

NAIRR Pilot

National Artificial Intelligence
Research Resource Pilot

Machine Learning vs. Deep Learning

Paola A. Buitrago

Director, AI and Big Data, PSC

April 3rd, 2025/ Track 2 – Intermediate or Advanced Participant

AI Workshop Denver, CO April 2-3, 2025



Machine Learning vs. Deep Learning

Track 2 – Intermediate or Advanced Participant

Welcome!

About this training!

- 1.5 hour of presentations + hands-on training.
- Assumes some background in Machine Learning.
- Assumes you are familiar with Jupyter notebooks, PyTorch, and OOD.
- Requires you to have set up your ACCESS ID.
- Requires access to DeltaAI and Expanse.
- Slack channel



Instructors and Support Staff

- Paola A. Buitrago, instructor
- Dana O'Connor
- Juliana Duncan
- Andrew Pasquale
- Vikram Gazula
- Devin Bayly (online)



Hands-On Segment of the Session

- How to ask for help?
 - Sticky notes on your laptop to signal that you need help!





Outline

This session provides a comparative overview of deep learning and machine learning, focusing on their distinct characteristics, applications, and requirements.

Key topics include:

1. Overview of Deep Learning vs. Machine Learning
2. Applications of Each Model
3. What You Need to Know Before Using These Models

By the end of the session, participants will understand when to use deep learning versus machine learning and the knowledge necessary to implement each effectively.



Outline

This session provides a comparative overview of deep learning and machine learning, focusing on their distinct characteristics, applications, and requirements.

Key topics include:

1. **Overview of Deep Learning vs. Machine Learning:**

Explanation of the differences between deep learning and traditional machine learning.

2. Applications of Each Model:

3. What You Need to Know Before Using These Models:

By the end of the session, participants will understand when to use deep learning versus machine learning and the knowledge necessary to implement each effectively.



Outline

This session provides a comparative overview of deep learning and machine learning, focusing on their distinct characteristics, applications, and requirements.

Key topics include:

1. Overview of Deep Learning vs. Machine Learning:

2. **Applications of Each Model:**

- Deep learning applications in image recognition, natural language processing, and speech recognition.
- Machine learning applications in predictive analytics, classification, and clustering.

3. What You Need to Know Before Using These Models:

By the end of the session, participants will understand when to use deep learning versus machine learning and the knowledge necessary to implement each effectively.



Outline

This session provides a comparative overview of deep learning and machine learning, focusing on their distinct characteristics, applications, and requirements.

Key topics include:

1. Overview of Deep Learning vs. Machine Learning:
2. Applications of Each Model:
3. **What You Need to Know Before Using These Models:**
 - Prerequisites for using DL and ML, such as data quality, computational power, and algorithm selection.
 - Considerations for choosing the right approach based on project requirements and available resources.

By the end of the session, participants will understand when to use deep learning versus machine learning and the knowledge necessary to implement each effectively.

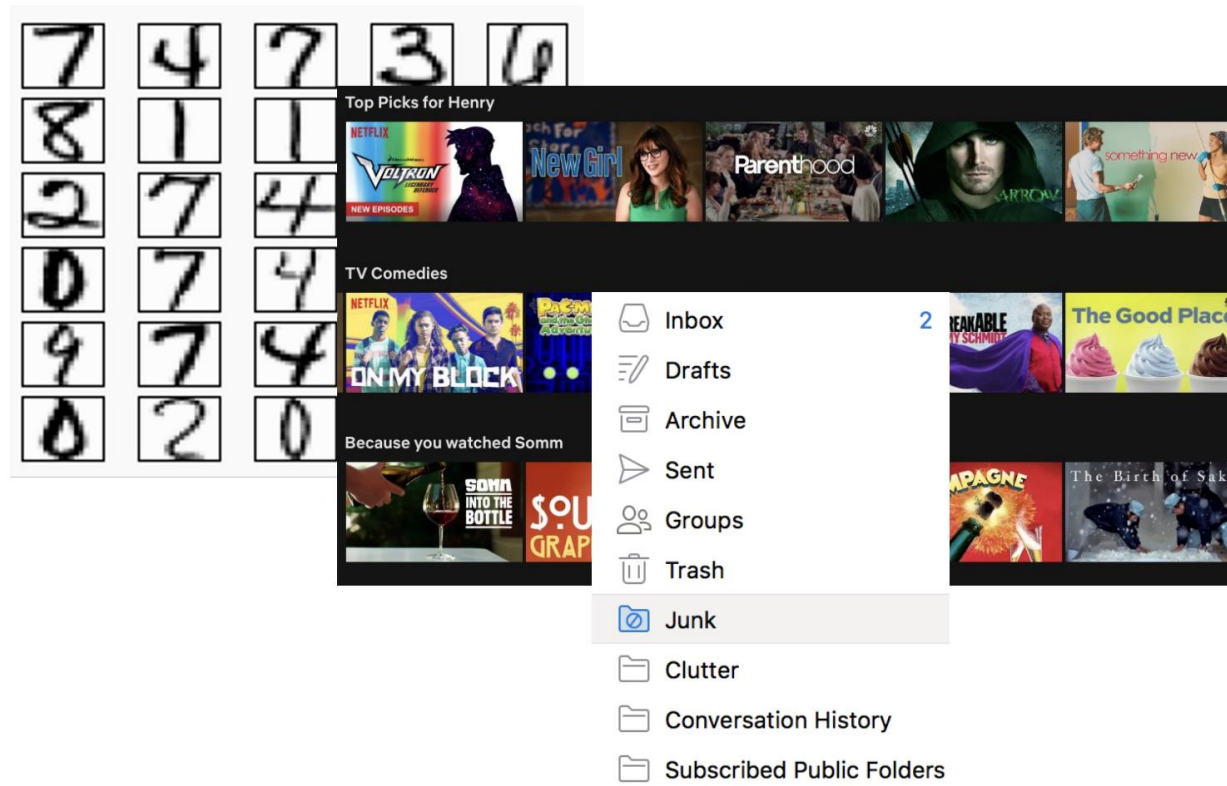


Outline

- 1. Overview of Deep Learning vs. Machine Learning**
2. Applications of Each Model
3. What You Need to Know Before Using These Models

What's is Machine Learning

Machine Learning a long time ago...



What's is Machine Learning

Machine Learning not so long ago...



What's is Machine Learning

Machine Learning not so long ago...





What's is Machine Learning

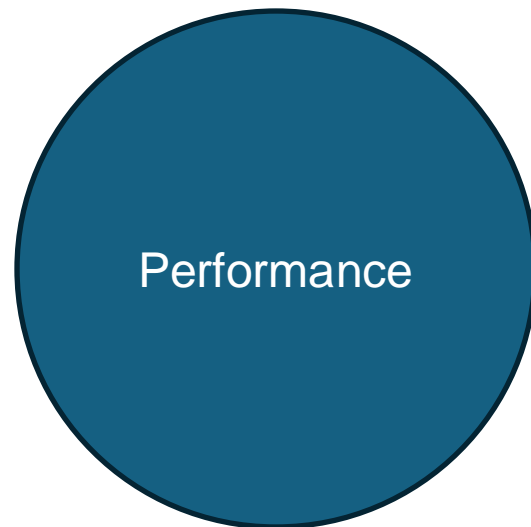
Machine learning definition:

Machine Learning (ML) is a subset of artificial intelligence (AI) that enables computers to learn from data and make predictions or decisions without being explicitly programmed. Instead of following a fixed set of rules, ML models identify patterns in data and improve their performance as they are exposed to more information.

