LECTURE 1: COURSE OVERVIEW

Ehsan Aryafar
earyafar@pdx.edu
http://web.cecs.pdx.edu/~aryafare/ML.html

Introductions

- Who am I?
  - Joined PSU in 2017
  - Worked at Intel Labs in Santa Clara from 2013-2017 on 5G/5G+
  - Rice PhD 2011 and 2 years of PostDoc at Princeton University
  - Research in “computer and information science and engineering”
- Some statistics about you
  - Mix of undergraduate, Masters and PhD students
- Single TA to help with course and Python code examples

Outline

- What is machine learning
- ML categories
- Notation
- ML Pipeline
- Course Logistics

A breakthrough in ML would be worth 10 Microsofts

- Bill Gates, Microsoft Co-Founder

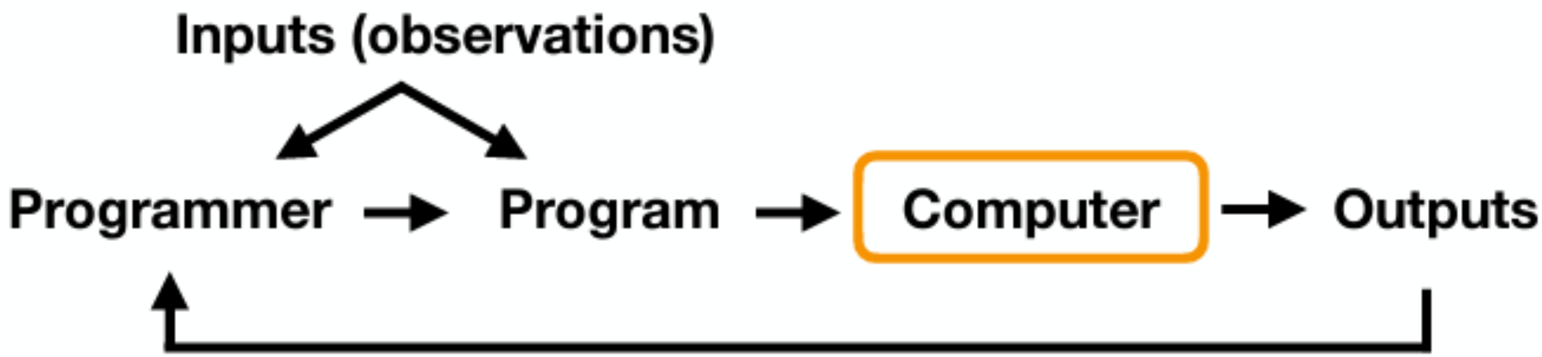
ML is a subcategory of CS, which allows you to implement AI. So it's kind of a mechanism to get you to AI.

- Rana el Kaliouby, CEO at Affectiva
- ML is generally considered a sub-field of AI

ML is the field of study that gives computers the ability to learn without being explicitly programmed

- Arthur L. Samuel, AI pioneer 1959
- Not exact quote but a paraphrased version

Traditional Programming Paradigm

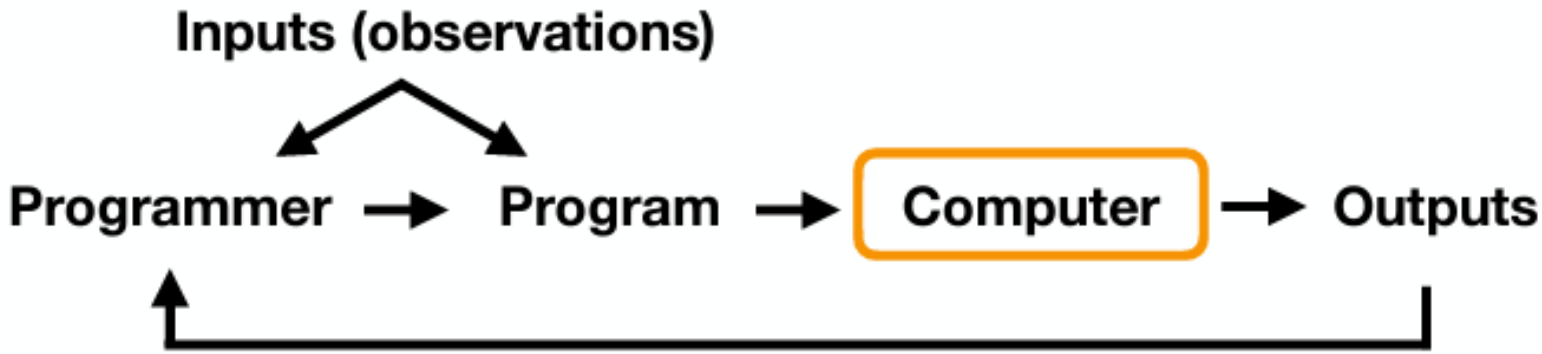


Create a program to filter spam emails (email program is given)

- Need to come up with some rules
- In above diagram, Input: emails, Program: Spam Filter, Outputs: Spam vs Not Spam
- Need to observe a huge amount of emails and see what makes an email spam
- Would come up with some rules (if ... else statements) to classify emails

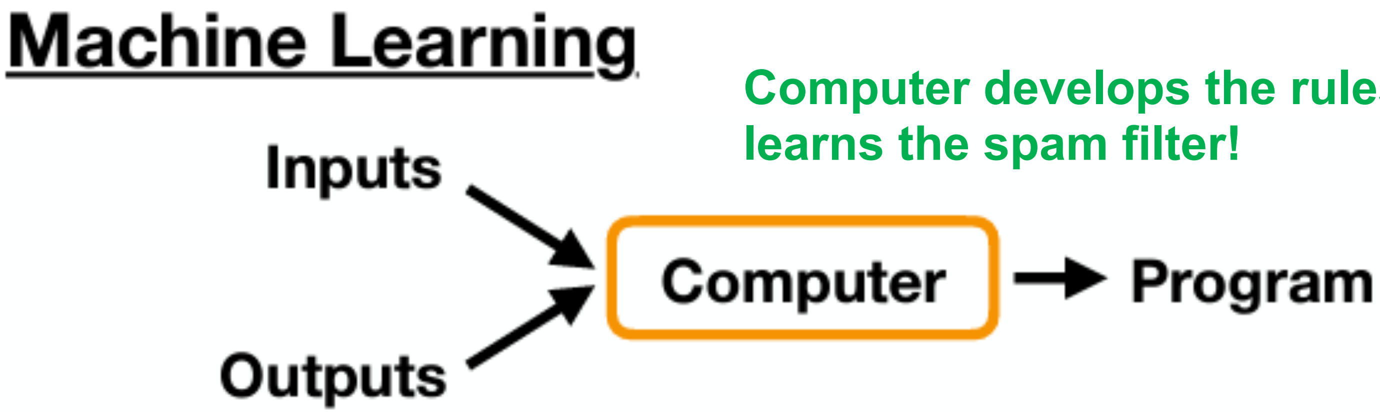
ML would help us achieve this goal much more conveniently in an automatic manner!

