A Crash Course in Practical ML

Day 1: Classical Methods

Gage DeZoort | October 27th, 2025

Prof. Andre Frankenthal
UCI Physics and Astro
asterenb@uci.edu

Dr. Gage DeZoort
Princeton ORFE / Physics
jdezoort@princeton.edu

Course Timeline

- **Day 1** (10/27, 5pm, ISEB 1010):
  - Linear regression, linear classification, clustering with K-means, clustering with DBScan

- **Day 2** (10/28, 5pm, Rowland Hall 160 + ISEB 4020)
  - Deep learning, introduction to neural networks (NNs)

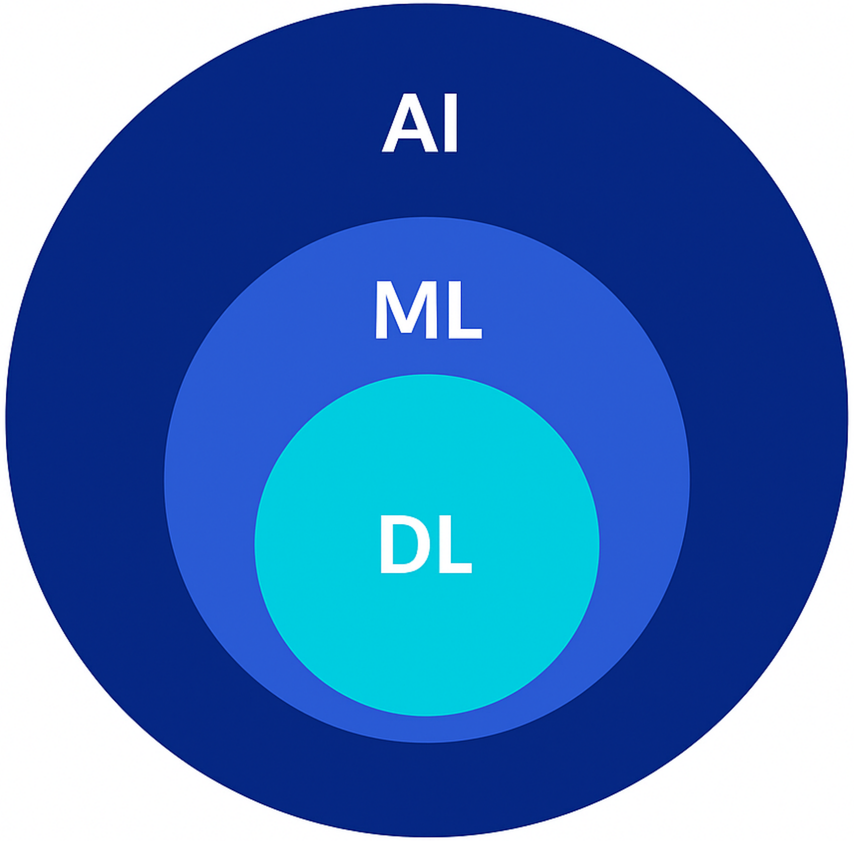
- **Day 3** (10/29, 5pm, ISEB 1010):
  - Convolutional NNs, Graph NNs, transformers and attention

- All materials will be posted in the course's Git repo:
  https://github.com/GageDeZoort/intro_ml_uci/tree/main