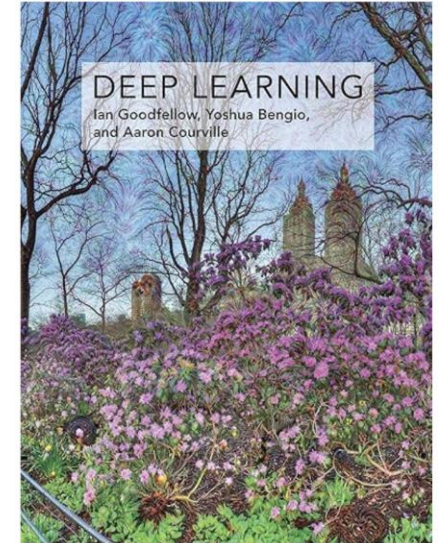
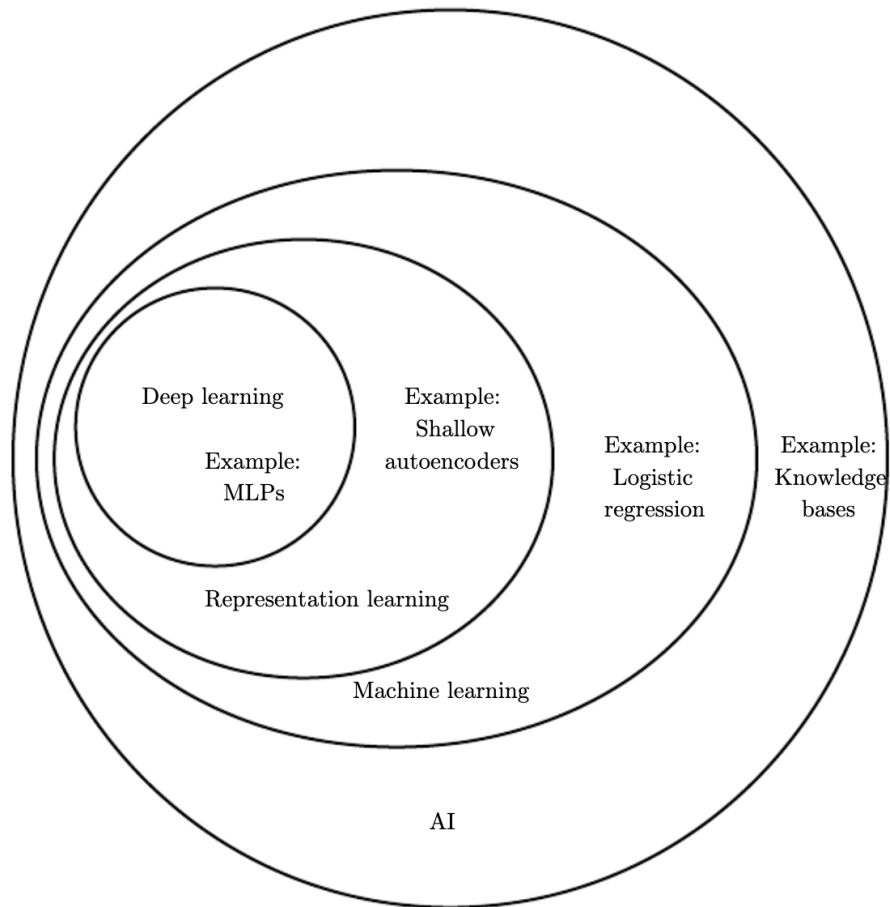
What's is Machine Learning

Machine learning definition:

- A computer program is said to **learn** if its *performance*, P , as some *task*, T , improves with *experience* E .
- Three components:



Machine Learning vs. Deep Learning



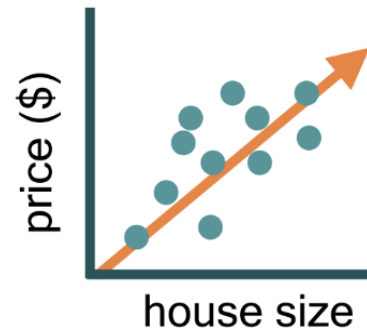
A Venn diagram showing how deep learning is a kind of representation learning, which is in turn a kind of machine learning, which is used for many but not all approaches to AI. Each section of the Venn diagram includes an example of an AI technology. Source: <https://www.deeplearningbook.org/contents/intro.html>

Machine Learning Pipeline



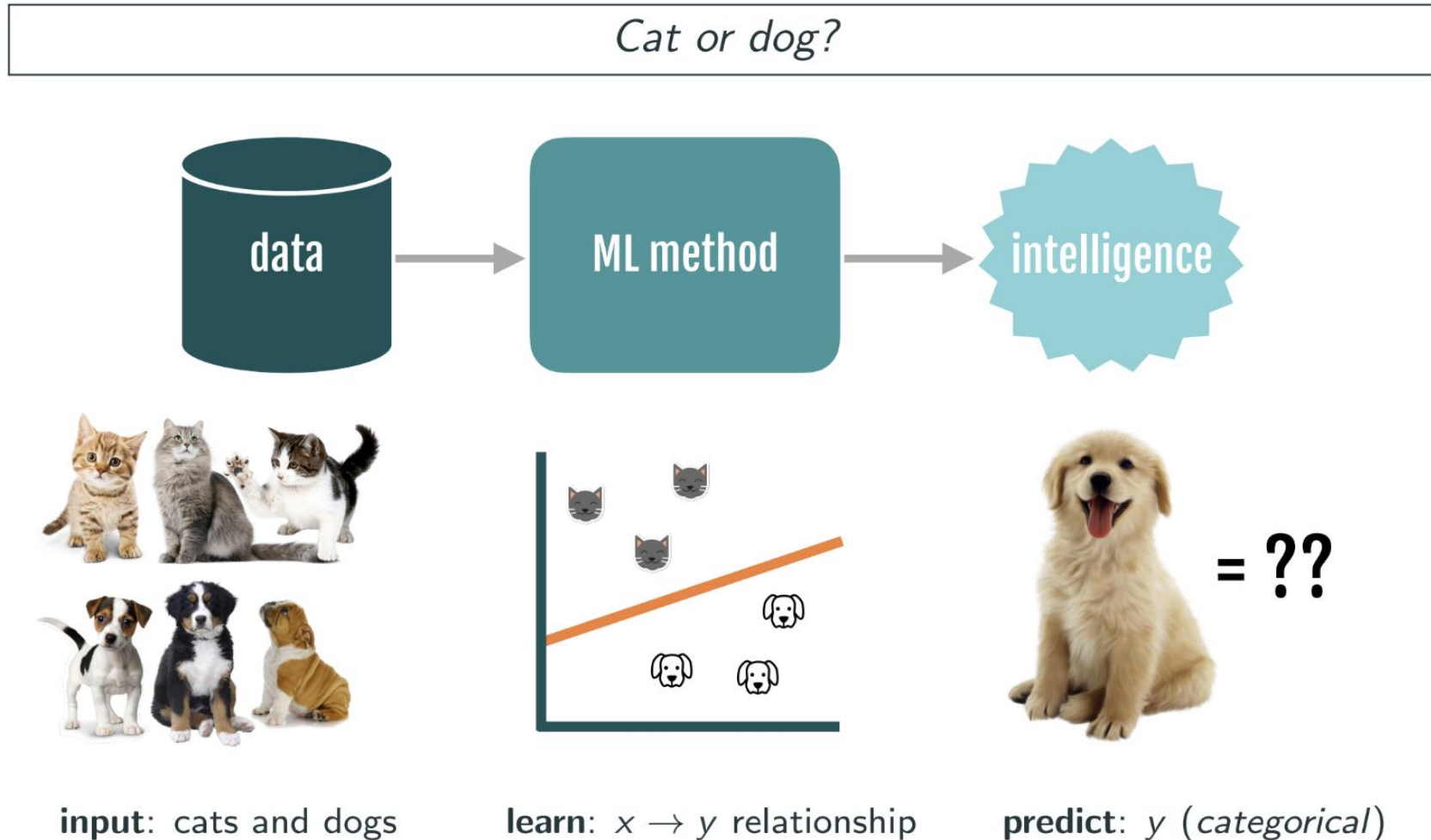
Task 1 - Regression

How much should you sell your house for?



input: houses & features **learn:** $x \rightarrow y$ relationship **predict:** y (*continuous*)

Task 2 - Classification



Task 3 - Clustering

How to segment an image?