Programmer develops the rules (programs)



ML is the field of study that gives computers the ability to learn without being explicitly programmed

- Arthur L. Samuel, AI pioneer 1959



Computer develops the rules; i.e., learns the spam filter!

Assume a small dataset of **labeled** emails: spam/not spam is output and emails are input

Program would be the spam filter

We will not only use the machines for their intelligence, we will also collaborate with them in ways that we cannot even imagine.

- Fei Fei Li, Director of Stanford's AI lab

- ML is not really replacing all human labor

Back to Definition of ML

- A computer program is said to learn from experience E with respect to some class of tasks T and performance measure P, if its performance at tasks in T, as measured by P, improves with experience E.
  - Tom Mitchell, CMU
  - Experience in the case of ML could be learning from data

Handwriting Recognition Example

Examples of digits from the MNIST database (50000 handwritten digits, 5000 per digit)

- Task T:?
- Performance measure P:?
- Training experience E:?

Handwriting Recognition Example

Examples of digits from the MNIST database (50000 handwritten digits, 5000 per digit)

- Task T: handwritten digit classification (ML term for recognition)
- Performance measure P: False negative, false positive, precision
- Training experience E: 50000 images, labeled

Some ML Applications

- Email spam/not-spam classification
- Sorting letters in post office (Fedex uses ML to recognize hand-written zip codes)
- Face recognition on smartphone to lock/unlock screen
- Zillow house price prediction
- Other suggestions?

Outline

- What is machine learning
- ML categories
- Notation
- ML Pipeline
- Course Logistics

ML Categories

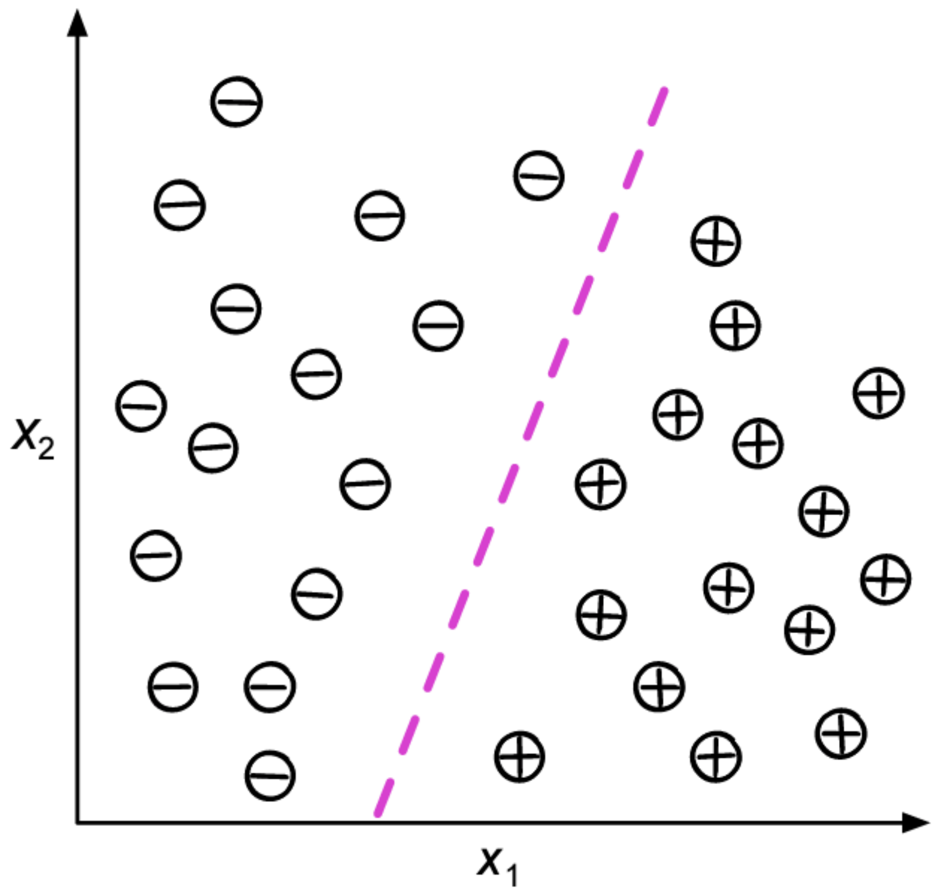
Supervised Learning

- Labeled data
- Direct feedback
- Predict outcome/future

- labeled data: dataset of emails and spam/not spam
- Predict outcome (spam or not spam) on future emails
- Direct feedback, if we make a wrong prediction we use that to improve and make a better classification

