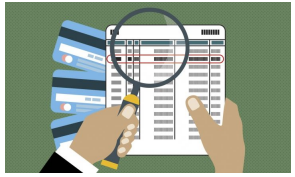
EXPLORE | DIGITAL SKILLS

Advanced Classification

Overview

What is Machine Learning?

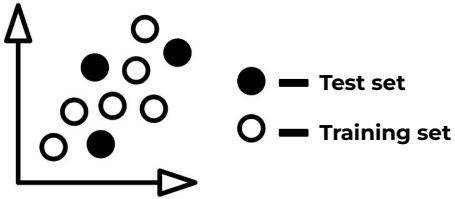
Machine learning is the study of **software that automatically learns** from experience.

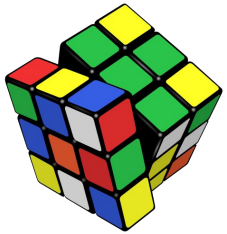
Types of machine learning		Where is it used?	
1 Supervised	Given sets of input-output pairs (x's and y's), the algorithm finds **hidden relationships** in the data (function approximation)	Prediction / Classification Credit & Insurance underwriting Cancer detection Speech recognition	
2 Unsupervised	Given input data only, the algorithm identifies **clusters**, **patterns** and **anomalies** in the data.	Clustering / Grouping Recommendation systems Fraud Detection	
3 Reinforcement	Given an environment and a reward function, the algorithm optimises its actions to **maximise its reward**.	Self-driving Cars Autonomous Robots Utilities (traffic management) Game Playing	

EXPLORE | DIGITAL SKILLS

Machine Learning in Practice

Preprocessing → **Model Building** → **Model Evaluation** → **Deployment**





- Data Cleaning
  - Impute
  - Normalise/Standardise
  - Label/Dummy encode
- Train-Test split / Kfold

- Model Selection
- Model Training
- Hyperparameter Tuning

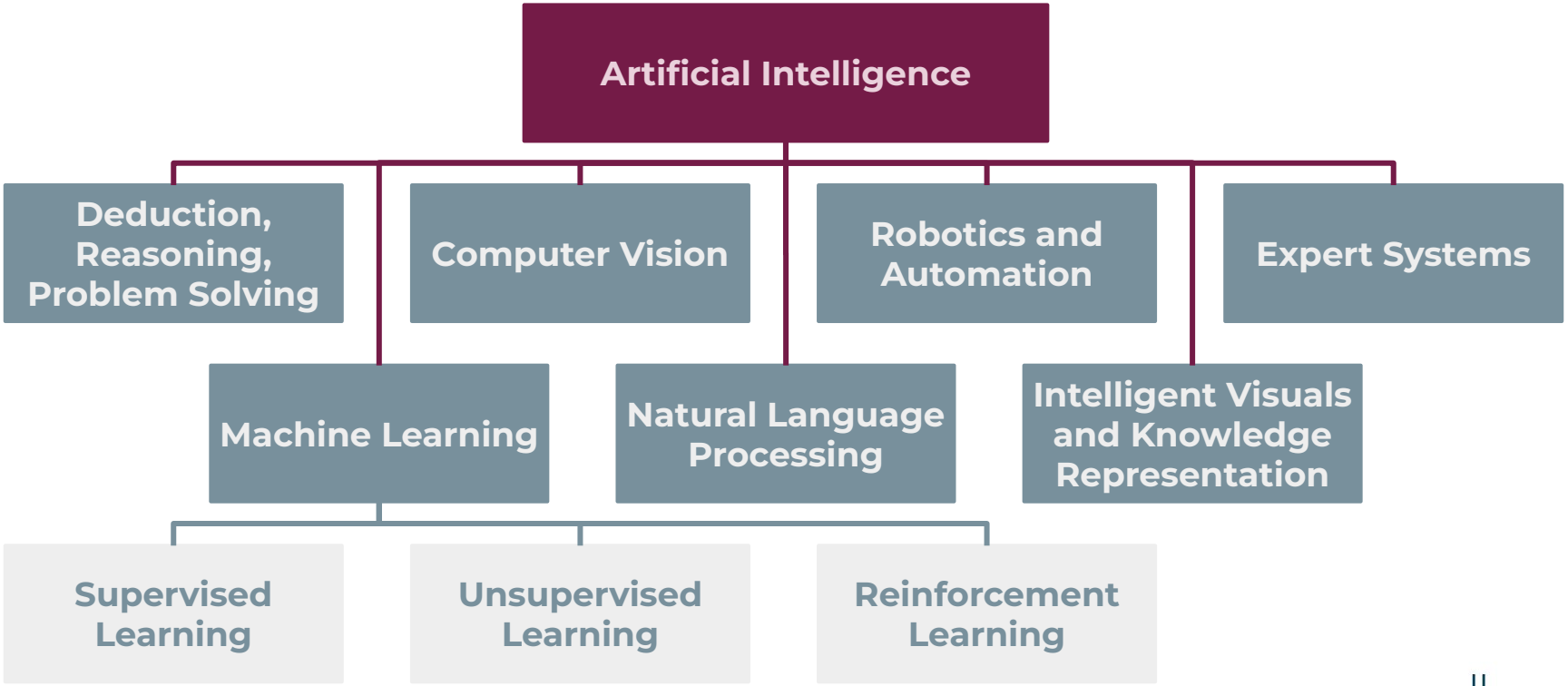
- Evaluate Model on Test set
- Report Performance metrics

