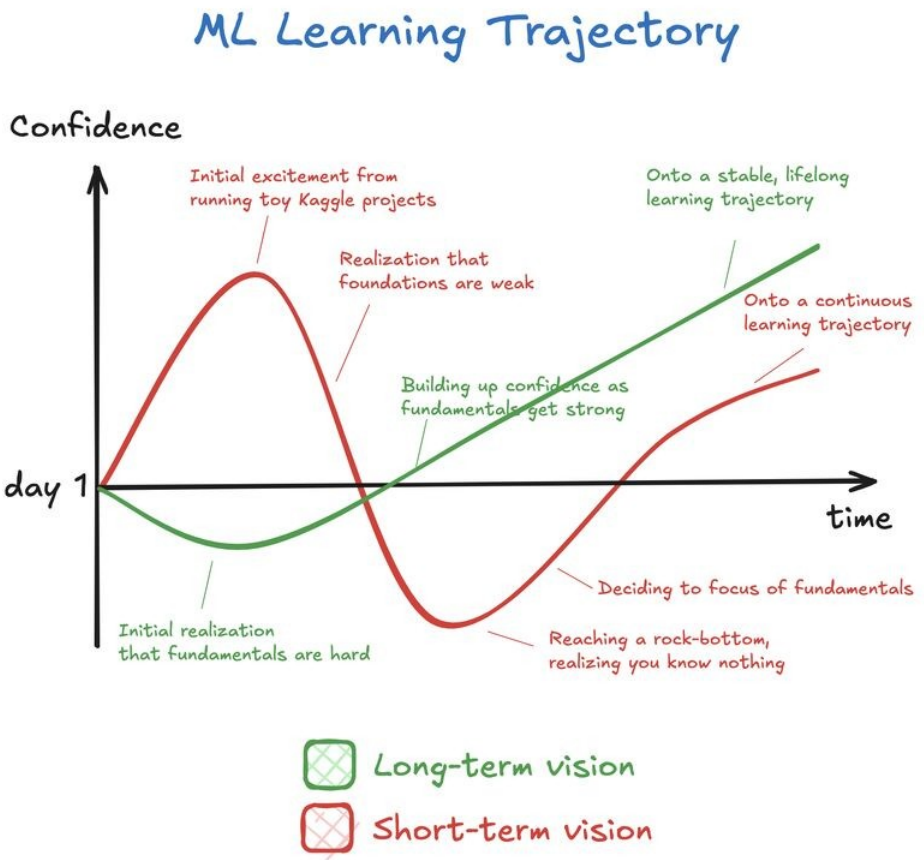
Machine Learning

Gagan Puri
linkedin.com/in/puri-gagan
higaganpuri@gmail.com

Contents

- AI Introduction
- Machine Learning Types
- Essential Libraries and Tools
- Anaconda and Environment Management
- How it looks like
- Next Class

Overwhelmingness

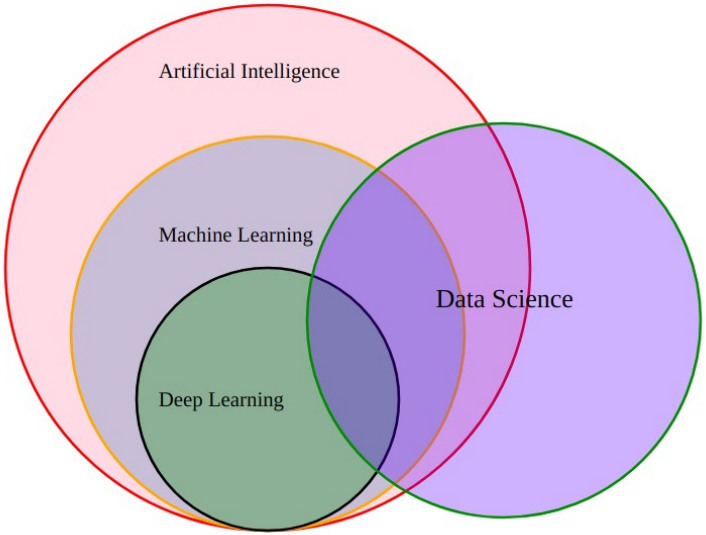


Structure

- Real quick revision of previous class
- Pre-Cap
- Short discussion on the DYOR topic if given previously
- Proceed ….
- Interactive Examples with Jupyter Notebook
- Over: with short DYOR