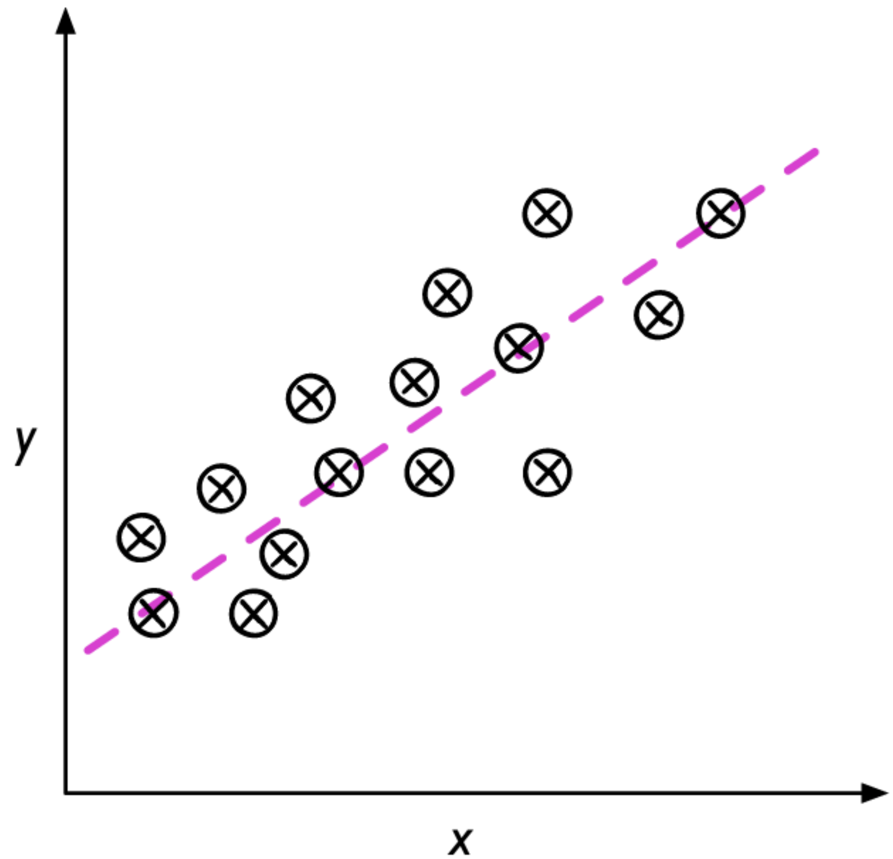
Supervised Learning: Classification

- Each data point has two feature values (x1 and x2) based on which we have to make a decision if the point is + or –
- Train ML model by showing it training dataset with labels
- ML algorithm learns to generate/draw a decision boundary (purple line)
- It you show a new data point, ML algorithm can now classify



Supervised Learning: Regression

- Regression aims at assigning a continuous value
- Target (y) in the previous example was + or – (binary classification task), here is a continuous number
- Here we have one input feature/variable
- A least square regression would try to find a line that minimizes the sum squared offsets (errors) on all dataset



ML Categories

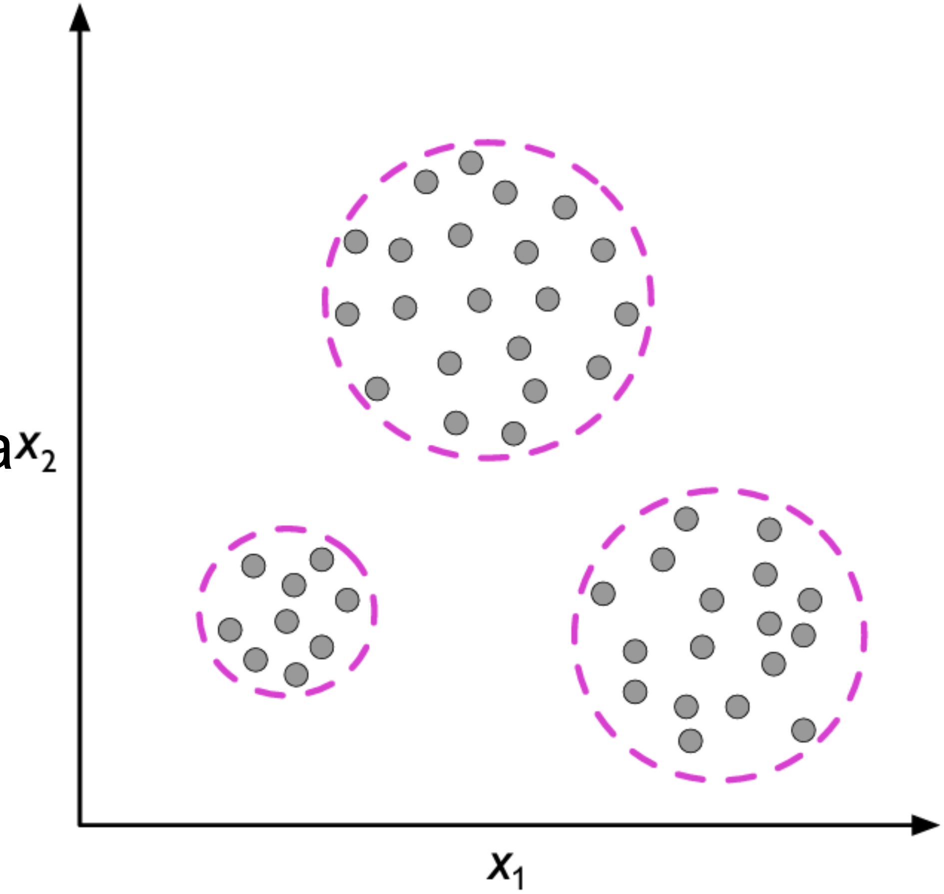
Unsupervised Learning

- No labels/targets
- No feedback
- Find hidden structure in data

- No feedback: we cannot measure how well the algorithm is doing
- Goal: find hidden structure

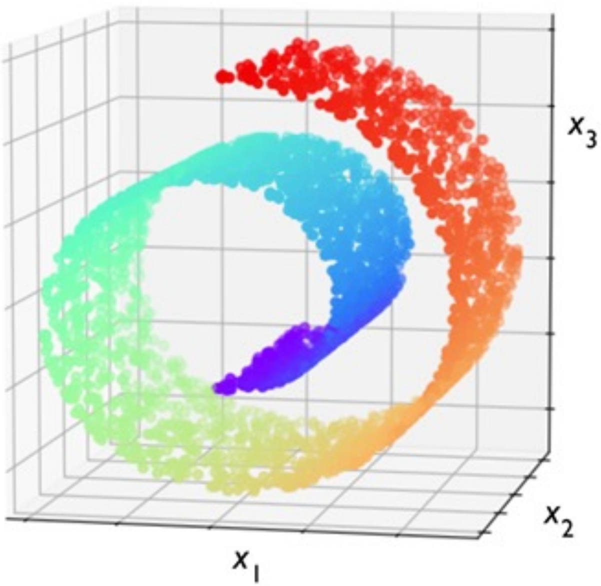
Unsupervised Learning: Clustering

- Consider a dataset with measurements x1 and x2
- No class or label information
- There are dots that are closer to one another (something about them that makes them related, called a cluster)