input: raw pixels $\{x\}$



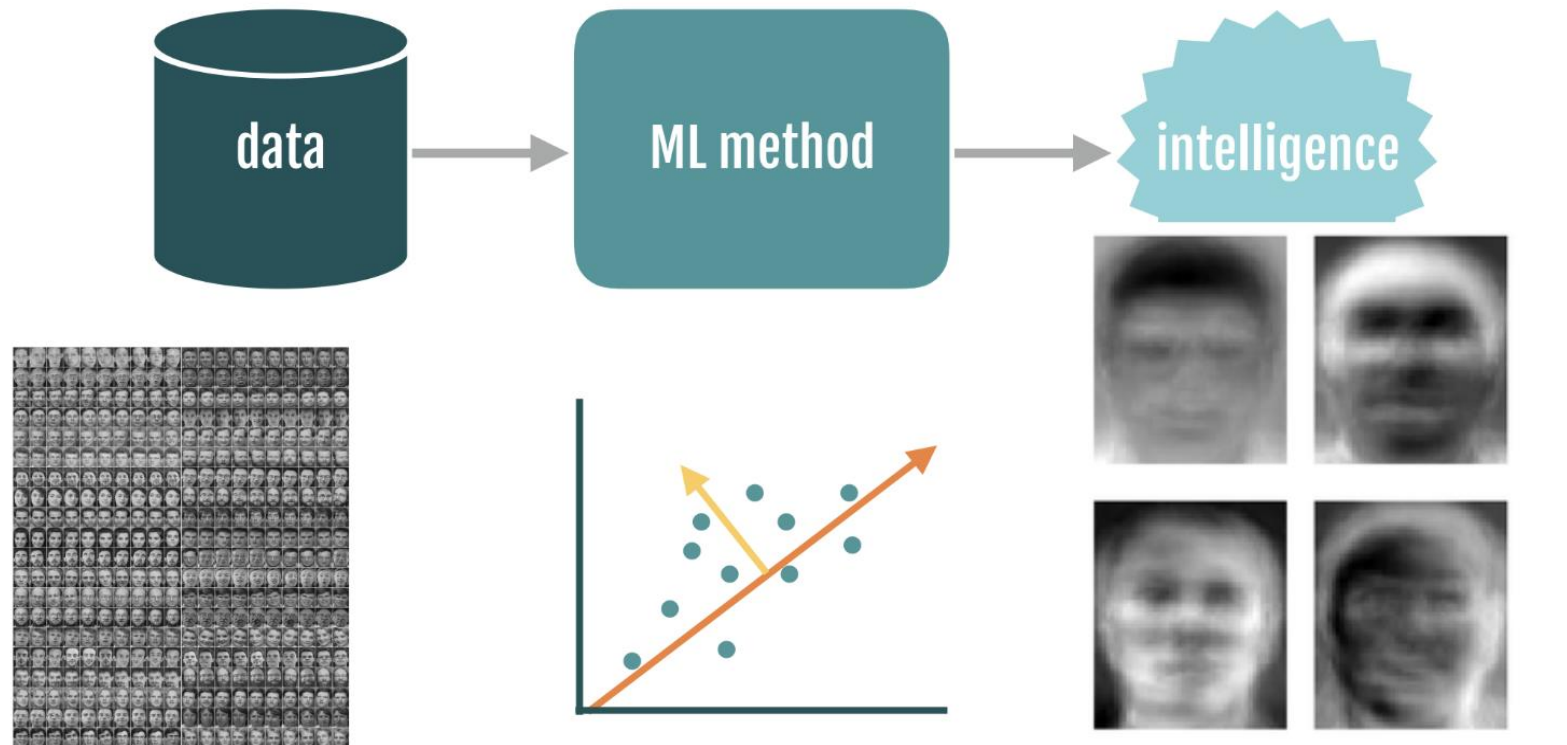
separate: $\{x\}$ into sets



output: cluster labels $\{z\}$

Task 4 - Embedding

How to efficiently represent data?



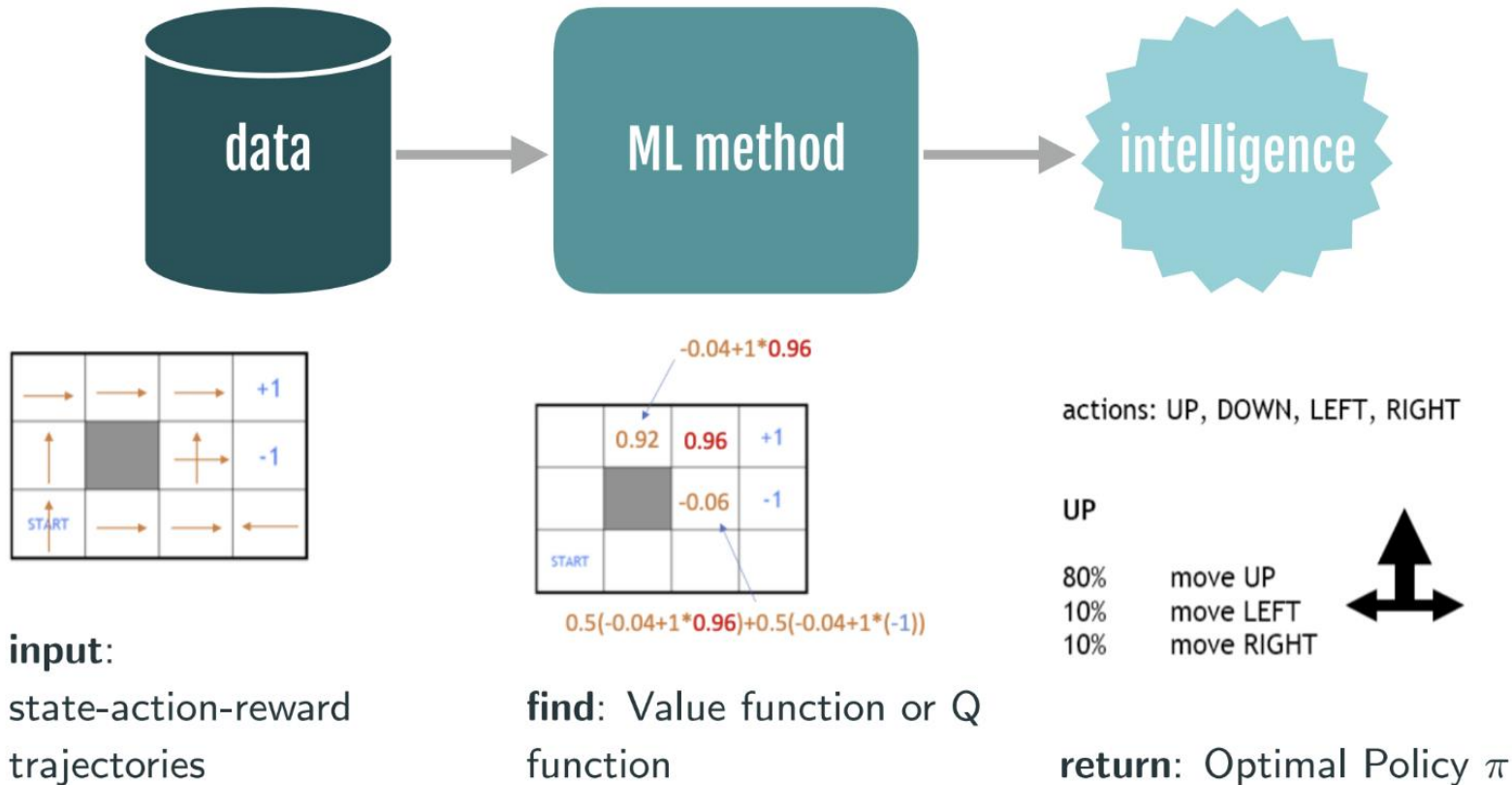
input: large dataset $\{x\}$

find: sources of variation

return: representation $\{z\}$

Task 5 - Reinforcement Learning

How to take the actions that maximize reward?



