Machine Learning Specialization

# A brief History of Modern AI and its Applications

[Machine Learning and Deep Learning - Part 1](https://www.coursera.org/learn/ibm-exploratory-data-analysis-for-machine-learning/lecture/3Pud9/machine-learning-and-deep-learning-part-1?trk_ref=coach_copy)

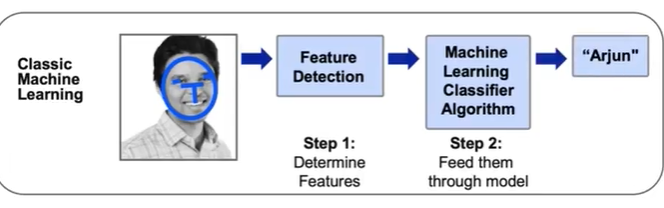
The lecture introduces **Machine Learning** as a subset of **Artificial Intelligence** that focuses on programs that learn from data rather than being explicitly programmed. Key points include:

* **Learning from Data**: Machine Learning algorithms improve as they are exposed to more data, although their performance may plateau after a certain point.
* **Features and Target**: In a dataset, features are the input variables (e.g., sepal length, sepal width) used for predictions, while the target is the outcome we want to predict (e.g., species of a flower).
* **Types of Machine Learning**:
  + **Supervised Learning**: Involves labeled data to predict outcomes (e.g., spam detection).
  + **Unsupervised Learning**: Involves unlabeled data to find patterns (e.g., customer segmentation).
* **Examples**:
  + Supervised Learning: Fraud detection using labeled transaction data.
  + Unsupervised Learning: Grouping customers for targeted marketing without predefined labels.

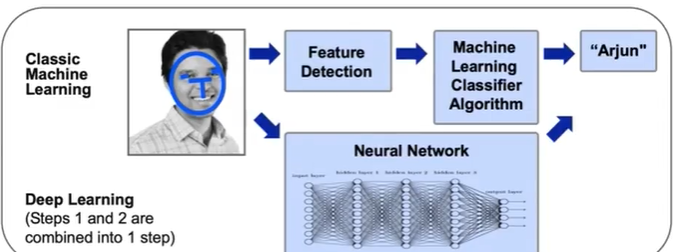
[Machine Learning and Deep Learning - Part 2](https://www.coursera.org/learn/ibm-exploratory-data-analysis-for-machine-learning/lecture/WfFsY/machine-learning-and-deep-learning-part-2?trk_ref=coach_copy)

In the current course content, the focus is on **Deep Learning** and its advantages over traditional **Machine Learning** techniques, particularly in image classification tasks. Here are the key points:

* **Feature Definition**: Identifying features in images (like distinguishing between a cat and a dog) is complex. Each pixel can be considered a feature, leading to a vast number of features (e.g., 65,000 for a 256x256 image).



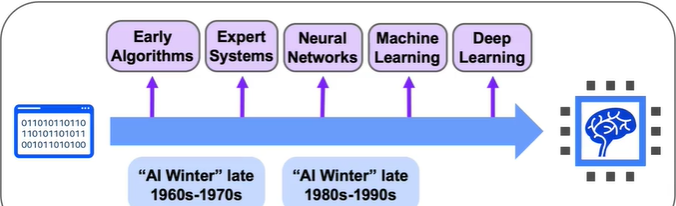
* **Spatial Relationships**: Traditional methods may overlook the spatial relationships between pixels, which are crucial for understanding images.
* **Deep Learning**: This approach uses deep neural networks to automatically learn and extract meaningful features from images, combining pixels to recognize patterns (like edges and shapes).



* **Comparison**: While Deep Learning excels with large datasets, traditional Machine Learning can outperform it with smaller, stable datasets or when data changes frequently.

[History of AI](https://www.coursera.org/learn/ibm-exploratory-data-analysis-for-machine-learning/lecture/Snx9k/history-of-ai?trk_ref=coach_copy)

The lecture discusses the **history of artificial intelligence (AI)**, highlighting its cyclical nature of excitement and disappointment, often referred to as "AI Winters."