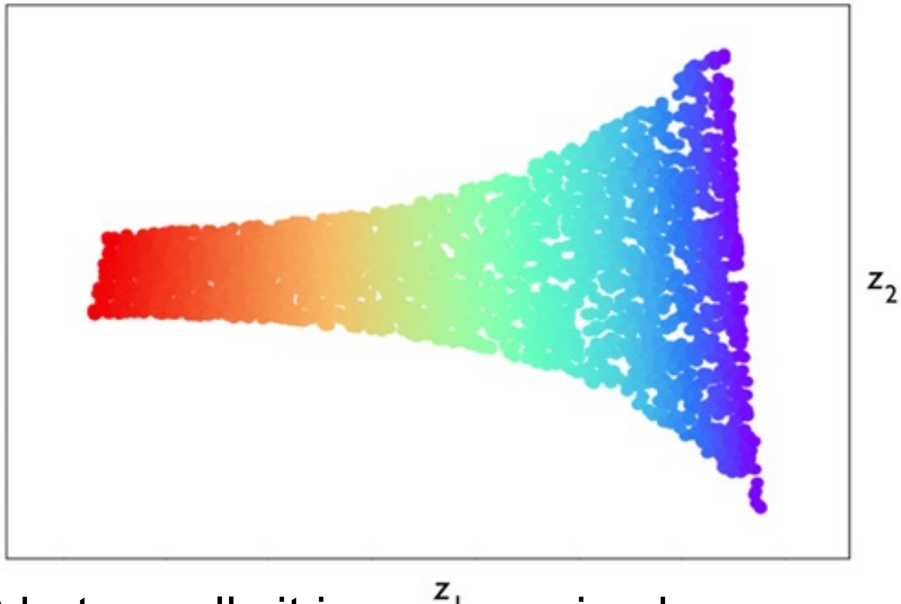


Unsupervised Learning: Dimensionality Reduction

Swiss Roll Dataset



Same colored points are still together



- There are supervised versions of DR but usually it is unsupervised
- We remove/reduce/compress some features (will have a dedicated lecture later)
- From [x1, x2, x3] -> [z1, z2] by dropping some features or combining/ or transforming them so that z captures most of the information
  - Feature extraction and selection
  - Also one of the goals of principle component analysis (PCA)

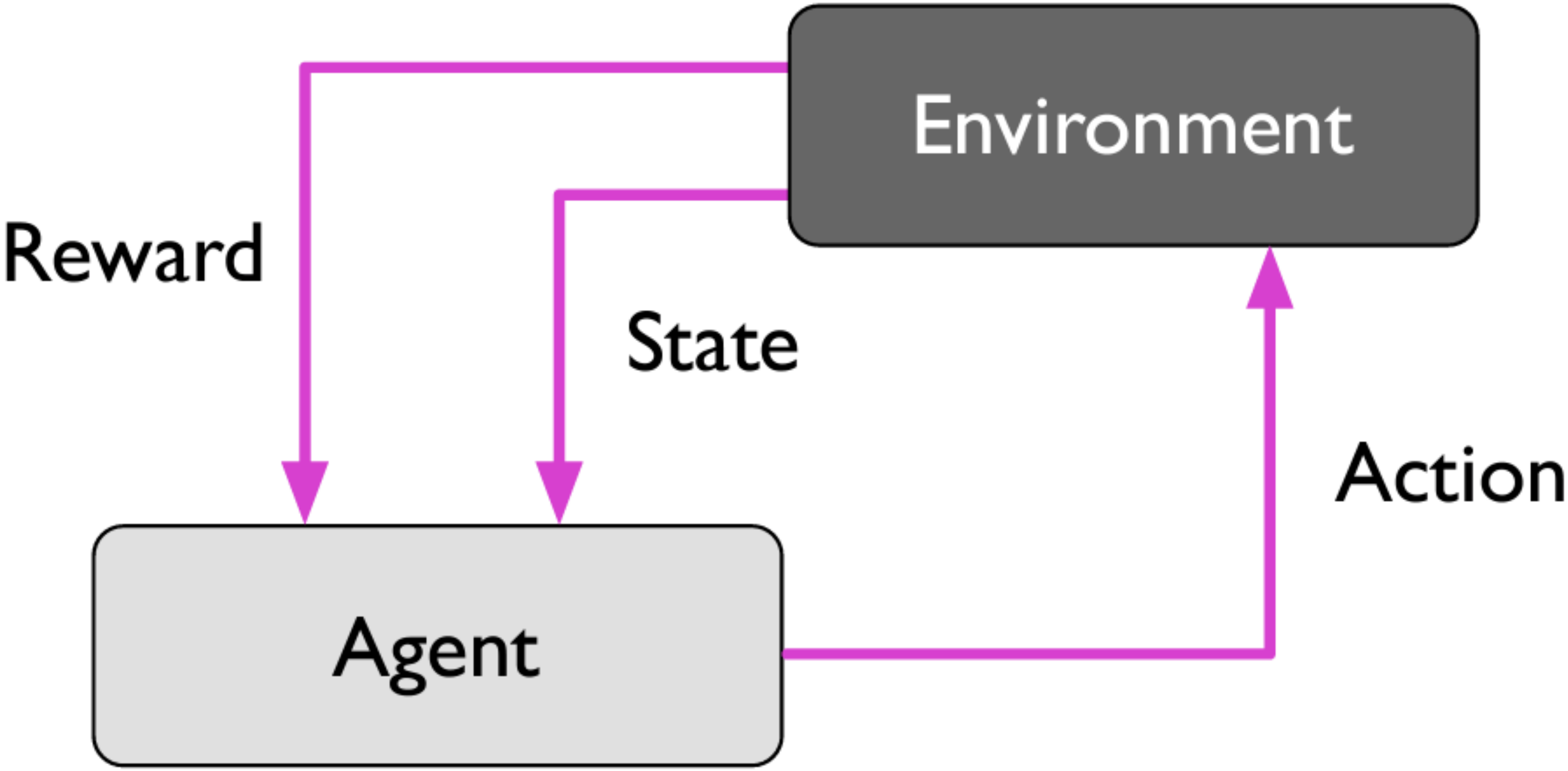
ML Categories

Reinforcement Learning

- Decision process
- Reward system
- Learn series of actions

- Sort of different from the other two
- Decision process based on taking a series of actions and getting rewards for them