Introduction

- AI is the broadest term of the four. It refers to the development of machines that can perform tasks that typically require human intelligence
- AI is about creating “smart” machines that can mimic human behaviors, make decisions, solve problems, understand language
- Autonomous Systems
- ML is a subset of AI that focuses on creating systems that can learn from data, identify patterns, and make decisions without explicit programming
- Once trained, the model can make predictions or decisions based on new, unseen data
- DL is a specialized branch of ML that focuses on neural networks algorithm that mimic the way the human brain processes information



Still… What’s difference? ML vs DL

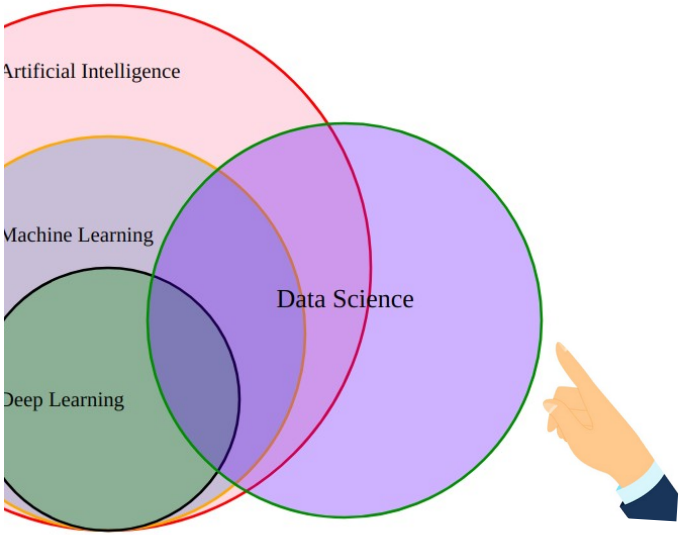
- How ML works:
  - Training: You feed a model with historical data (sample data)
  - Learning: The model identifies patterns in the data
  - Prediction: The model makes predictions about new data
- How Deep Learning works:
  - Input Layer: Raw data, such as an image, is fed into the system
  - Hidden Layers: Multiple layers of neurons process the data by applying mathematical transformations
  - Output Layer: The network generates a prediction, such as Classification: e.g. whether a photo contains a cat or a dog

And what about Data Science

- Data Science is an interdisciplinary field that focuses on extracting insights from large datasets
- Use a combination of statistics, programming, machine learning, and domain knowledge
- Analyze data and drive decision-making

Why it’s certain domain is outside of AI:

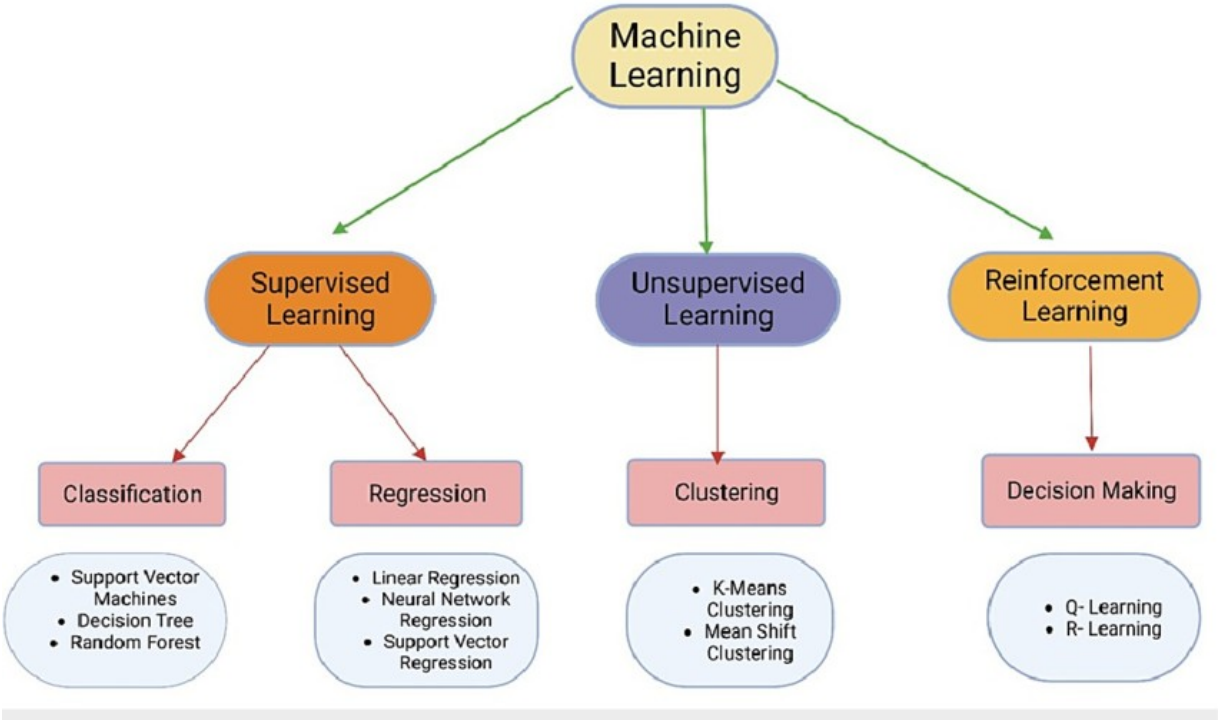
- Data Science is not about creating intelligent systems
- It’s about understanding and interpreting data to inform business strategies, product development, and more
- Data analysis, statistics, and domain expertise to derive actionable insights from data
- Can use AI