Here are the key points:

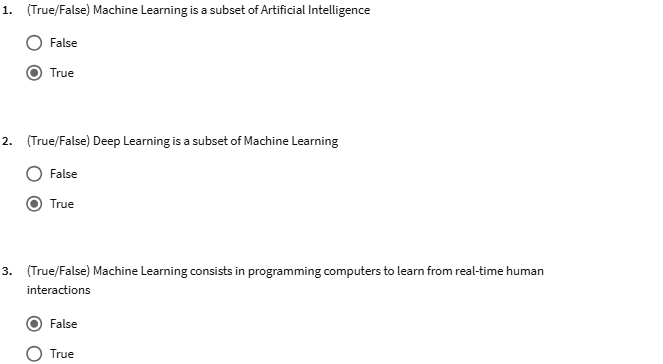
* **1950s**: The term "artificial intelligence" was coined at the Dartmouth Conference. Alan Turing developed the Turing test to assess machine intelligence. Frank Rosenblatt introduced the perceptron algorithm, a precursor to modern neural networks.
* **1960s-1970s**: The first AI Winter occurred due to unmet expectations, leading to reduced funding and interest in AI research.
* **1980s**: A resurgence in AI occurred with the rise of expert systems, which used programmed rules to mimic human decision-making. Geoffrey Hinton's work on the Backpropagation algorithm advanced neural networks.
* **Late 1980s-1990s**: Another AI Winter followed as expert systems failed to deliver practical solutions, leading to decreased investments.
* **Late 1990s-2000s**: Machine learning gained traction with successful applications in speech recognition and search algorithms.
* **Present Day**: Breakthroughs in deep learning have allowed AI to excel in complex tasks, such as image classification and machine translation, marking a new era of excitement in the field.

[History of Machine Learning and Deep Learning](https://www.coursera.org/learn/ibm-exploratory-data-analysis-for-machine-learning/lecture/0QG1n/history-of-machine-learning-and-deep-learning?trk_ref=coach_copy)

In the lecture, the focus is on the evolution of artificial intelligence (AI) from the 1990s to the 2010s. Key points include:

* **1990s-2000s**: AI solutions gained traction in areas like speech recognition and robotics. Notable achievements included Deep Blue defeating a world chess champion in 1996 and Google's PageRank algorithm revolutionizing search engines.
* **2006**: Advancements in deep learning emerged, overcoming issues like exploding and vanishing gradients, leading to deeper neural networks.
* **2009**: The introduction of the ImageNet database provided millions of labeled images, enhancing machine learning capabilities.
* **2012**: AlexNet, a convolutional neural network, achieved significant success in visual recognition tasks, marking a turning point for deep learning applications.
* **2013-2014**: Deep learning was applied to natural language processing and machine translation, improving web search and document summarization.
* **2015**: TensorFlow was released, making deep learning more accessible.
* **2016-2019**: Notable developments included DeepMind's AlphaGo defeating a Go master and IBM's Project Debater showcasing advanced argumentation skills.

**Quiz**:



[Modern AI](https://www.coursera.org/learn/ibm-exploratory-data-analysis-for-machine-learning/lecture/YFcPU/modern-ai?trk_ref=coach_copy)

The current lecture discusses the **AI landscape** today, highlighting significant advancements in **computer vision** and **natural language processing**. Key points include:

* **Growth Areas**:
  + **Computer Vision**: Used in self-driving cars and medical imaging for diagnosing illnesses.
  + **Natural Language Processing**: Improved capabilities in translation, sentiment analysis, and content generation.
* **Factors Driving Change**:
  + **Larger Datasets**: Access to diverse and extensive datasets, facilitated by cloud storage.
  + **Faster Computers**: Powerful hardware is now more accessible, enabling complex computations.
  + **Neural Networks**: Recent innovations in deep learning have led to practical applications across industries.
* **Industry Applications**:
  + **Healthcare**: AI aids in medical imaging and drug discovery.
  + **Finance**: Used for algorithmic trading and fraud detection.
  + **Transportation**: Autonomous vehicles and optimized logistics.
  + **Government**: Enhances public safety and city management.

[Applications](https://www.coursera.org/learn/ibm-exploratory-data-analysis-for-machine-learning/lecture/VMU0j/applications?trk_ref=coach_copy)

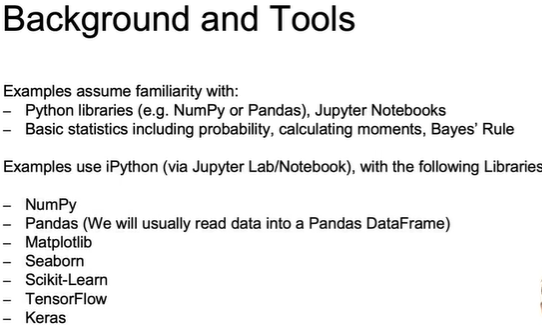
The lecture discusses the **current applications of artificial intelligence (AI)** in everyday life, particularly in personal transportation and social media. Key points include:

* **Navigation**: AI enhances navigation apps like Google Maps and Waze by considering traffic data and historical patterns to find the fastest routes.
* **Ride-sharing**: Companies like Uber and Lyft use AI to adjust pricing based on real-time supply and demand.
* **Social Media**: AI helps in content personalization, targeted advertising, image recognition, and sentiment analysis to improve user experience.
* **Voice Recognition**: Products like Siri and Alexa utilize natural language processing to understand and respond to user commands.
* **Computer Vision**: AI is crucial for self-driving cars and applications like face recognition on platforms like Facebook.