Other Tasks

Computer Vision

- Depth Estimation
- Image Classification
- Object Detection
- Image Segmentation
- Text-to-Image
- Image-to-Text
- Image-to-Image
- Image-to-Video
- Unconditional Image Generation
- Video Classification
- Text-to-Video
- Zero-Shot Image Classification
- Mask Generation
- Zero-Shot Object Detection
- Text-to-3D
- Image-to-3D
- Image Feature Extraction
- Keypoint Detection

Natural Language Processing

- Text Classification
- Token Classification
- Table Question Answering
- Question Answering
- Zero-Shot Classification
- Translation
- Summarization
- Feature Extraction
- Text Generation
- Text2Text Generation
- Fill-Mask
- Sentence Similarity

Audio

- Text-to-Speech
- Text-to-Audio
- Automatic Speech Recognition
- Audio-to-Audio
- Audio Classification
- Voice Activity Detection



Other Tasks

Multimodal

 Audio-Text-to-Text

 Image-Text-to-Text

 Visual Question Answering

 Document Question Answering

 Video-Text-to-Text

 Visual Document Retrieval

 Any-to-Any

Tabular

 Tabular Classification

 Tabular Regression

 Time Series Forecasting

Reinforcement Learning

 Reinforcement Learning

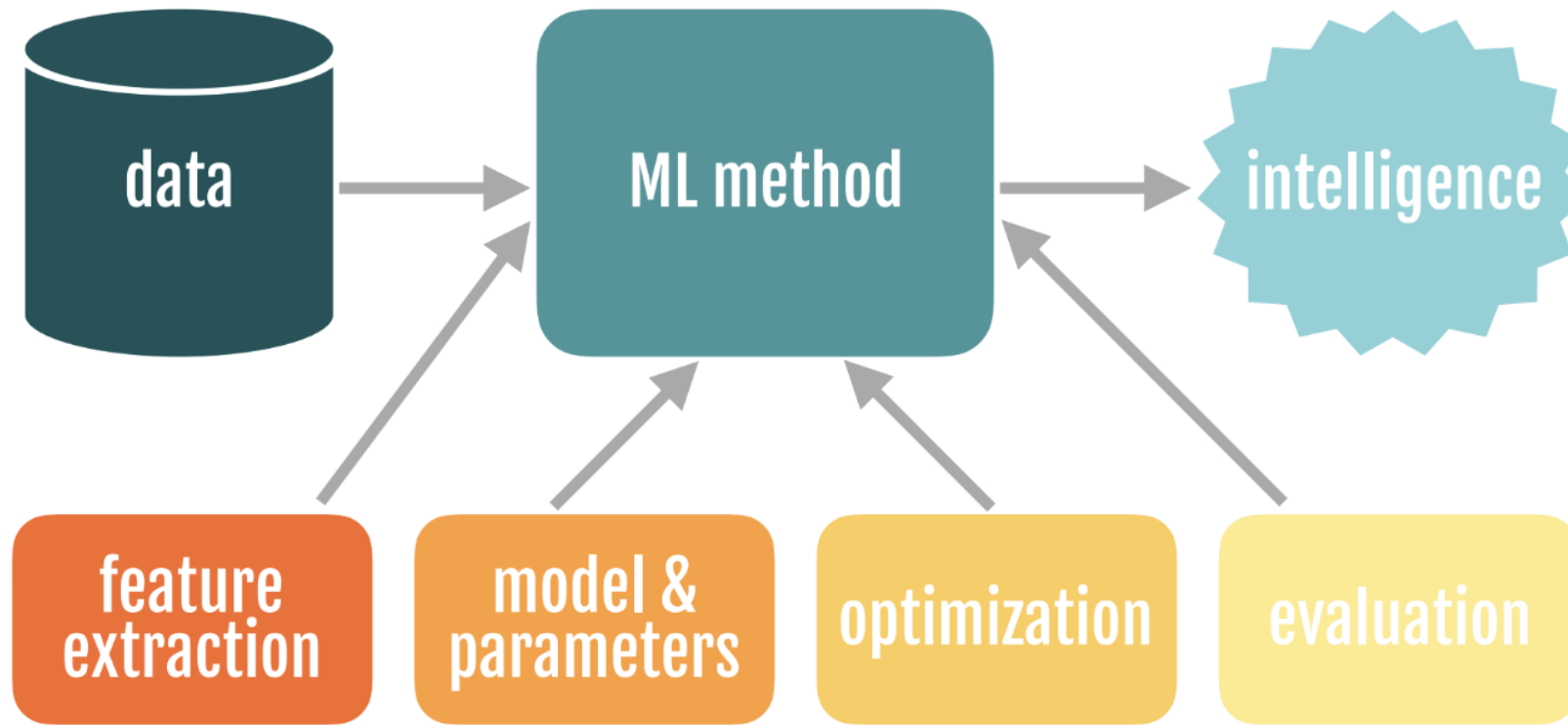
 Robotics

Other

 Graph Machine Learning



Machine Learning Pipeline





Outline

1. Overview of Deep Learning vs. Machine Learning
- 2. Applications of Some Models**
3. What You Need to Know Before Using These Models

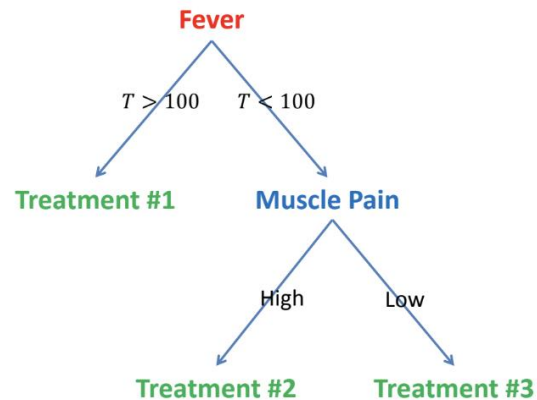
Decision Trees

Task: Multiclass Classification

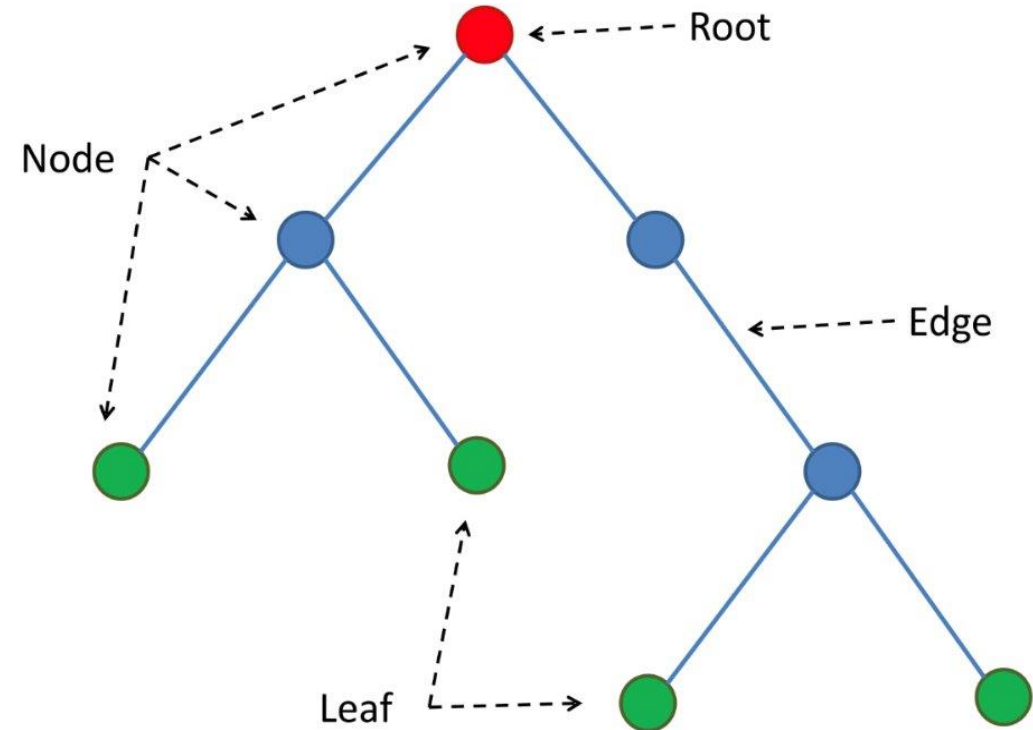
Advantages:

- Explain the reasoning in clear terms.