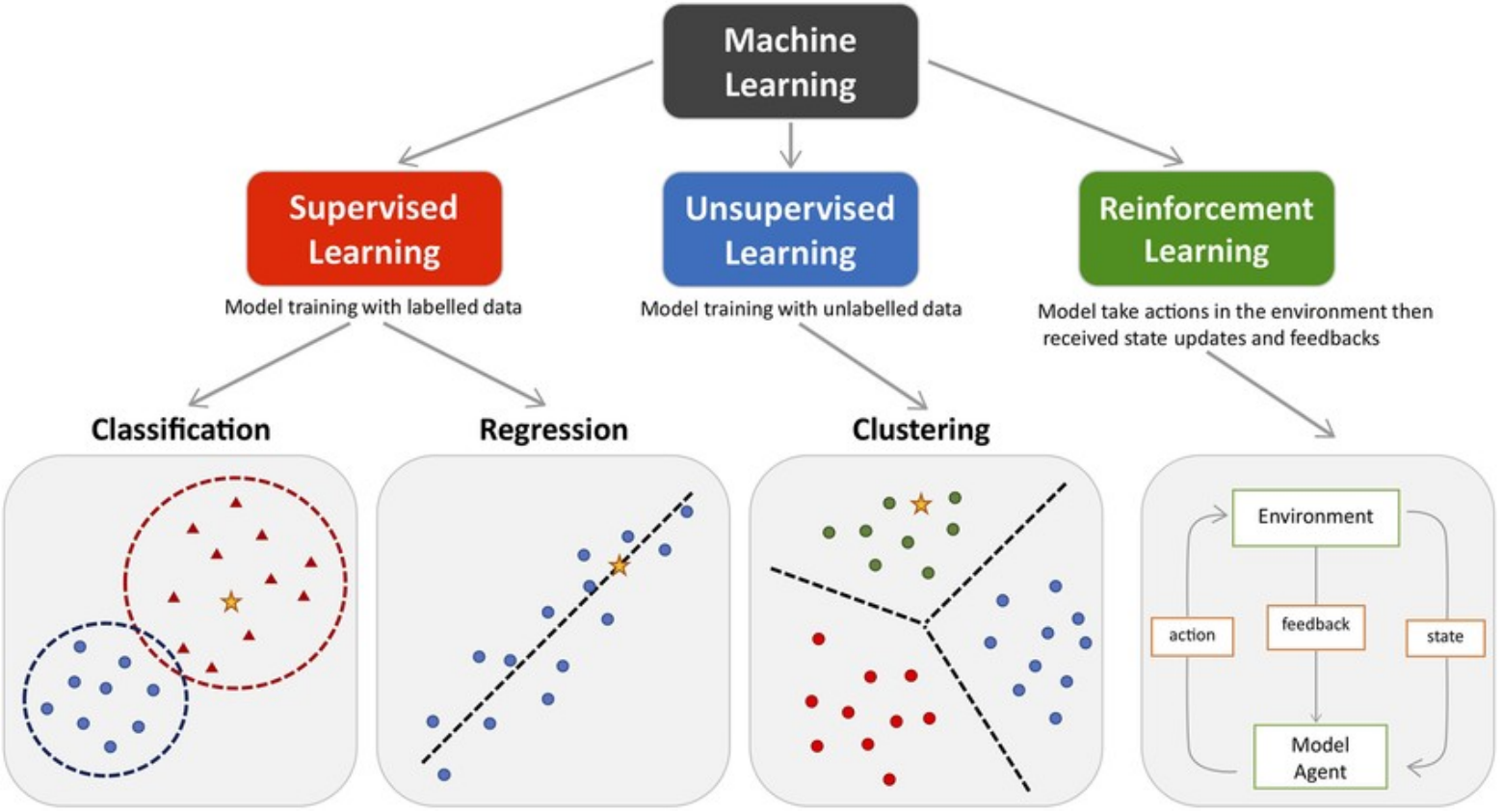
Machine Learning

- Teaches computers to recognize patterns and make decisions automatically using data and algorithms