Course Format

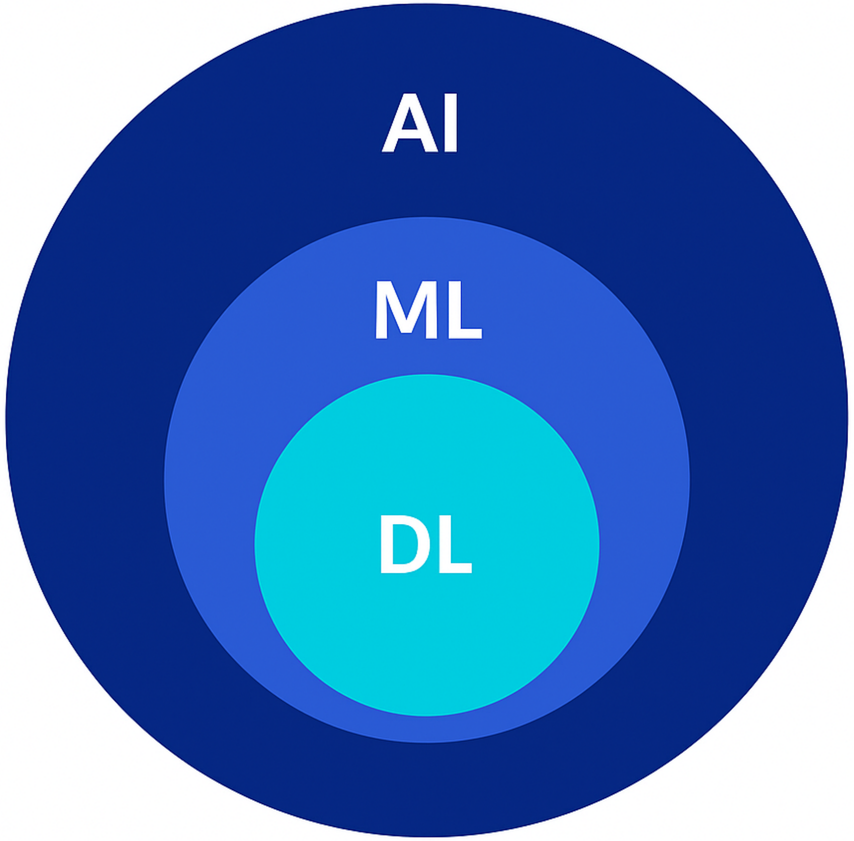
- We'll be using **Jupyter notebooks** in Google Colab: https://colab.research.google.com/
- Our notebooks will contain both text-based exposition and python code
  - Though we've scattered exercises throughout the notebooks, you can still participate without significant coding experience!
  - The exposition is designed to help you review the notebooks after the course; in real time, we'll talk through it

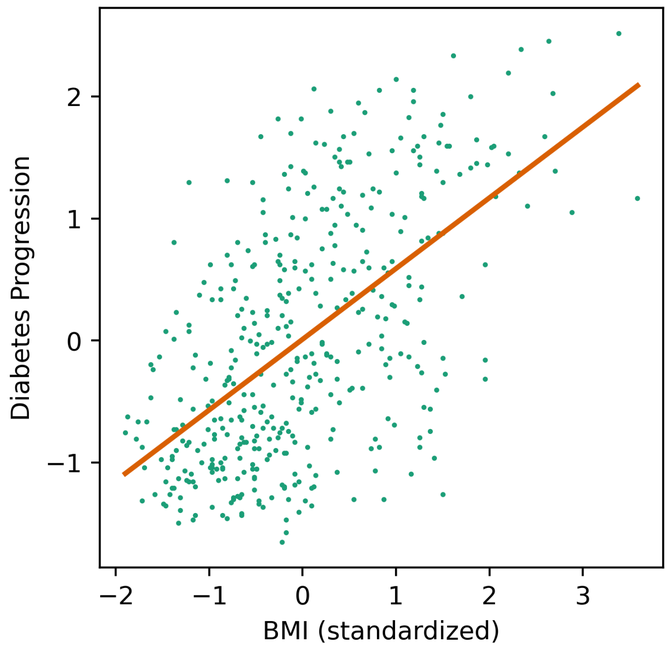
Artificial intelligence (AI), the broad endeavor to build systems that reason, plan, perceive, problem-solve, navigate etc.