Medical treatment

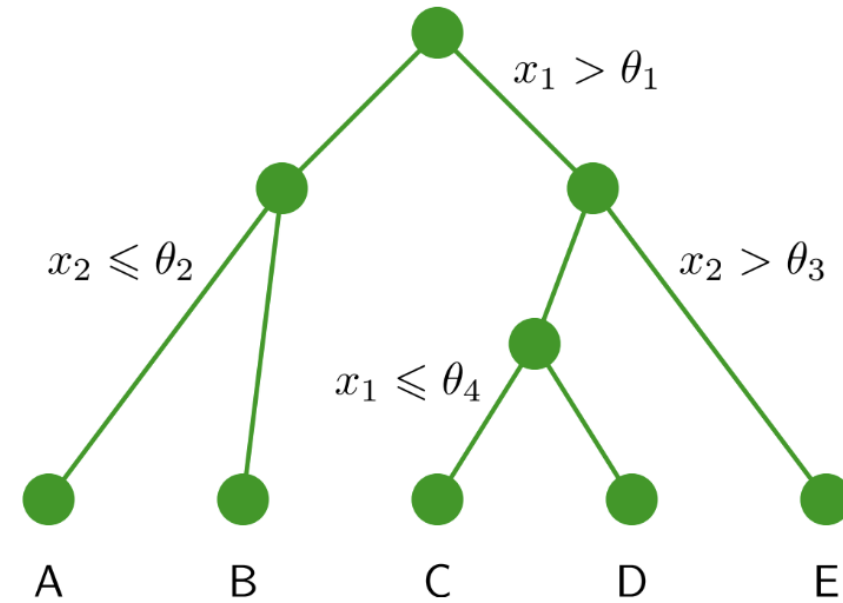
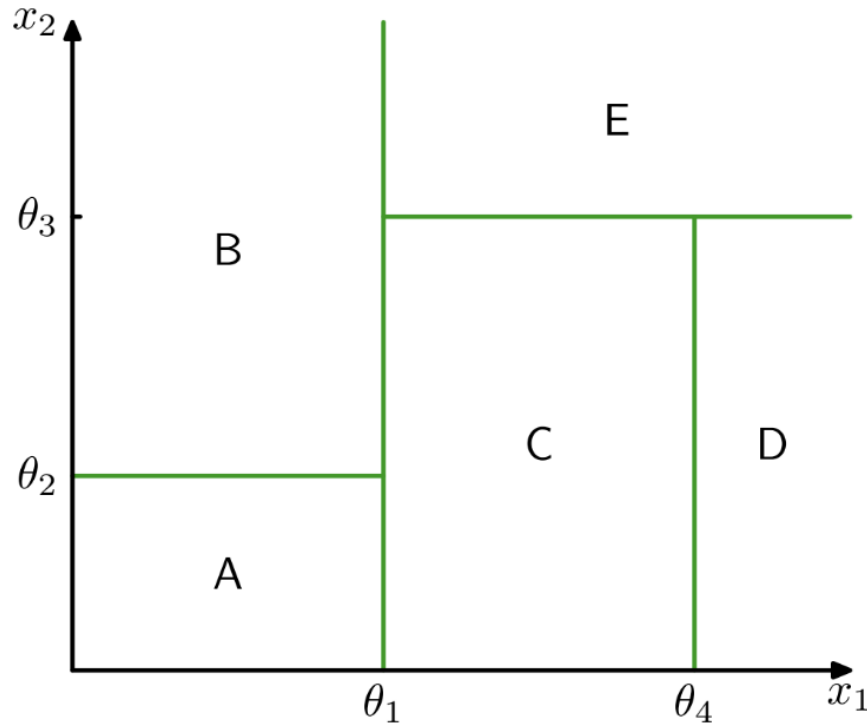


Model structure:



Decision Trees

Feature space partitioning



Decision Trees

Data example

Attributes										Target
<i>Alt</i>	<i>Bar</i>	<i>Fri</i>	<i>Hun</i>	<i>Pat</i>	<i>Price</i>	<i>Rain</i>	<i>Res</i>	<i>Type</i>	<i>Est</i>	<i>WillWait</i>
<i>T</i>	<i>F</i>	<i>F</i>	<i>T</i>	<i>Some</i>	<i>\$\$\$</i>	<i>F</i>	<i>T</i>	<i>French</i>	<i>0-10</i>	<i>T</i>
<i>T</i>	<i>F</i>	<i>F</i>	<i>T</i>	<i>Full</i>	<i>\$</i>	<i>F</i>	<i>F</i>	<i>Thai</i>	<i>30-60</i>	<i>F</i>
<i>F</i>	<i>T</i>	<i>F</i>	<i>F</i>	<i>Some</i>	<i>\$</i>	<i>F</i>	<i>F</i>	<i>Burger</i>	<i>0-10</i>	<i>T</i>
<i>T</i>	<i>F</i>	<i>T</i>	<i>T</i>	<i>Full</i>	<i>\$</i>	<i>F</i>	<i>F</i>	<i>Thai</i>	<i>10-30</i>	<i>T</i>
<i>T</i>	<i>F</i>	<i>T</i>	<i>F</i>	<i>Full</i>	<i>\$\$\$</i>	<i>F</i>	<i>T</i>	<i>French</i>	<i>>60</i>	<i>F</i>
<i>F</i>	<i>T</i>	<i>F</i>	<i>T</i>	<i>Some</i>	<i>\$\$</i>	<i>T</i>	<i>T</i>	<i>Italian</i>	<i>0-10</i>	<i>T</i>
<i>F</i>	<i>T</i>	<i>F</i>	<i>F</i>	<i>None</i>	<i>\$</i>	<i>T</i>	<i>F</i>	<i>Burger</i>	<i>0-10</i>	<i>F</i>
<i>F</i>	<i>F</i>	<i>F</i>	<i>T</i>	<i>Some</i>	<i>\$\$</i>	<i>T</i>	<i>T</i>	<i>Thai</i>	<i>0-10</i>	<i>T</i>
<i>F</i>	<i>T</i>	<i>T</i>	<i>F</i>	<i>Full</i>	<i>\$</i>	<i>T</i>	<i>F</i>	<i>Burger</i>	<i>>60</i>	<i>F</i>
<i>T</i>	<i>T</i>	<i>T</i>	<i>T</i>	<i>Full</i>	<i>\$\$\$</i>	<i>F</i>	<i>T</i>	<i>Italian</i>	<i>10-30</i>	<i>F</i>
<i>F</i>	<i>F</i>	<i>F</i>	<i>F</i>	<i>None</i>	<i>\$</i>	<i>F</i>	<i>F</i>	<i>Thai</i>	<i>0-10</i>	<i>F</i>
<i>T</i>	<i>T</i>	<i>T</i>	<i>T</i>	<i>Full</i>	<i>\$</i>	<i>F</i>	<i>F</i>	<i>Burger</i>	<i>30-60</i>	<i>T</i>

Task

- Decide whether to wait or not to wait for a service (e.g. in a restaurant).



Decision Trees

Advantages

- Can be interpreted by humans (as long as the tree is not too big)
- Computationally efficient (for shallow trees)
- Handles both numerical and categorical features.
- Can be used for both classification and regression

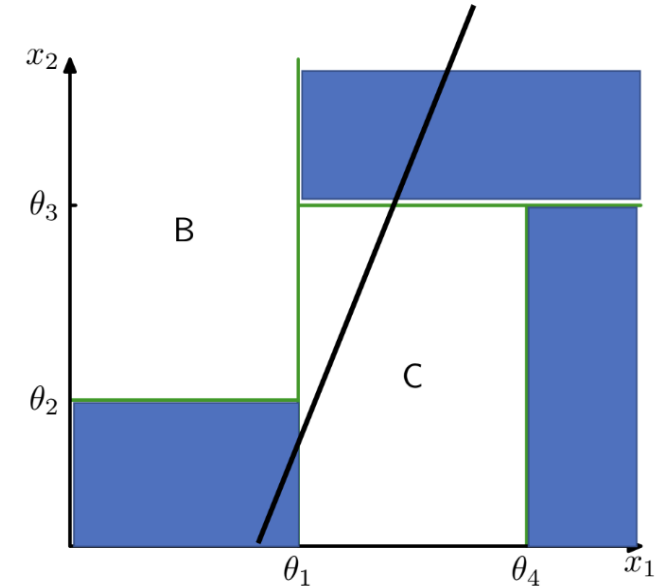
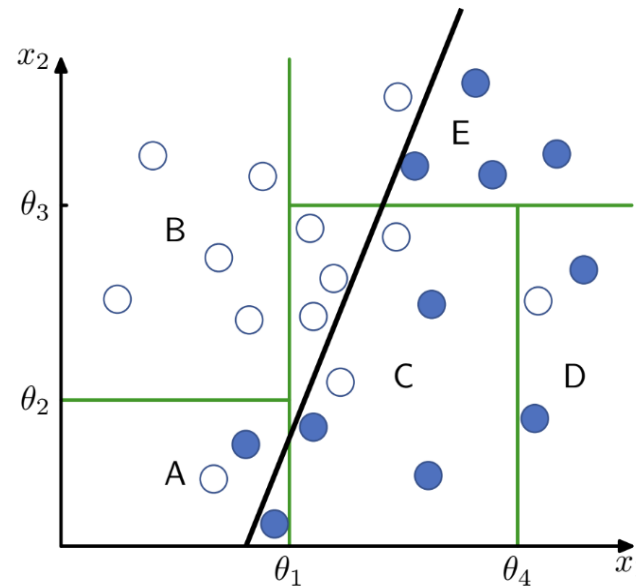
Decision Trees

Advantages

- Can be interpreted by humans (as long as the tree is not too big)
- Computationally efficient (for shallow trees)
- Handles both numerical and categorical features.
- Can be used for both classification and regression

Disadvantages

- Binary decision trees find it hard to learn linear boundaries.
- Decision trees are prone to overfitting!