- Hosting and Versioning
- Dashboards
- Containerization
  - Docker
  - Kubernetes

EXPLORE || DIGITAL SKILLS

Things to consider - the larger AI spectrum

Artificial Intelligence is the broad term used to **describe machines with cognitive ability**.



EXPLORE | DIGITAL SKILLS

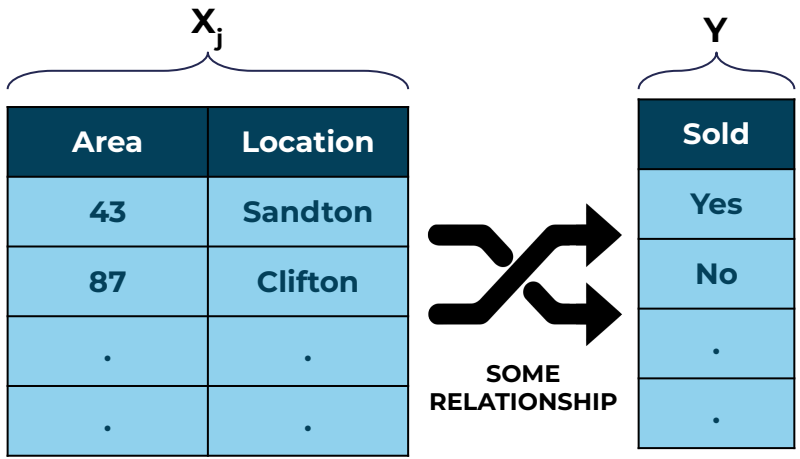
Introduction to classification

Classification is finding relationships between:

- A categorical, qualitative variable *Y*
  - Like whether a loan was taken, or a house sold
- And, a number of independent predictor variables, each known as an X_j
  - Like the size and and location of the house, type of tree, etc.

It's a type of supervised learning:

- Supervised learning is a subset of machine learning - simply the process of working out how some inputs relate to some outputs.
  - *Inputs*: independent X_j's described above;
  - *Outputs*: dependent *Y*, as above.



EXPLORE | DIGITAL SKILLS

Classification applications

CLASSIFICATION is the problem of identifying to which set of categories a new observation belongs

Regression	Predicting the value of a continuous variable	Usually numerical independent variables (can be categorical). Model for value of dependent variable.
Classification	Predicting the value of a **categorical** variable	Numerical and/or categorical independent variables. Model for category, or probability of belonging to category.

Health: Is a tumour malignant or benign, given information about its shape and other features?
ML classic: What species of iris is this, given sepal and petal length and width?
Illustration: Will a student graduate, given their aptitude test score and interview score?
Personality: What personality type is a person based on analysis of what they've written?

Classification algorithms

What algorithms can we use?

- Logistic Regression
- K-Nearest Neighbours
- Support Vector Machines
- Naïve Bayes
- Tree-Based Methods
- LDA / QDA
- Neural Networks

Is there a cat in this photo?

Implementation

Model Selection
Choosing the algorithm suitable for the task and dataset.

Training
Fitting the model to the many data points we have.

Assessment
Is the model actually any good?

Hyperparameters
Optimising the performance of the selected algorithm.

EXPLORE | DIGITAL SKILLS

Natural language processing (NLP)

NLP is the study of how computers can interpret and parse human language.

SPAM FILTERS

- Scan the text of each email.
- Attempt to gain context or understanding.
- Determine whether spam or not.

ALGORITHMIC TRADING

- Read and digest masses of news and articles relevant to stocks.
- Combined with ML, determines buy/hold/sell positions.

ANSWERING QUESTIONS

- Major use-case: have search engines understand what we mean.
- Bonus: respond in the same language, tone, etc.
- Used widely in Siri, Google Assistant, Alexa, etc.

SUMMARIZING INFORMATION

- Far too much info out there for us to process wholly.
- Using NLP we can parse large document volumes.
- Attempt to understand meaning and generate summaries.

https://web.archive.org/web/20171204090801/http://blog.neospeech.com/what-is-natural-language-processing/

Your EGAD Classification Sprint Heatmap

	Explain	Gather	Analyse	Deploy
DRAFT	Problem Statement	Problem Landscape	Equation of Value	Project Management
DO	Story Telling	Databases	Programming	Version Control
DELIVER	Communication	Data Engineering	Solution Governance	Production
DECOMPRESS	Feedback	Insights	Performance Metrics	Maintenance

EXPLORE DIGITAL SKILLS