Reinforcement Learning

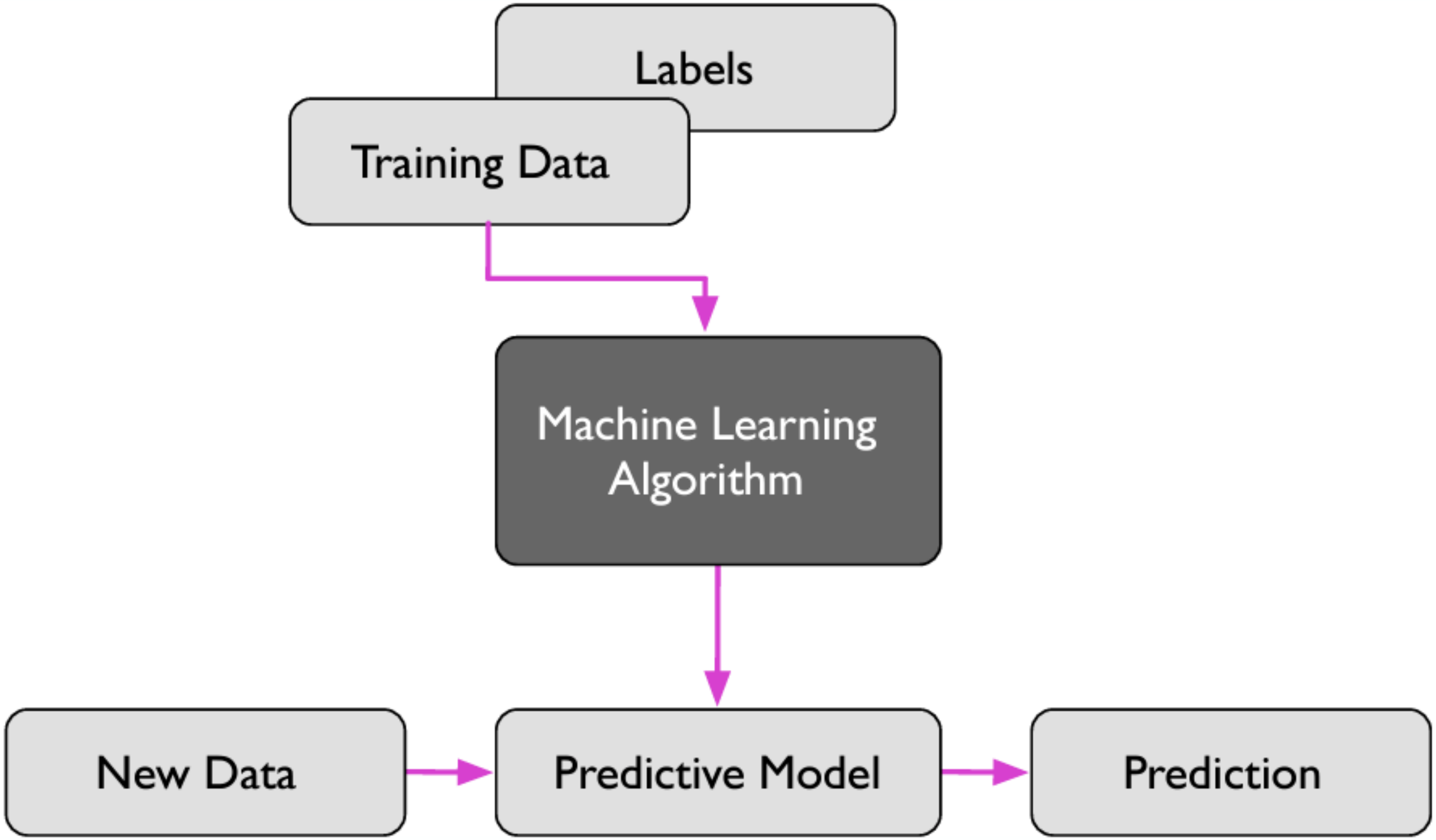


RL Example: DeepMind Playing Atari

Outline

- What is machine learning
- ML categories
- Notation
- ML Pipeline
- Course Logistics

Supervised Learning Workflow



Supervised Learning Notation

y: label

Training set: $\mathscr{D} = \{\langle \mathbf{x}^{[i]}, y^{[i]} \rangle, i = 1, \ldots , n\}$

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_m \end{bmatrix}$$

$$\mathbf{X} = \begin{bmatrix} \mathbf{x}_1^T \\ \mathbf{x}_2^T \\ \vdots \\ \mathbf{x}_n^T \end{bmatrix} = \begin{bmatrix} x_1^{[1]} & x_2^{[1]} & \cdots & x_m^{[1]} \\ x_1^{[2]} & x_2^{[2]} & \cdots & x_m^{[2]} \\ \vdots & \vdots & \ddots & \vdots \\ x_1^{[n]} & x_2^{[n]} & \cdots & x_m^{[n]} \end{bmatrix}$$

Feature vector

Dataset Samples

The goal is to find the function *f* that can best match *f*(**x**) to y

Example: Zillow House DataSet

	# of Rooms	SQFT	Zip Code	Price (label)
1	3	140	12321	$700k
2	3	230	18903	$350k
3	2	100	34501	$200k
		…..		
120	6	200	90213	$1M

