



Revision Notes: Introduction to Machine Learning

Understanding Machine Learning

Definition and Fundamental Concepts

- **Machine Learning (ML)** is the field of study that enables computers to learn from data without being explicitly programmed [4:19+source] . Arthur Samuel defines it as the capability for machines to learn from experience and improve their performance on specific tasks using that experience [4:4+source] [4:19+source] .
- **ML vs. Classical Programming:** Traditional programming involves writing explicit instructions for a computer to perform tasks. In contrast, ML relies on data and algorithms allowing the machine to deduce rules and patterns independently [4:18+source] .

ML Pipeline

1. **Input:** Training data, which is labeled to indicate outcomes (e.g., an email is labeled as spam or not spam).
2. **Processing:** Application of a learning algorithm to this data.
3. **Output:** A trained model that can predict or classify new data [4:0+source] .

Key Terms

- **Training Data:** The dataset used to train models. It includes features and labels to guide the learning process [4:0+source] .
- **Learning Algorithm:** A method or a set of rules followed in problem-solving operations, usually by a computer, to learn from data [4:0+source] .
- **Model:** An abstraction representing a mechanism through which data is processed to yield outcomes.



Supervised Learning

- Involves learning a function that maps an input to an output based on example input-output pairs **【4:5+source】** **【4:9+source】** .
- Tasks include **Regression** (predicting continuous values) and **Classification** (predicting categories) **【4:9+source】** .
- Examples:
 - Email spam detection.
 - Classifying images of fruits into categories like orange, kinu, or mosmein based on labeled data **【4:16+source】** .

Unsupervised Learning

- The model does not learn from labeled data. Instead, it tries to learn the structure from the input data **【4:5+source】** **【4:9+source】** .
- Common tasks include **Clustering**, where similar data points are grouped without prior labeling **【4:15+source】** .
- Example: Grouping customers into segments based on purchasing behavior without labeled data indicating the customer segments **【4:8+source】** .

Example Machine Learning Tasks

Spam Classification

- **Task:** Given an email, predict whether it is spam.
- **Experience (Training Data):** Historical emails labeled as spam or not.
- **Performance Metric:** Accuracy of spam detection **【4:4+source】** **【4:18+source】** .

Predicting Stock Prices

- **Task:** Predict future stock prices based on historical price data.
- **Experience:** Historical stock prices.
- **Performance Metric:** Mean squared error between predicted and actual stock prices **【4:16+source】** .



- **Plotting Data:** Various plots such as scatter plots for numerical data or bar charts for categorical data can be used to understand and visualize different types of data 【4:17+source】 .
- **Heatmaps:** Used for showing the intensity of data points or correlations between variables 【4:1+source】 【4:3+source】 .

Practical Considerations

- **Evaluation Metrics:** Useful for assessing how well a model performs. Common metrics include precision, recall, F1-score, and accuracy 【4:16+source】 .
- **Data Preprocessing:** Utilizes tools like NumPy and Pandas for data manipulation and Matplotlib for visualization 【4:17+source】 .

By completing this session, learners will have a conceptual understanding of basic machine learning techniques and their application in real-world scenarios using example problems discussed in classes.