***Machine Learning (ML)*, algorithms that *learn* (without explicit programming) to model relationships in data and use them to make predictions**

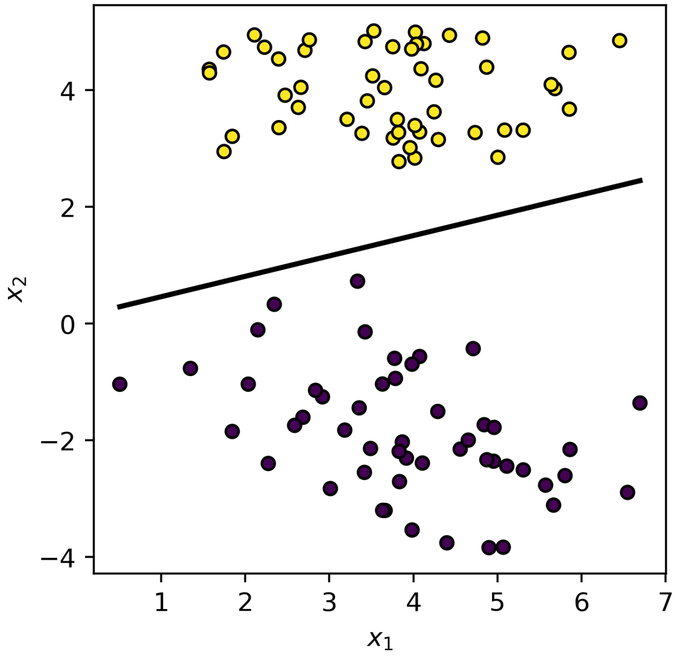
- *Pattern recognition* is key; something beyond a hard-coded set of rules
- Goal is not to memorize data, but to *generalize* to unseen inputs