Some ML Terminology

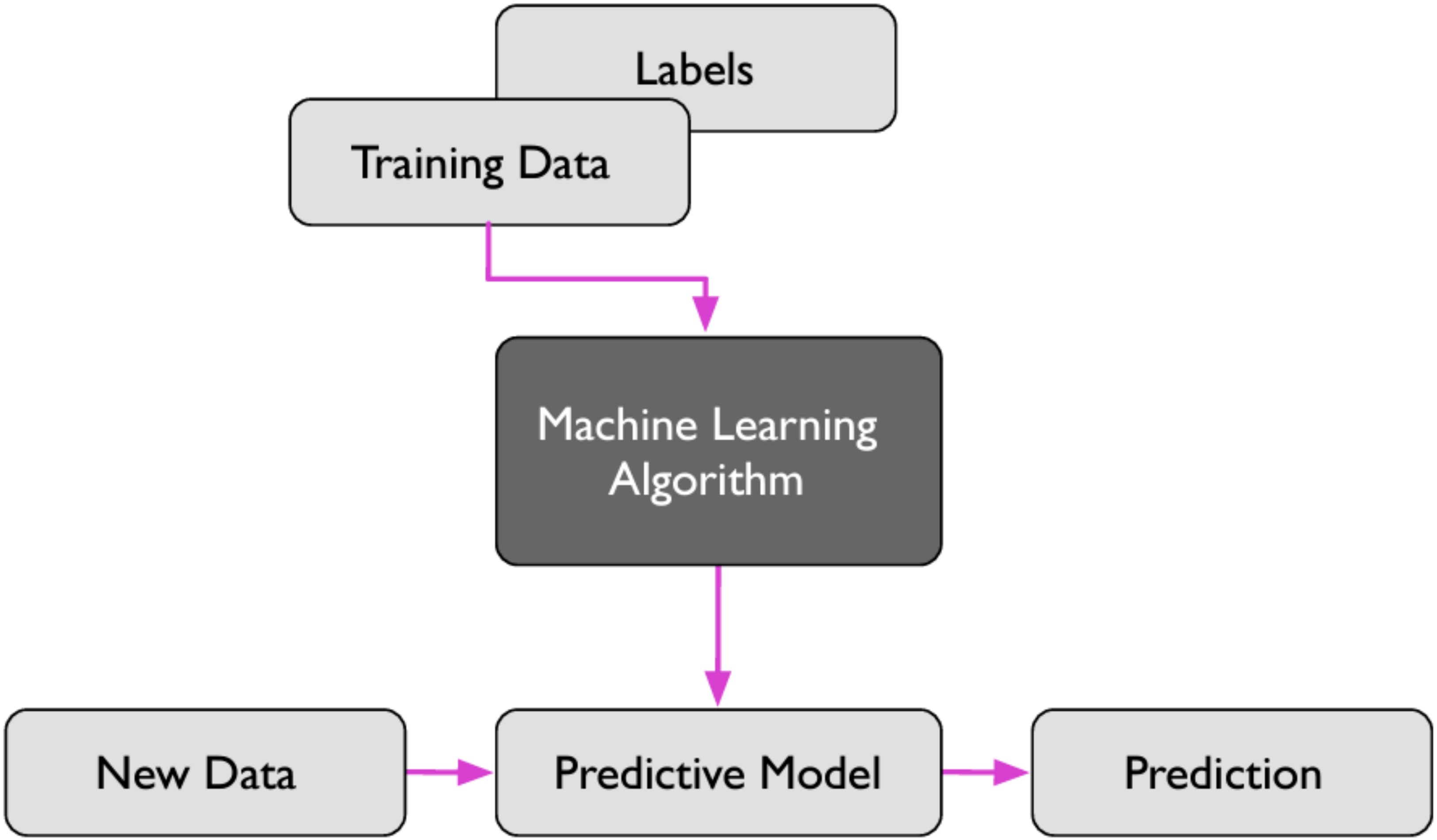
	# of Rooms	SQFT	Zip Code	Price (label)
1	3	140	12321	$700k
2	3	230	18903	$350k
3	2	100	34501	$200k
		…..		
120	6	200	90213	$1M

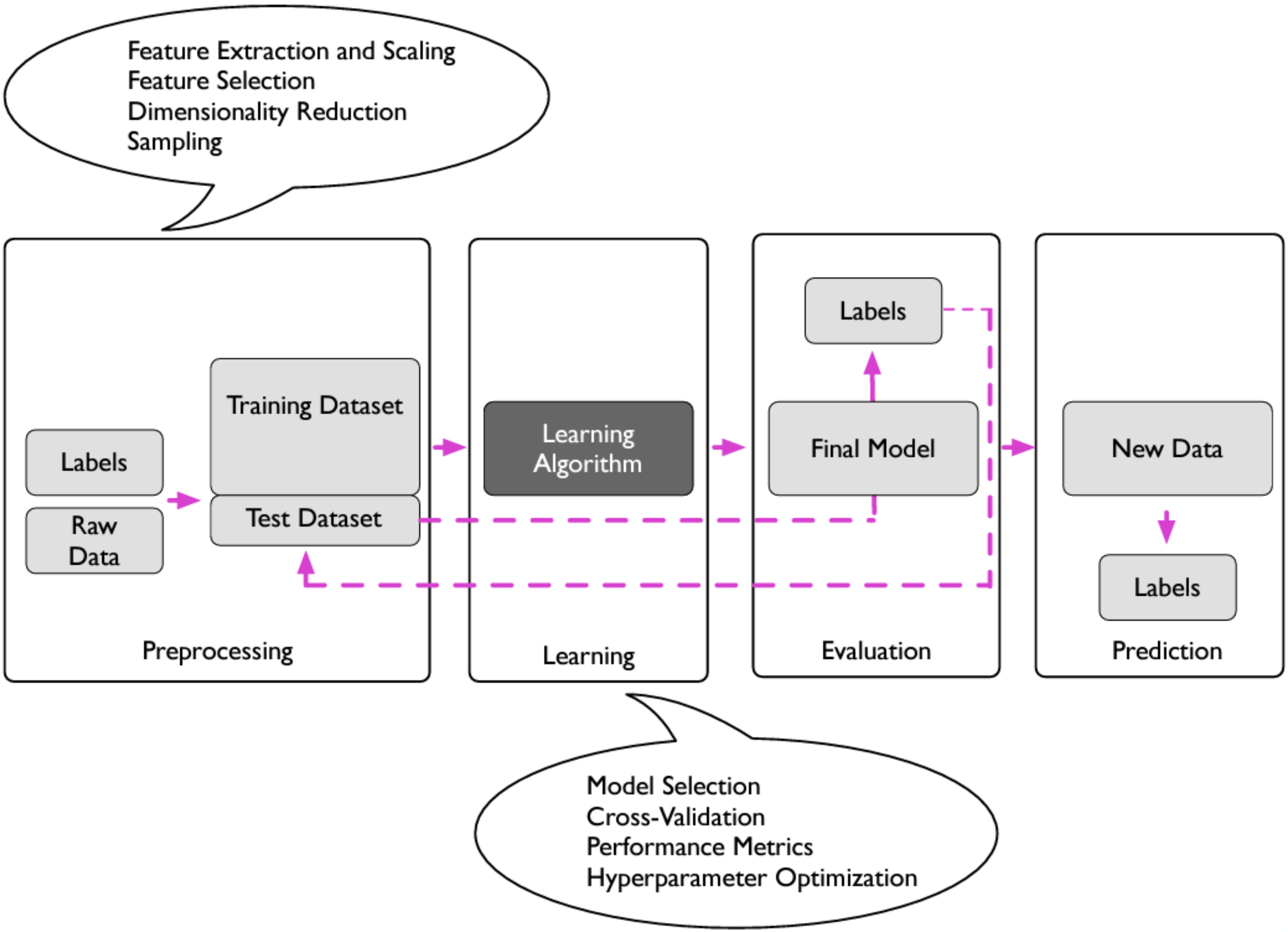
Training Example: a row in the table representing the dataset
Features: variables representing the dataset
Output (prediction): output of the model
Target: what we want to predict