Decision Trees

Advantages

- Can be interpreted by humans (as long as the tree is not too big)
- Computationally efficient (for shallow trees)
- Handles both numerical and categorical features.
- Can be used for both classification and regression

Disadvantages

- Binary decision trees find it hard to learn linear boundaries.
- Decision trees are prone to overfitting!

Strategies to avoid overfitting

- Stop growing when data split is not statistically significant.
- Acquire more training data.
- Remove irrelevant attributes (manual process, not always possible).
- Grow full tree, then post-prune.
- **Use tree ensembles**



Decision Trees

Advantages

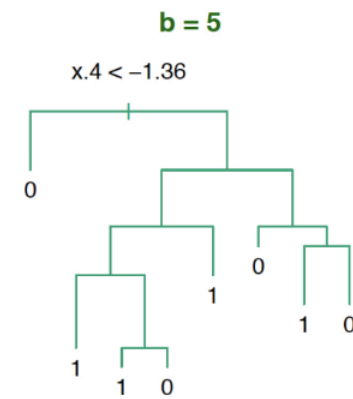
- Can be interpreted by humans (as long as the tree is not too big)
- Computationally efficient (for shallow trees)
- Handles both numerical and categorical features.
- Can be used for both classification and regression

Disadvantages

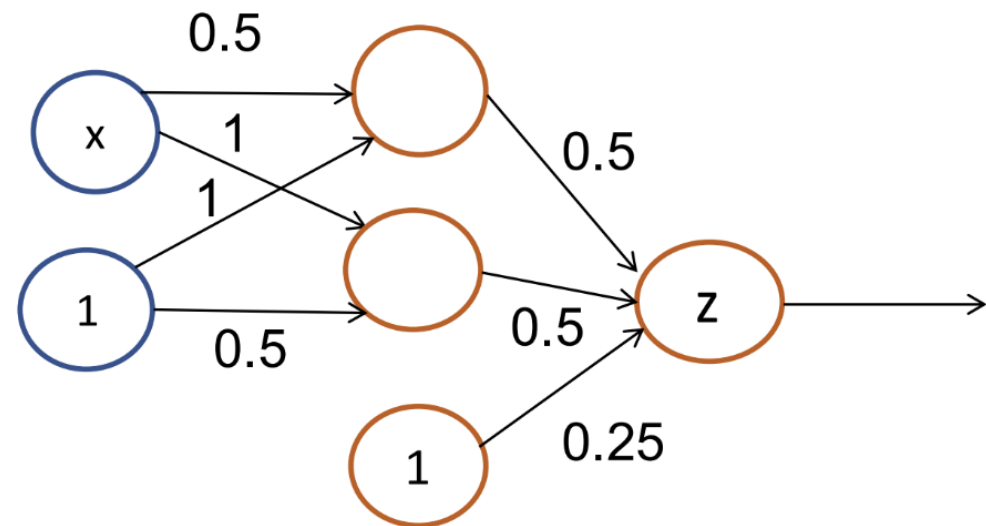
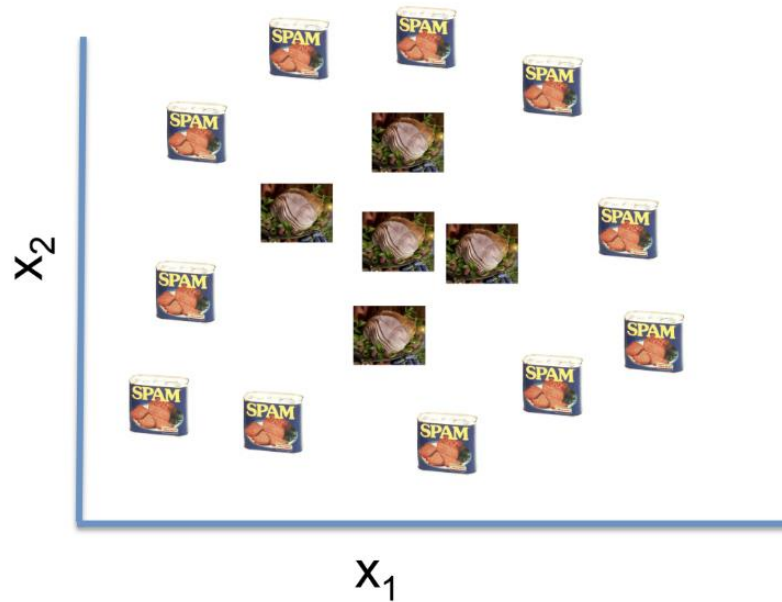
- Binary decision trees find it hard to learn linear boundaries.
- Decision trees are prone to overfitting!

Strategies to avoid overfitting

- Stop growing when data split is not statistically significant.
- Acquire more training data.
- Remove irrelevant attributes (manual process, not always possible).
- Grow full tree, then post-prune.
- **Use tree ensembles**