- **Broadly categorized into three types**

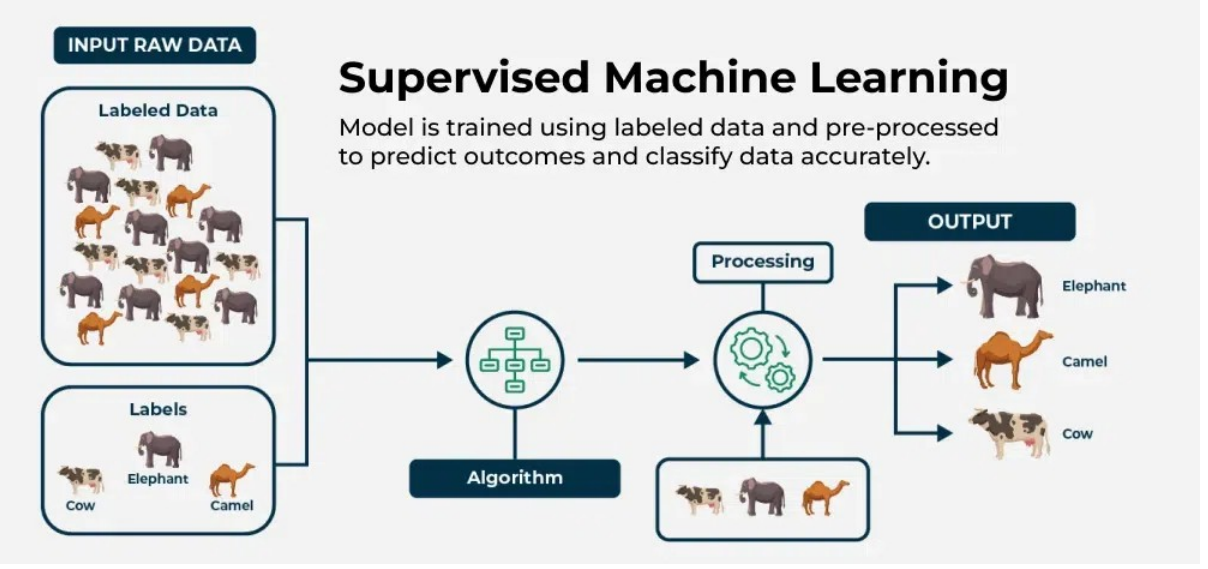


Types


Machine Learning
Supervised Learning
Unsupervised Learning
Reinforcement Learning
Model training with labelled data
Model training with unlabelled data
Model take actions in the environment then received state updates and feedbacks
Classification
Regression
Clustering
Environment
action
feedback
state
Model Agent

Supervised Machine Learning

- A model is trained on labeled data: each input is paired with the correct output
- Learning a class from examples: providing the model with examples where the correct label is known



Supervised Machine Learning Types

- Applied to two **main types** of problems
- Classification: Where the output is a categorical variable (e.g., spam vs. non-spam emails, yes vs. no)
- Regression: Where the output is a continuous variable (e.g., predicting house prices, stock prices)

User ID	Gender	Age	Salary	Purchased
15624510	Male	19	19000	0
15810944	Male	35	20000	1
15668575	Female	26	43000	0
15603246	Female	27	57000	0
15804002	Male	19	76000	1
15728773	Male	27	58000	1
15598044	Female	27	84000	0
15694829	Female	32	150000	1
15600575	Male	25	33000	1
15727311	Female	35	65000	0
15570769	Female	26	80000	1
15606274	Female	26	52000	0
15746139	Male	20	86000	1
15704987	Male	32	18000	0
15628972	Male	18	82000	0
15697686	Male	29	80000	0
15733883	Male	47	25000	1

Figure A: CLASSIFICATION

Temperature	Pressure	Relative Humidity	Wind Direction	Wind Speed
10.69261758	986.882019	54.19337313	195.7150879	3.278597116
13.59184184	987.8729248	48.0648859	189.2951202	2.909167767
17.70494885	988.1119385	39.11965597	192.9273834	2.973036289
20.95430404	987.8500366	30.66273218	202.0752869	2.965289593
22.9278274	987.2833862	26.06723423	210.6589203	2.798230886
24.04233986	986.2907104	23.46918024	221.1188507	2.627005816
24.41475295	985.2338867	22.25082295	233.7911987	2.448749781
23.93361956	984.8914795	22.35178837	244.3504333	2.454271793
22.68800023	984.8461304	23.7538641	253.0864716	2.418341875
20.56425726	984.8380737	27.07867944	264.5071106	2.318677425
17.76400389	985.4262085	33.54900114	280.7827454	2.343950987
11.25680746	988.9386597	53.74139903	68.15406036	1.650191426
14.37810685	989.6819458	40.70884681	72.62069702	1.553469896
18.45114201	990.2960205	30.85038484	71.70604706	1.005017161
22.54895853	989.9562988	22.81738811	44.66042709	0.264133632
24.23155922	988.796875	19.74790765	318.3214111	0.329656571