Outline

- What is machine learning
- ML categories
- Notation
- ML Pipeline
- Course Logistics

Supervised Learning Workflow




Feature Extraction and Scaling
Feature Selection
Dimensionality Reduction
Sampling
Labels
Raw Data
Training Dataset
Test Dataset
Preprocessing
Learning Algorithm
Learning
Labels
Final Model
Evaluation
New Data
Labels
Prediction
Model Selection
Cross-Validation
Performance Metrics
Hyperparameter Optimization

5 Steps for Approaching an ML Application

- Define the problem to be solved
- Collect (labeled) data
- Choose an algorithm class
- Choose an optimization metric for learning the model
- Choose a metric or measure for evaluating the model
  - Could be the same as for learning the model

Example Metric: Misclassification Error or Loss Function

Model prediction

True Outcome

$$L(\hat{y}, y) = \begin{cases} 0 & \textbf{if } \hat{y} = y \\ 1 & \textbf{if } \hat{y} \neq y \end{cases}$$

$$ERR_{\mathscr{D}_{\textbf{test}}} = \frac{1}{n} \sum_{i=1}^{n} L\left(\hat{y}^{[i]}, y^{[i]}\right)$$

Some of the Key Topics

Models

- Linear Regression
- Linear classification: logistic regression, SVM
- Nonlinear models: kernels, neural networks & deep learning, decision trees
- Nearest neighbors, clustering

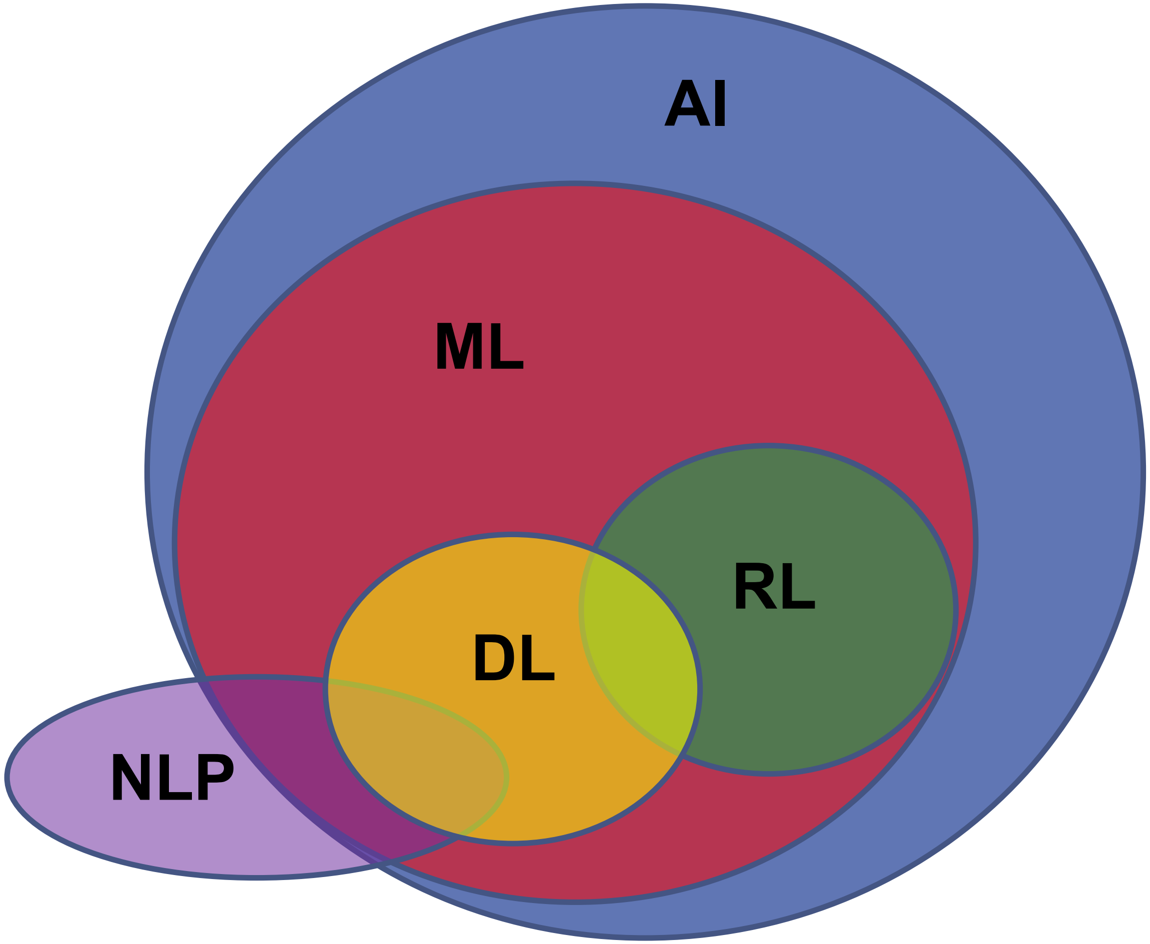
Methods

- Gradient descent
- Boosting
- k-means
- PCA
- EM

Concepts

- Point estimation, MLE, MAP
- Loss functions, bias-variance tradeoff, cross-validation
- Sparsity, overfitting, model selection
- Types of ML (supervised, unsupervised, reinforcement)