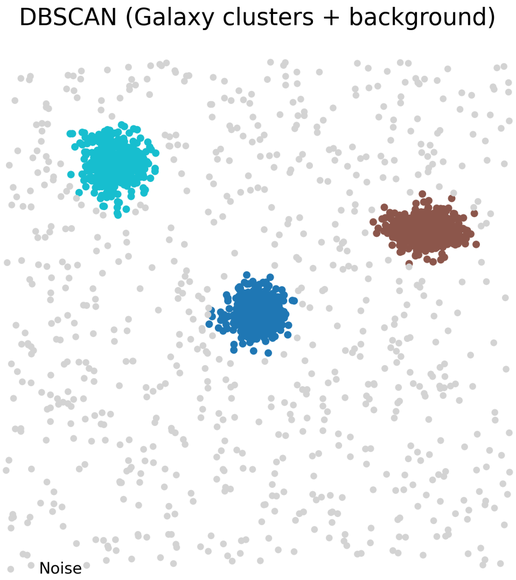
Regression

Predict input data values



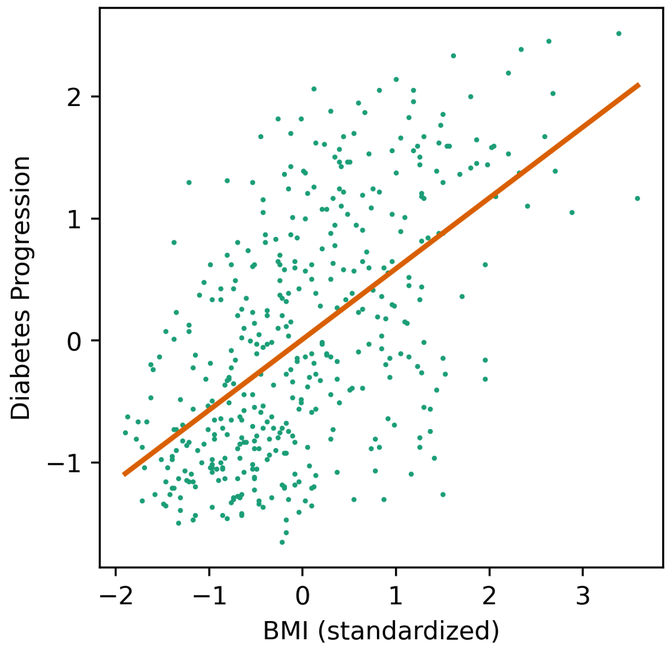
Classification

Predict input data labels



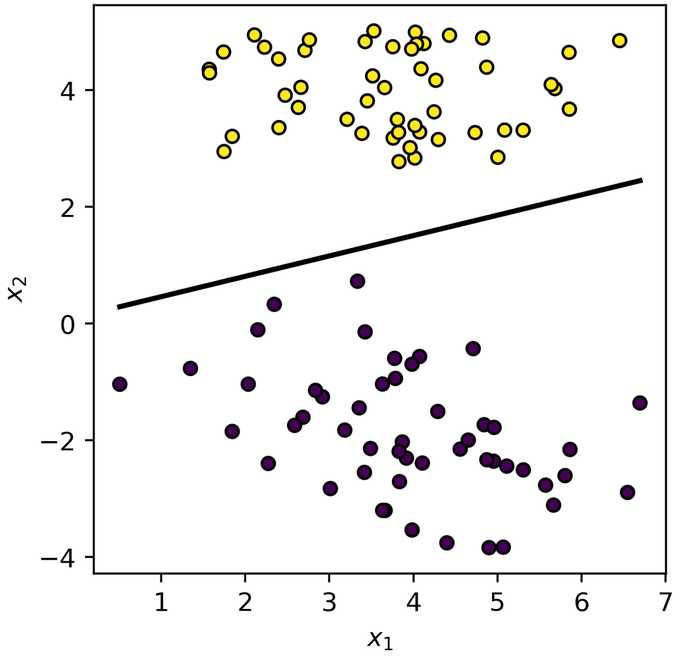
Clustering

Group input data



Regression

Predict input data values



Classification

Predict input data labels

Supervised Methods

Learn to model the relationship between inputs and outputs

Dataset $\{(\mathbf{x_i}, y_i)\}_{i=1}^{N}$

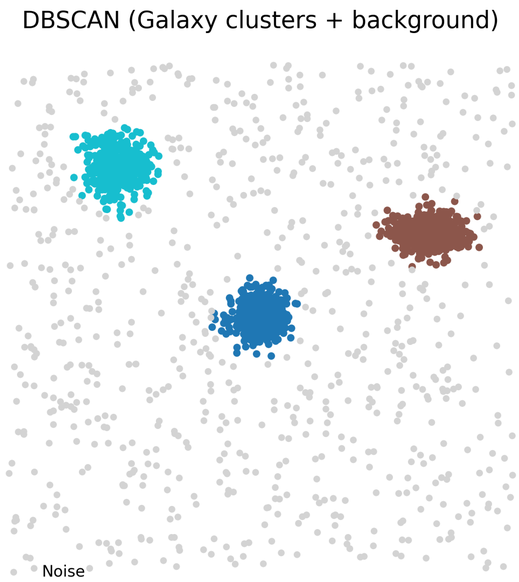
Every input $\mathbf{x}_i$ has a corresponding *truth target* y_i

Unsupervised Methods

Learn patterns and groupings without explicit targets

Dataset $\{\mathbf{x_i}\}_{i=1}^{N}$

Unlabeled data $\mathbf{x}_i$, goal is to find structure in it



Clustering

Group input data

Dataset Terminology

- **Inputs** $\mathbf{x}_i = [x_{i,1}, x_{i,2}, \ldots, x_{i,n}]$ are called **features**. These are the independent variables.

  Examples: height (continuous), $V0_2$ max (continuous), sex (categorical)

- **Outputs** y_i are called **truth targets**. These are the dependent variables.

  Examples: disease status (categorical), mile running time (continuous)

- **Model** predicts $\hat{y}_i$ for each input $\mathbf{x}_i$, which should match y_i

Training Supervised ML Algorithms

Select a model that produces predictions $\hat{y}_i$ for input i. The model will have **learnable parameters** or **weights** called $\boldsymbol{\theta}$. When we **train** the model, we are tuning these parameters to better fit the data. In this sense, our model is a function of our inputs, conditioned on the parameters we tune:

$$\hat{y}_i = \hat{y}(\mathbf{x}_i | \boldsymbol{\theta})$$

Split your data into a **train set** and a **test set**.

- Tune, or **train**, your model to fit the train set by minimizing a **loss function** describing how well $\hat{y}_i$ matches y_i
- Once your model is trained, test its **generalization** using the test set

https://github.com/GageDeZoort/intro_ml_uci/tree/main