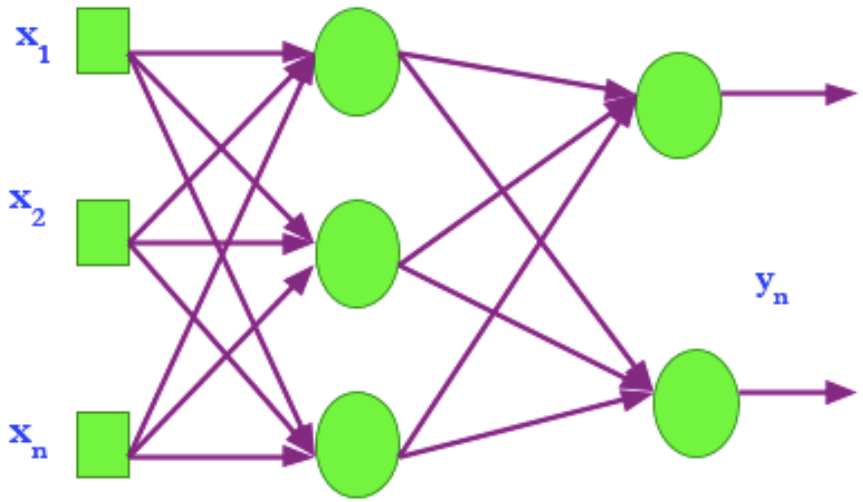


McCulloch-Pitts' Artificial Neuron

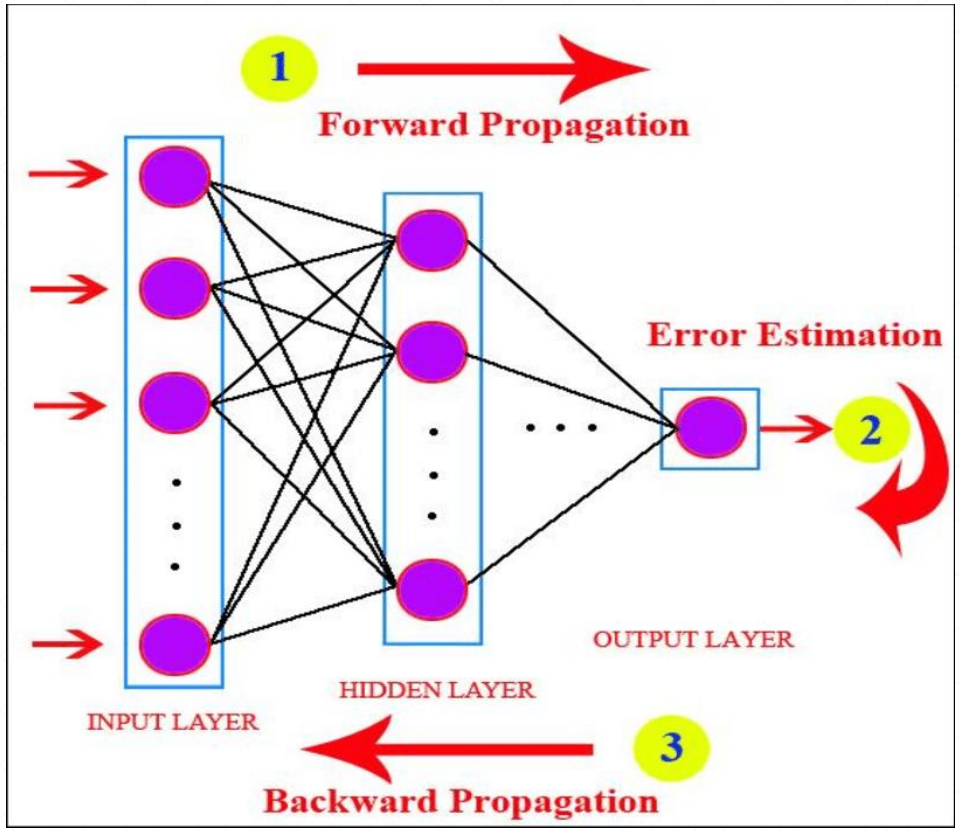
$$\sum_{i=1}^{n} x_i \, w_i \geq \theta$$



Neural network



Feed forward and error propagation



Machine learning

- **Feature selection**
- **Model selection**
- **Model construction** 
- Training
- Prediction

Machine learning

- **Feature selection**
- **Model selection**
- **Model construction**
- **Training** 
- Prediction

Machine learning

- Feature selection
- Model selection
- Model construction
- Training
- Prediction 

Assignment

How is the assignment going so far?