ML, AI, DL, RL, NLP



AI: a non-biological system that is intelligent through rules

ML: Algorithms that learn rules or models automatically from data

DL: Multiple layers of neurons that learn data representation

NLP: Builds machines that respond to text or voice data similar to humans

RL: machines make sequence of decisions to maximize reward

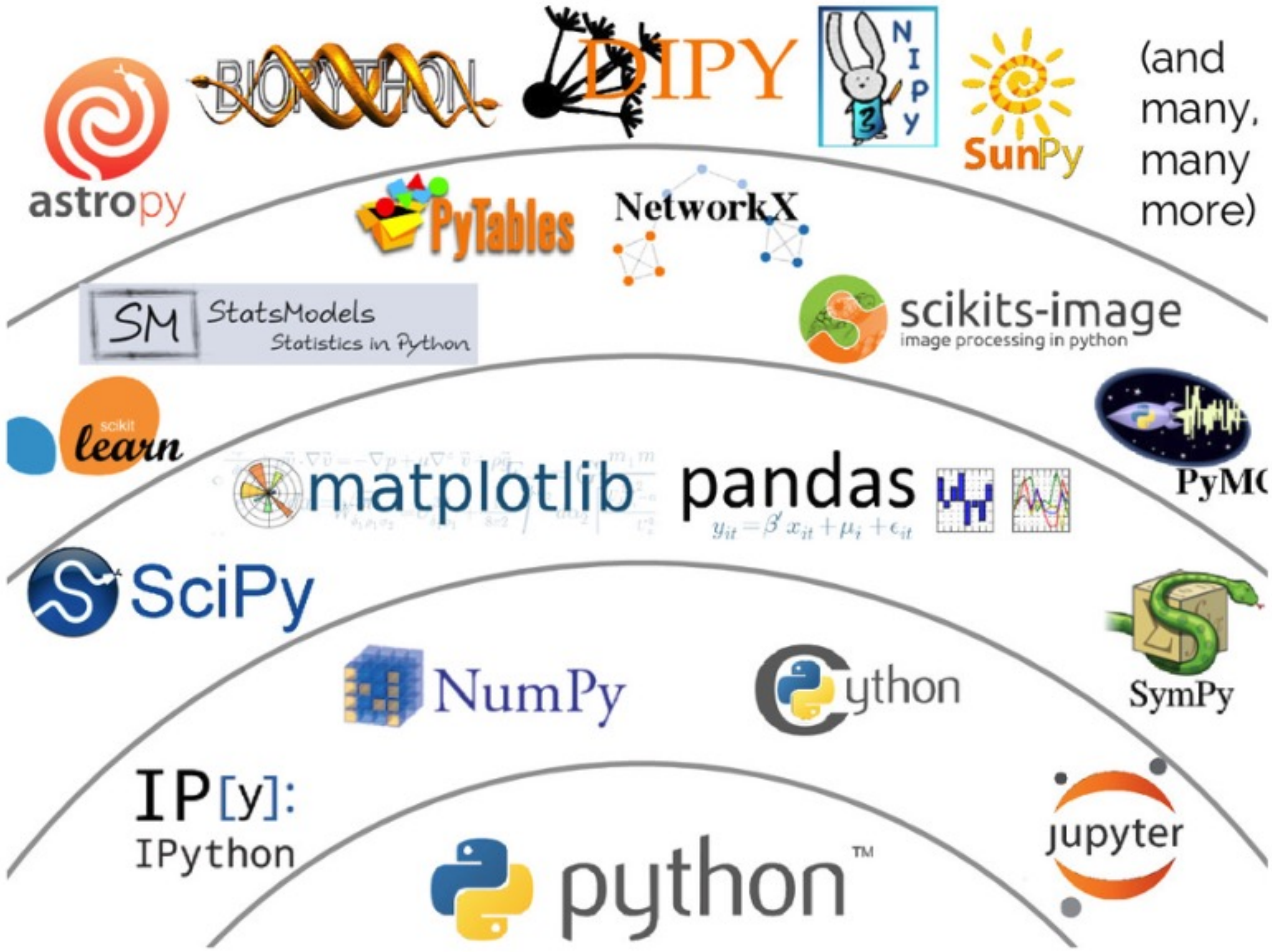


Image by Jake VanderPlas; Source:
https://speakerdeck.com/jakevdp/the-state-of-the-stack-scipy-2015-keynote?slide=8)

Outline

- What is machine learning
- ML categories
- Notation
- ML Pipeline
- Course Logistics

Course Structure:

- **Lectures:** 19 lectures covered by the instructor and TA
- **Exam:** One exam day, covering the lecture material

Detailed schedule is available on canvas.pdx.edu

Course Evaluation I

- There will be four homeworks covering the lecture material and coding assignments based on Python
  - 60% of your total grade
  - We use Jupyter notebooks for all homeworks

- Exam
  - There will be a single exam. The exam will be **based on all the lecture material.** The exam will **be open notes and open books.**
  - 30% of your total grade

Course Evaluation II

- Participation is 10% of your grade
  - Need to inform instructor if you have missed more than four sessions
  - Need to inform the instructor if you are taking the class mostly in an asynchronous manner
  - Need to respect your classmates with respect at all times

Course Evaluation Summary

- Homework: 60% of Total Grade
- Exam: 30% of Total Grade
- Participation: 10% of Total Grade