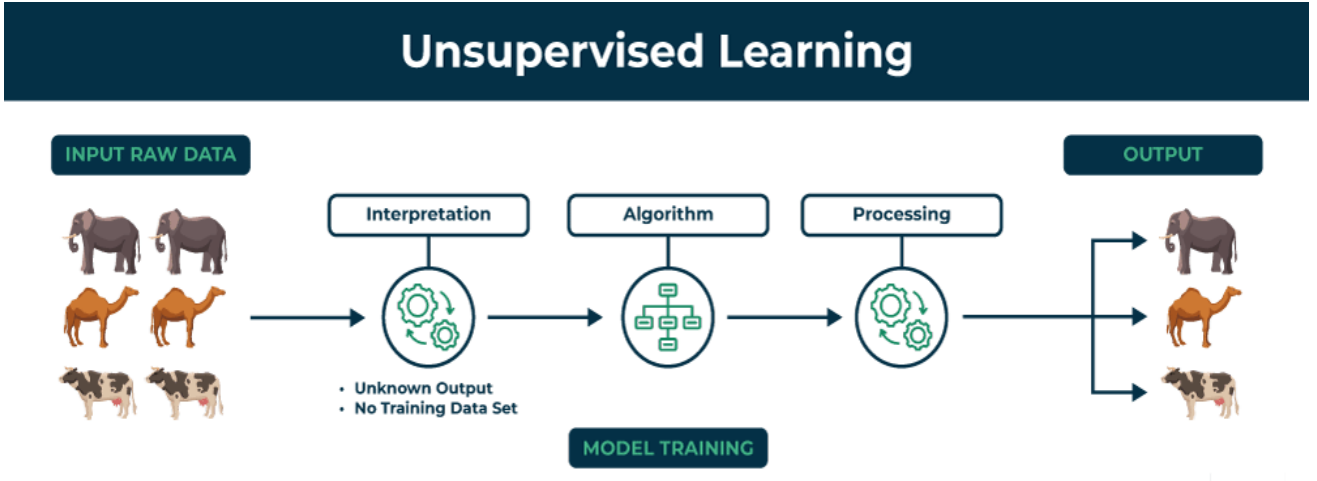
Figure B: REGRESSION

Supervised Machine Learning Algorithms

Algorithm	Regression, Classification	Purpose	Method	Use Cases
Linear Regression	Regression	Predict continuous output values	Linear equation minimizing sum of squares of residuals	Predicting continuous values
Logistic Regression	Classification	Predict binary output variable	Logistic function transforming linear relationship	Binary classification tasks
Decision Trees	Both	Model decisions and outcomes	Tree-like structure with decisions and outcomes	Classification and Regression tasks
Random Forests	Both	Improve classification and regression accuracy	Combining multiple decision trees	Reducing overfitting, improving prediction accuracy
SVM	Both	Create hyperplane for classification or predict continuous values	Maximizing margin between classes or predicting continuous values	Classification and Regression tasks
KNN	Both	Predict class or value based on k closest neighbors	Finding k closest neighbors and predicting based on majority or average	Classification and Regression tasks, sensitive to noisy data
Gradient Boosting	Both	Combine weak learners to create strong model	Iteratively correcting errors with new models	Classification and Regression tasks to improve prediction accuracy
Naive Bayes	Classification	Predict class based on feature independence assumption	Bayes' theorem with feature independence assumption	Text classification, spam filtering, sentiment analysis, medical

Unsupervised Machine Learning

- Tasked with finding patterns and relationships within the data without any prior knowledge of the data's meaning
- The training model has only input parameter values and discovers the groups or patterns on its own
- Finds hidden patterns and data without any human intervention: we don't give output to our model



Unsupervised Machine Learning Types

- Applied to 3 **main types** of Algorithms
- Clustering
  - K-means Clustering: Groups data into K clusters based on how close the points are to each other
- Association Rule Learning
  - Apriori Algorithm: Finds patterns by exploring frequent item combinations step-by-step
- Dimensionality Reduction
  - PCA: Reduces dimensions by transforming data into uncorrelated principal components.