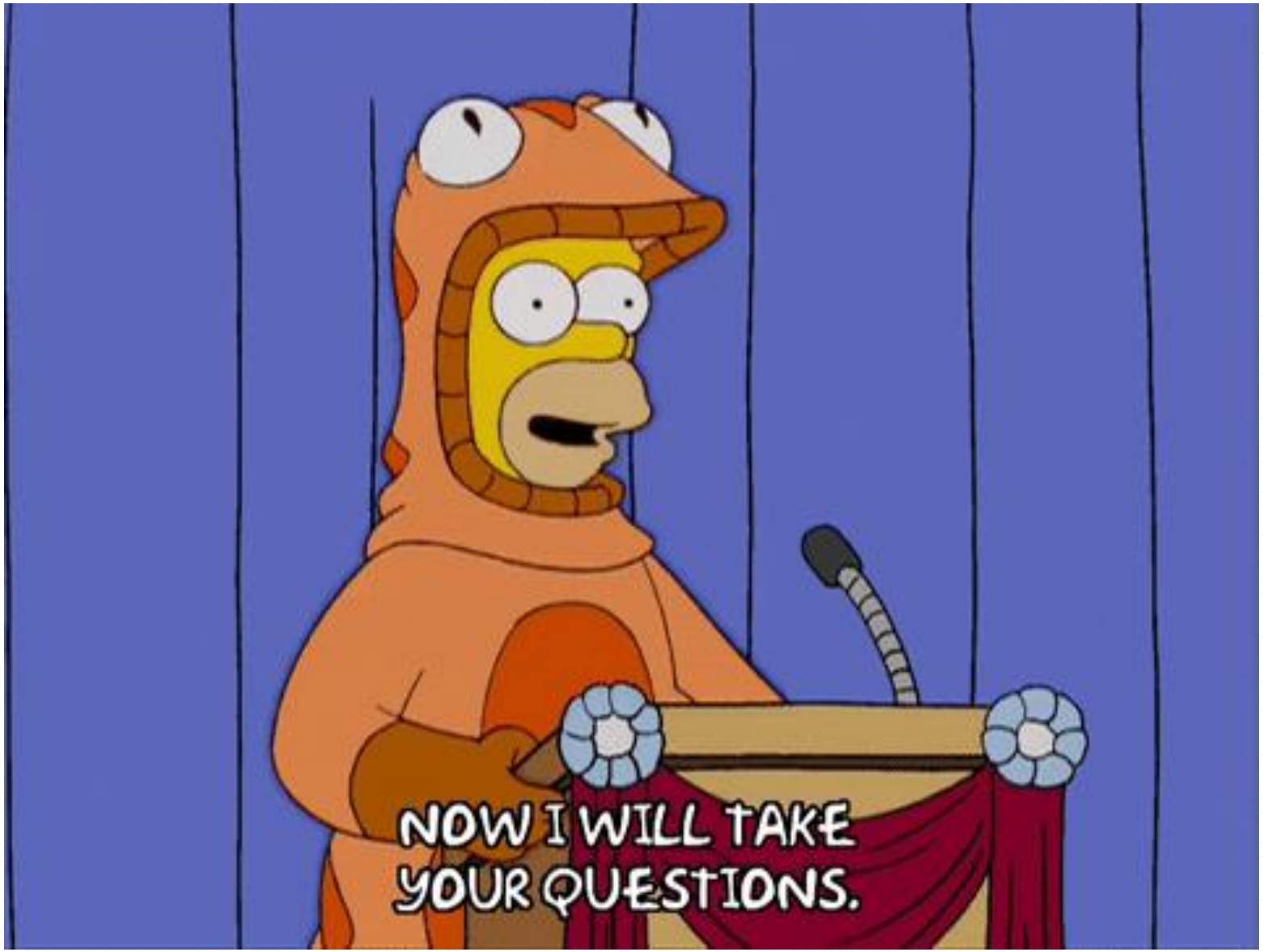
Questions?

Problems?

Challenges?

Next Session !

- Assignment workshop


NOW I WILL TAKE
YOUR QUESTIONS.