Neural Networks





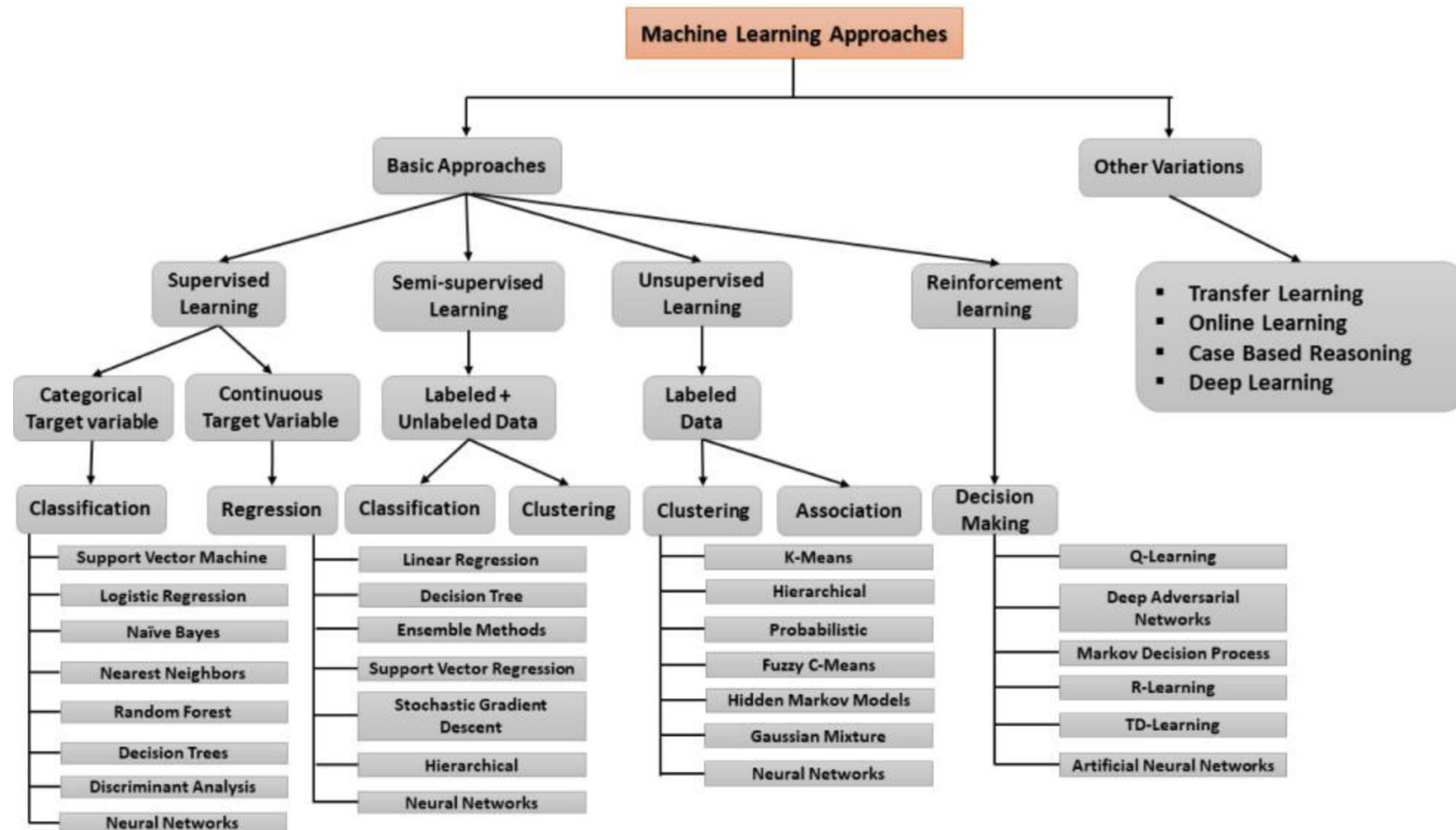
Hands-On ML and DL



Outline

1. Overview of Deep Learning vs. Machine Learning
2. Applications of Each Model
- 3. What You Need to Know Before Using These Models**

What You Need to Know Before Using These Models





What You Need to Know Before Using These Models

1. Identify your task
2. Characterize the type of data available (supervised vs. unsupervised)
3. **Identify approaches that are ideally suited for your task and your data.**
 - **Aim to use the simplest approach first.**



Q&A



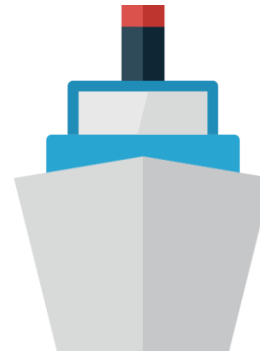
Hands-on Portion

Hands-on Portion

- Supervised Binary classification and Multiclass classification
- Jupyter notebook



Datasets:



Hands-on Portion

NCSA Delta

Go to: <https://openondemand.delta.ncsa.illinois.edu/>



Consent to Attribute Release

[Delta Open OnDemand](#) requests access to the following information. If you do not approve this request, do not proceed.

- Your CILogon user identifier
- Your name
- Your email address
- Your username and affiliation from your identity provider

Selected Identity Provider

National Center for Supercomputing Applications

☐ Remember this selection

Log On

By selecting "Log On", you agree to the [privacy policy](#).

Hands-on Portion

NCSA Delta

Go to: <https://openondemand.delta.ncsa.illinois.edu/>



NATIONAL CENTER FOR
SUPERCOMPUTING APPLICATIONS
UNIVERSITY OF ILLINOIS AT URBANA-CHAMPAIGN

NCSA Web Authentication

Log in to CILogon

NCSA Username

NCSA Kerberos Password

Login



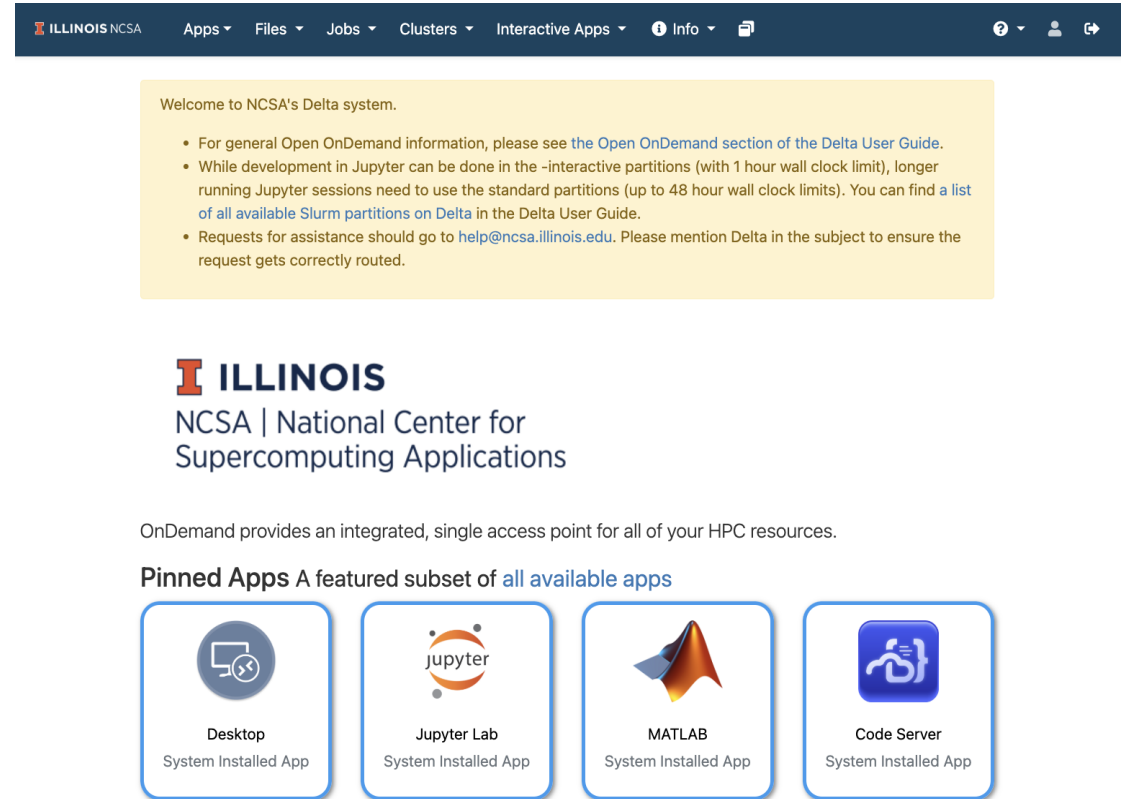
CILogon facilitates secure access to
CyberInfrastructure (CI).

- Enroll In Duo
- Forgot Your Username?
- Forgot Your Password?
- Send Email To Get Help

Hands-on Portion

NCSA Delta

Go to: <https://openondemand.delta.ncsa.illinois.edu/>



The screenshot shows the NCSA Delta Open OnDemand web interface. At the top is a dark blue navigation bar with the 'ILLINOIS NCSA' logo and menu items: Apps, Files, Jobs, Clusters, Interactive Apps, Info, and a user profile icon. Below the navigation bar is a yellow welcome message box that reads: 'Welcome to NCSA's Delta system.' followed by three bullet points: 'For general Open OnDemand information, please see the Open OnDemand section of the Delta User Guide.', 'While development in Jupyter can be done in the -interactive partitions (with 1 hour wall clock limit), longer running Jupyter sessions need to use the standard partitions (up to 48 hour wall clock limits). You can find a list of all available Slurm partitions on Delta in the Delta User Guide.', and 'Requests for assistance should go to help@ncsa.illinois.edu. Please mention Delta in the subject to ensure the request gets correctly routed.' Below the welcome message is the 'ILLINOIS NCSA | National Center for Supercomputing Applications' logo. A paragraph states: 'OnDemand provides an integrated, single access point for all of your HPC resources.' This is followed by a section titled 'Pinned Apps A featured subset of all available apps' which contains four app tiles: 'Desktop System Installed App' with a monitor icon, 'Jupyter Lab System Installed App' with the Jupyter logo, 'MATLAB System Installed App' with the MATLAB logo, and 'Code Server System Installed App' with a code editor icon.

ILLINOIS NCSA | National Center for Supercomputing Applications

OnDemand provides an integrated, single access point for all of your HPC resources.

Pinned Apps A featured subset of [all available apps](#)

- Desktop**
System Installed App
- Jupyter Lab**
System Installed App
- MATLAB**
System Installed App
- Code Server**
System Installed App


Hands-on Portion

NCSA Delta - Go to: <https://openondemand.delta.ncsa.illinois.edu/>


Home / My Interactive Sessions / Jupyter Lab

Interactive Apps


Jupyter

 Jupyter Lab


MATLAB


 MATLAB

Servers

 Code Server

Visualization

 TensorBoard

 Desktop

Jupyter Lab

This app will launch a Jupyter Lab server on one compute node.