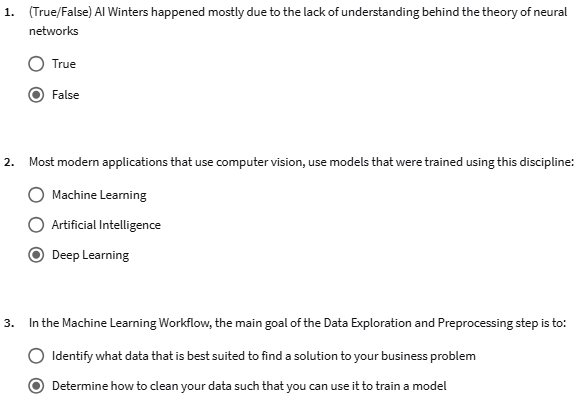
[Machine Learning Workflow](https://www.coursera.org/learn/ibm-exploratory-data-analysis-for-machine-learning/lecture/Mx9s2/machine-learning-workflow?trk_ref=coach_copy)

In this section, the lecture covers **basic vocabulary and workflow for Machine Learning**. Key points include:

* **Learning Goals**: Understanding fundamental concepts, tools, and the typical machine learning workflow.
* **Assumed Knowledge**: Familiarity with Python libraries (NumPy, Pandas), Jupyter Notebooks, and basic statistics (probability, moments, Bayes' rule).
* **Machine Learning Workflow**:
  1. **Problem Statement**: Define the problem to solve (e.g., image classification).
  2. **Data Collection**: Gather necessary data (e.g., labeled images).
  3. **Data Exploration and Preprocessing**: Clean and prepare data for modeling.
  4. **Modeling**: Build and test a model.
  5. **Validation**: Check model performance with a holdout set.
  6. **Decision-Making and Deployment**: Communicate results and implement the model.
* **Key Vocabulary**:
  1. **Target Variable**: The value to predict (e.g., species of iris).
  2. **Features**: Other variables used for prediction (e.g., sepal and petal dimensions).
  3. **Example/Observation**: A single row in the dataset.
  4. **Label**: The value of the target variable for an observation.



**Quiz**



**Summary of Module 1**

**Introduction to Artificial Intelligence and Machine Learning**

Artificial Intelligence is a branch of computer science dealing with the simulation of intelligent behavior in computers. Machines mimic cognitive functions such as learning and problem solving.

Machine learning is the study of programs that are not explicitly programmed, but instead these algorithms learn patterns from data.

Deep learning is a subset of machine learning in which multilayered neural networks learn from vast amounts of data.

**History of AI**

AI has experienced cycles of AI winters and AI booms.

AI solutions include speech recognition, computer vision, assisted medical diagnosis, robotics, and others.

**Modern AI**

Factors that have contributed to the current state of Machine Learning are: bigger data sets, faster computers, open source packages, and a wide range of neural network architectures.

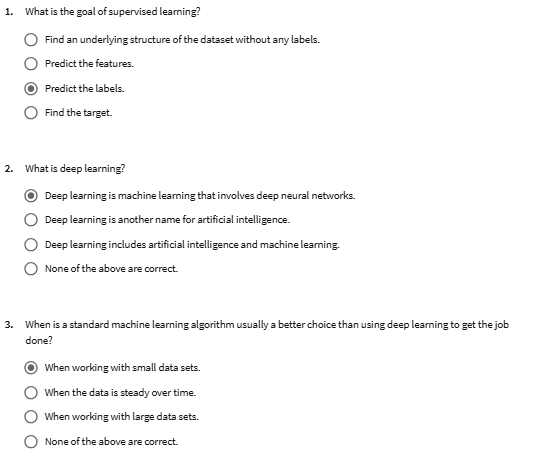
**Machine Learning Workflow**

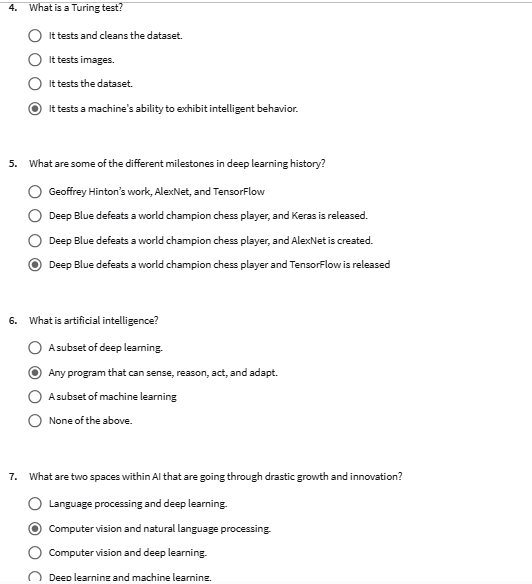
The machine learning workflow consists of:

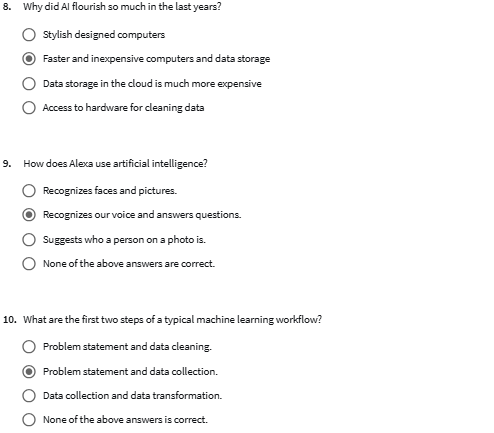
* Problem statement
* Data collection
* Data exploration and preprocessing
* Modeling
* Validation
* Decision Making and Deployment

This is a summary of the common taxonomy for data in open source packages for Machine Learning:

* target: category or value you are trying to predict
* features: explanatory variables used for prediction
* example: an observation or single data point within the data
* label: the value of the target for a single data point







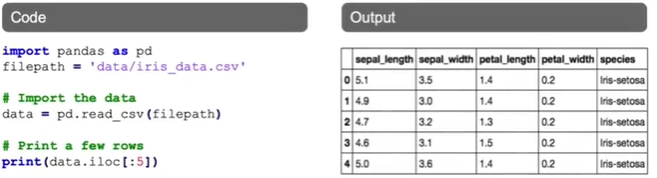
# Machine Learning Workflow

## Retrieving and Cleaning Data

[Retrieving Data from CSV and JSON Files](https://www.coursera.org/learn/ibm-exploratory-data-analysis-for-machine-learning/lecture/Lt8V6/retrieving-data-from-csv-and-json-files?trk_ref=coach_copy)

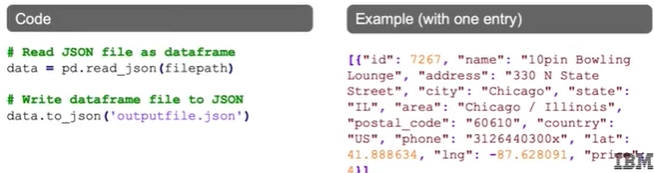
In this section, the focus is on **retrieving data from various sources**. Key points include:

* **Data Sources**: You will learn to pull data from SQL databases, NoSQL databases, APIs, and Cloud data sources.
* **CSV Files**:
  + CSV stands for **Comma Separated Values**.
  + You can read CSV files in Pandas using pd.read\_csv().
  + Important arguments include specifying separators, handling headers, and defining null values.





* **JSON Files**:
  + JSON stands for **JavaScript Object Notation** and is commonly used for data storage and APIs.
  + You can read JSON files using pd.read\_json().
  + Familiarity with the structure of JSON files is important for effective data retrieval.



[Retrieving Data from Databases, APIs, and the Cloud](https://www.coursera.org/learn/ibm-exploratory-data-analysis-for-machine-learning/lecture/nUznC/retrieving-data-from-databases-apis-and-the-cloud?trk_ref=coach_copy)

The lecture discusses working with **SQL** and **NoSQL** databases, highlighting the following key points:

* **SQL Databases**:
  + SQL (Structured Query Language) is used for relational databases with a fixed schema.
  + Examples include Microsoft SQL Server, Postgres, MySQL, and Oracle DB.
  + Python libraries like sqlite3, SQLAlchemy, and Psycopg2 are used to connect and query these databases.
  + A sample code snippet demonstrates how to connect to an SQLite database and retrieve data using a SQL query.
* **NoSQL Databases**:
  + These are non-relational databases that can store data in various formats, often using JSON.
  + Examples include document databases (like MongoDB) and graph databases.
  + The lecture provides a code example for connecting to a MongoDB database and retrieving data.
* **APIs and Cloud Data Access**:
  + Data can also be accessed through APIs, such as Twitter or Amazon.
  + A brief example shows how to read data from a CSV file using Pandas.