Example Application: Unsupervised Machine Learning

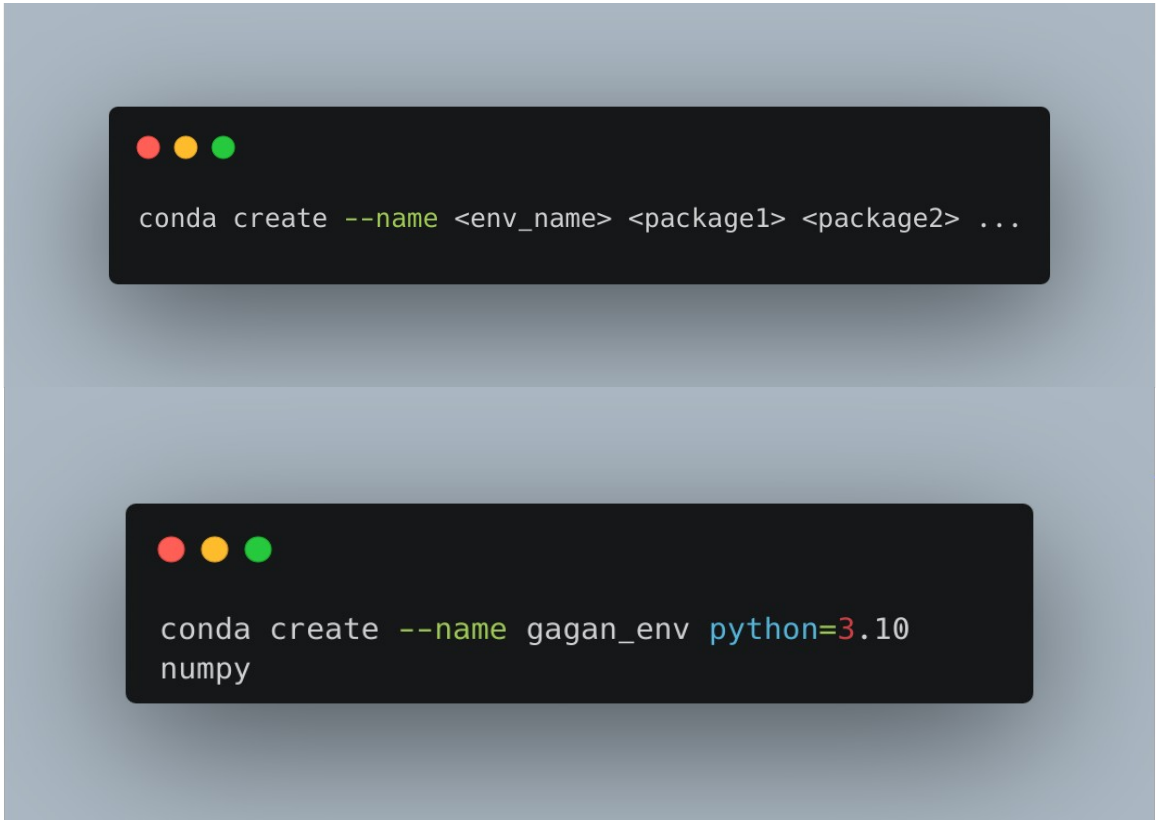
- You have 1,000 news articles from various sources, covering a range of topics like politics, sports, technology, health, and entertainment without any manual labeling. After clustering:
  - Group 1: Articles about the latest political developments (e.g., "Senator proposes new healthcare bill," "President addresses climate change issues").
  - Group 2: Articles about football and basketball games (e.g., "Champions League final preview," "NBA playoffs predictions").
  - Group 3: Articles on advancements in AI and new tech gadgets (e.g., "AI models surpass human-level performance," "New smartphone launches this year").