Name of account

Chargeable account of the form abcd-delta-cpu or abcd-delta-gpu. Replace abcd with your allocation code.

Partition

Interactive partitions are limited to one hour.

Duration of job

Slurm format: DD-HH:MM:SS

Hands-on Portion

NCSA Delta - Go to: <https://openondemand.delta.ncsa.illinois.edu/>

Name of reservation (leave empty if none)

Number of CPUs

Amount of RAM

Use Slurm format, e.g. 4096M, 10G. If left blank, 1000 MB will be allocated per CPU core requested.

Number of GPUs

☒ I would like to receive an email when the session starts

Working Directory

Select your project directory; defaults to \$HOME

Select Path

Launch

* The Jupyter Lab session data for this session can be accessed under the [data root directory](#).

Hands-on Portion

NCSA Delta - Go to: <https://openondemand.delta.ncsa.illinois.edu/>

Jupyter Lab (8787041)

1 node | 1 core | Running

Host: `>_ gpub059.delta.internal.ncsa.edu`

✕ Delete

Created at: 2025-04-01 09:19:49 CDT

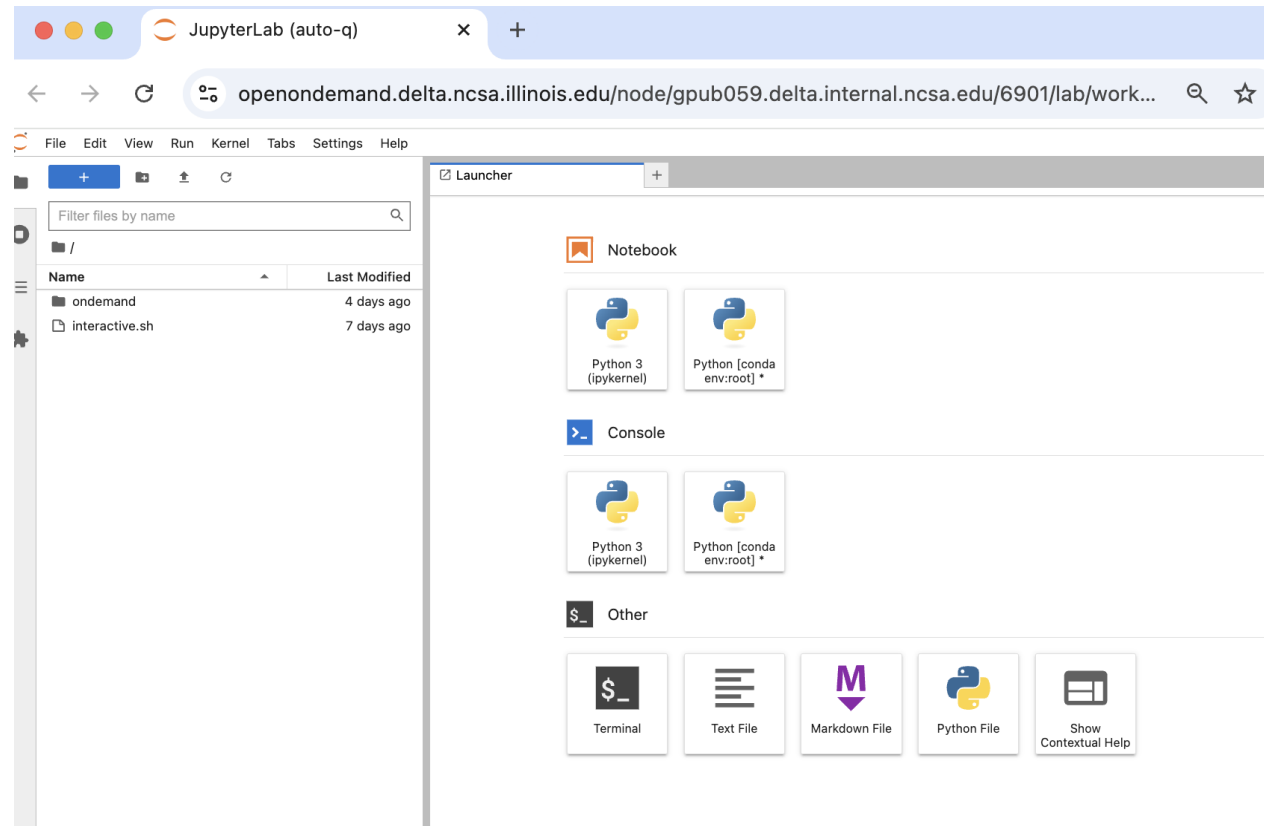
Time Remaining: 2 hours and 40 minutes

Session ID: `e5c75ba9-f6a0-403c-9929-91198795ac99`

Connect to Jupyter

Hands-on Portion

NCSA Delta - Go to: <https://openondemand.delta.ncsa.illinois.edu/>



Hands-on Portion

NCSA Delta - Go to: <https://openondemand.delta.ncsa.illinois.edu/>

The screenshot displays the NCSA Delta Open OnDemand web interface. On the left is a file browser sidebar with a search bar and a list of files in the '/ondemand/' directory. The main area shows a Jupyter notebook titled 'ml_model_comparison_training (1).ipynb'. The notebook content includes a title, a description, and three sections: '1. Setup', '2. Dataset Selection', and '3. Load Dataset', each with corresponding Python code cells.

File Browser:

Name	Last Modified
data	8 days ago
ml_model_comparison_training (1).ipynb	4 days ago
Untitled.ipynb	4 days ago

ML Classification: Comparing Decision Tree, Random Forest, and Neural Network

This notebook allows you to load one of four popular datasets, preprocess it, and train three classification models (Decision Tree, Random Forest, and a Neural Network) for comparison. It also includes training and inference timing, evaluation metrics, and visualizations.

1. Setup

```
[108]: import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt
import tensorflow as tf
import time
from sklearn.model_selection import train_test_split
from sklearn.preprocessing import LabelEncoder, StandardScaler
from sklearn.metrics import accuracy_score, precision_score, recall_score, f1_score, confusion_matrix, classification_report
from sklearn.tree import DecisionTreeClassifier, plot_tree
from sklearn.ensemble import RandomForestClassifier
from sklearn.tree import plot_tree
from tensorflow.keras import Sequential
from tensorflow.keras.layers import Dense
```

2. Dataset Selection

```
[135]: # Choose from: 'heart', 'wine', 'mushroom', 'titanic'
dataset_name = 'heart' # Change as needed
print('You selected the dataset: ' + dataset_name)

You selected the dataset: heart
```

3. Load Dataset

```
[136]: if dataset_name == 'heart':
df = pd.read_csv('https://raw.githubusercontent.com/sharmaroshan/Heart-UCI-Dataset/master/heart.csv')
```



Let's work with the Jupyter notebook!



Considerations

- Target Expanse and DeltaAI
- Prepare materials so they can run on their machine
 - Environment
 - Download in advance.
 - Jupyter notebooks with solutions.
 - PDF.
- Share in the Slack channel.
 - TODO: Who is sending this message?
- Hands on exercises – even after the workshop the attendees can leverage the material.
 - Allocation will be active for 1 year.
- **Are we prepopulating the accounts with the material?**
- Resource available: ACCESS github.



Considerations

- Hands on exercises – even after the workshop the attendees can leverage the material.
 - Allocation will be active for 1 year.
- **Are we prepopulating the accounts with the material?**
- Resource available: ACCESS github.

Next steps:

- DONE: Create Slack channel for the session.
- Use ACCESS account to get access to Expanse (direct login on OOD).
- Ask Laurie to help us find a time to go through the hands-on portion.



Timeline

- Intro portion of the talk (30 mins)
- Hands on portion (50 mins)
 - 5 mins to show how to get the jupyter notebook up and running
 - Classification trees (10 mins)
 - 5 mins to look at results of classification
 - Random forest (10 mins)
 - 5 mins to look at results of Random forest
 - Neural networks (10 mins)
 - 5 mins to look at results of Neural networks
- Wrap up and some considerations & Q&A (10 mins)



Aptos or Arial 28 pt black - Heading

Aptos or Arial 18 pt black Text Body

Use black, blue and white color palette