## Data Cleaning

Introduction

In this section, the focus is on the **importance of data cleaning** for machine learning. Key points include:

* **Garbage-in, garbage-out**: The quality of data directly affects the performance of machine learning models. If the data is messy, the results will be unreliable.
* **Common issues with messy data**:
  + **Duplicates**: Can skew model predictions if not handled properly.
  + **Inconsistent text and typos**: Variations in spelling or formatting can lead to misclassification of data.
  + **Missing data**: Can hinder the use of potentially valuable features.
  + **Outliers**: Can disproportionately affect the model's ability to find true relationships in the data.
  + **Data sourcing issues**: Challenges in integrating data from various sources can lead to mismatches.

The section emphasizes that ensuring clean data is the first step for organizations looking to leverage artificial intelligence or machine learning effectively. The next topic will cover handling missing values and outliers.

[Handling Missing Values and Outliers](https://www.coursera.org/learn/ibm-exploratory-data-analysis-for-machine-learning/lecture/uTXEf/handling-missing-values-and-outliers?trk_ref=coach_copy)

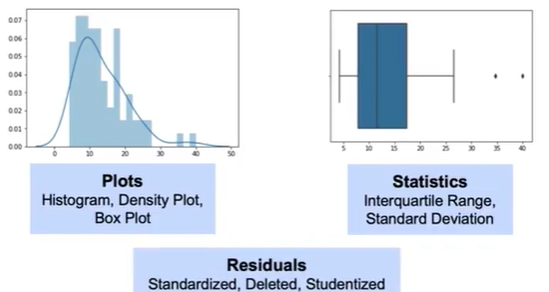
In the current lecture, the focus is on **handling missing values** and **working with outliers** in datasets. Here are the key points:

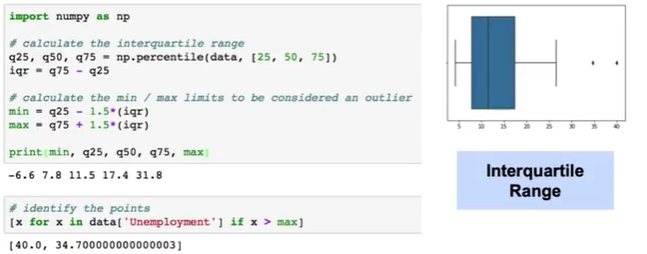
### Handling Missing Values:

1. **Removing Data**:
   * **Pros**: Quickly cleans the dataset.
   * **Cons**: May lose important information if many rows are removed.
2. **Imputing Data**:
   * Replacing missing values with mean, median, or estimates.
   * **Pros**: Retains all rows and columns.
   * **Cons**: Introduces uncertainty due to estimation.
3. **Masking Data**:
   * Treating missing values as a separate category.
   * **Pros**: Preserves all data.
   * **Cons**: Adds uncertainty by assuming all missing values are similar.

### Working with Outliers:

* **Definition**: Outliers are data points that differ significantly from others.
* **Impact**: They can skew predictions (e.g., a sales figure of 3,000 when most are between 10 and 50).
* **Detection Methods**:
  + **Plots**: Histograms, density plots, and box plots.
  + **Mathematical Methods**: Using percentiles and interquartile range to define outliers.





Understanding these concepts is crucial for preparing data for analysis and ensuring the accuracy of predictive models.

[Handling Missing Values and Outliers using Residuals](https://www.coursera.org/learn/ibm-exploratory-data-analysis-for-machine-learning/lecture/DIidO/handling-missing-values-and-outliers-using-residuals?trk_ref=coach_copy)

In the current lecture, the focus is on **residuals** in the context of model evaluation and outlier detection. Here are the key points:

* **Residuals** are the differences between actual values and predicted values from a model, indicating model failure.
* **Standardized Residuals**: These are calculated by dividing the residual by the standard error, allowing for comparison across different outcome ranges.
* **Deleted Residuals**: This involves removing an observation from the dataset to see how it affects the model's predictions.
* **Studentized Residuals**: Similar to deleted residuals, but these are standardized after removing an observation.

When outliers are detected, several approaches can be taken:

* **Remove the outlier**: Eliminates its effect but may lose important data.
* **Assign a different value**: Keeps the row intact but may lose valuable information.
* **Transform the data**: Techniques like log transformation can reduce the impact of outliers.
* **Predict the outlier's value**: Using regression or similar observations to estimate what the value should be.
* **Keep the outlier**: Use models that are resistant to outliers.