Essential Libraries and Tools

- Conda
- Jupyter Notebook
- NumPy
- Pandas
- Matplotlib / Seaborn
- Sklearn
- TensorFlow
- PyTorch
- XGBoost & LightGBM
- SciPy
- NLTK / Spacy
- OpenCV

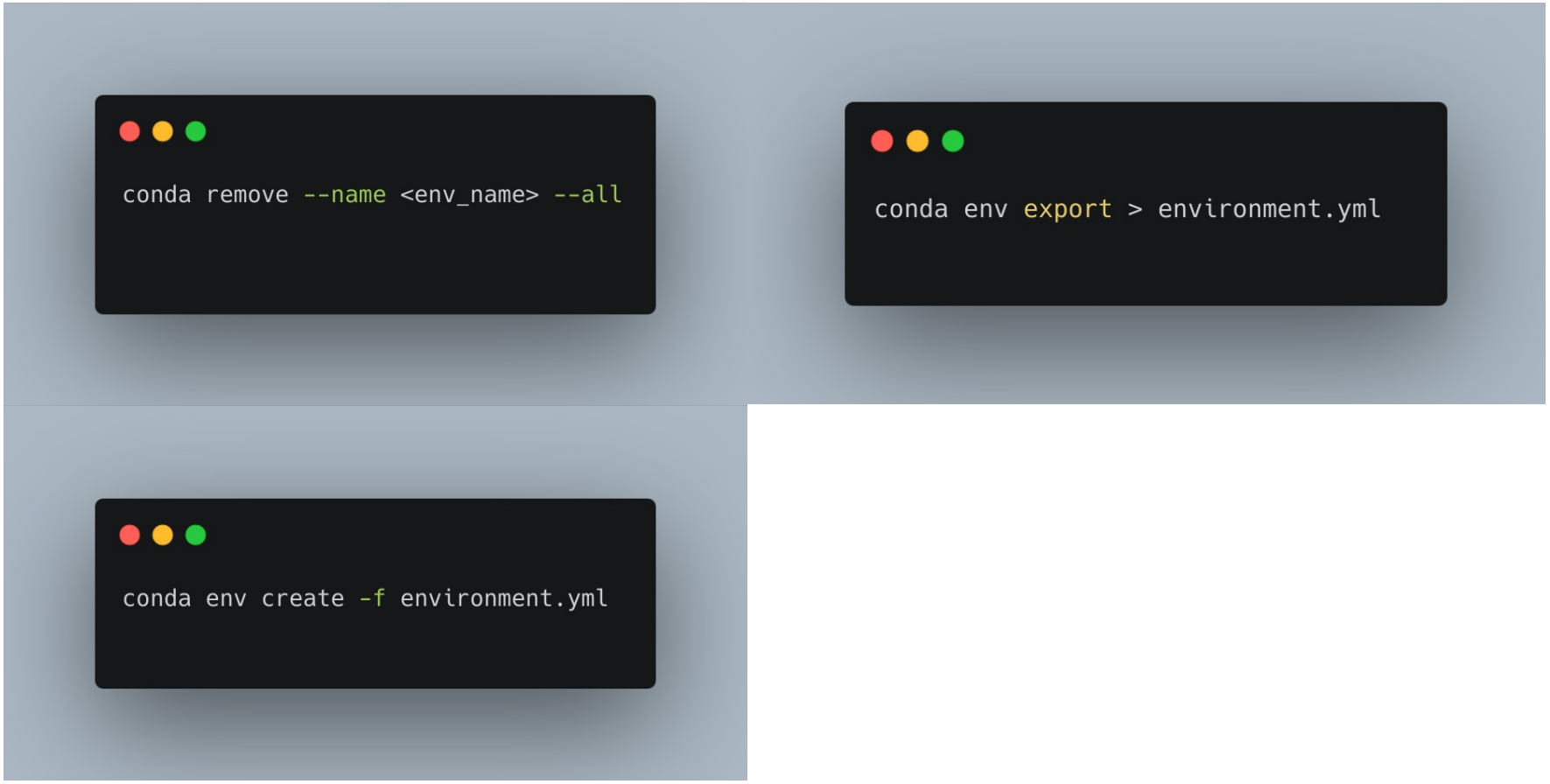
Installation

Anaconda / Mini Conda Environments



Managing Environments

Managing Environments



Codes?

Next Class

- Datasets
- Numpy
- Scikit Learn
- Matplotlib
- Pandas