Chapter 1

ML Introduction

August 24, 2016

Math/stats background

- Partial derivatives, e.g.:

$$\frac{\partial(x^3 + y^2 + 1)}{\partial x}$$

- Matrix and vector operations

$$\begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix} \times \begin{bmatrix} 7 \\ 8 \\ 9 \end{bmatrix} = \begin{bmatrix} 1 \times 7 + 2 \times 8 + 3 \times 9 \\ 4 \times 7 + 5 \times 8 + 6 \times 9 \end{bmatrix} = \begin{bmatrix} 50 \\ 122 \end{bmatrix}$$

- Basic probability and statistics
  - Conditional probability
  - Normal distribution

Programming background

- Python
- NumPy
- Matplotlib

What is ML? Ask Wikipedia

Machine learning is a subfield of computer science (more particularly soft computing) that evolved from the study of pattern recognition and computational learning theory in artificial intelligence. In 1959, Arthur Samuel defined machine learning as a "Field of study that gives computers the ability to learn without being explicitly programmed". Machine learning explores the study and construction of algorithms that can learn from and make predictions on data.

More definitions here

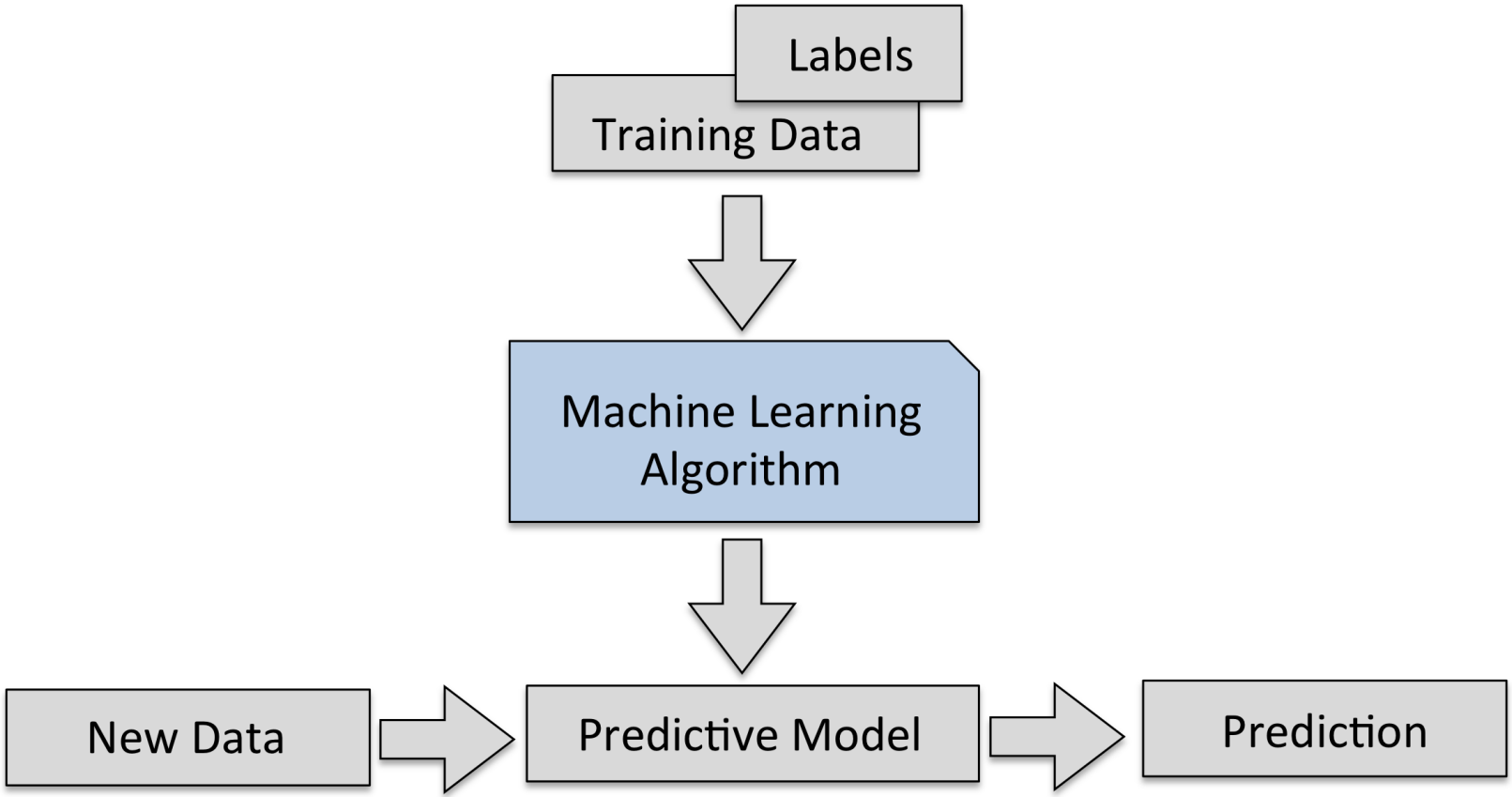
Applications of ML

- Image recognition
- Spam classification
- Web search engines
- Voice recognition
- Link to Quora

Three types of ML

- Supervised learning
- Unsupervised learning
- Reinforcement learning

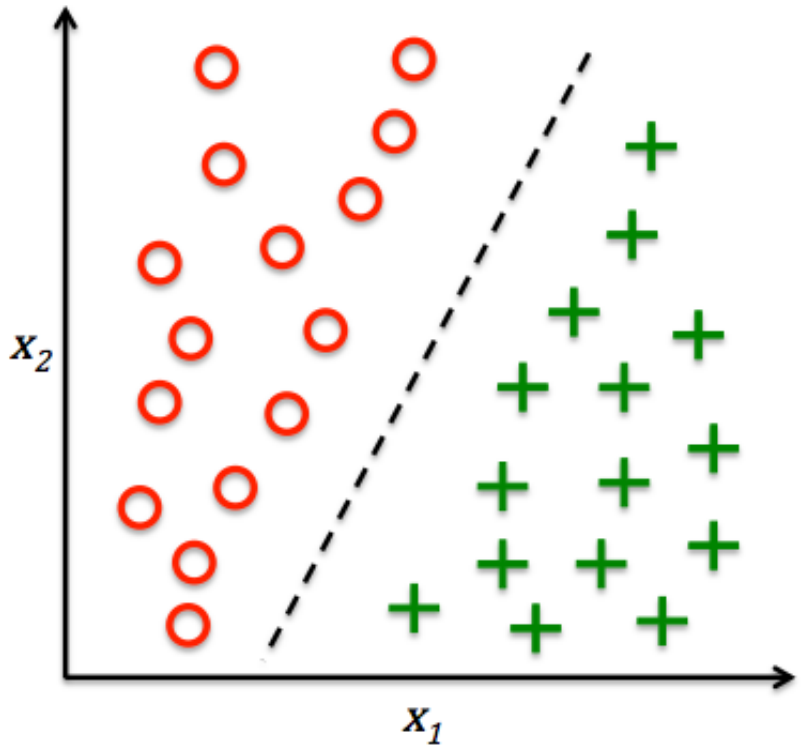
Supervised learning



- Predicting the future with supervised learning
- Classification vs. Regression

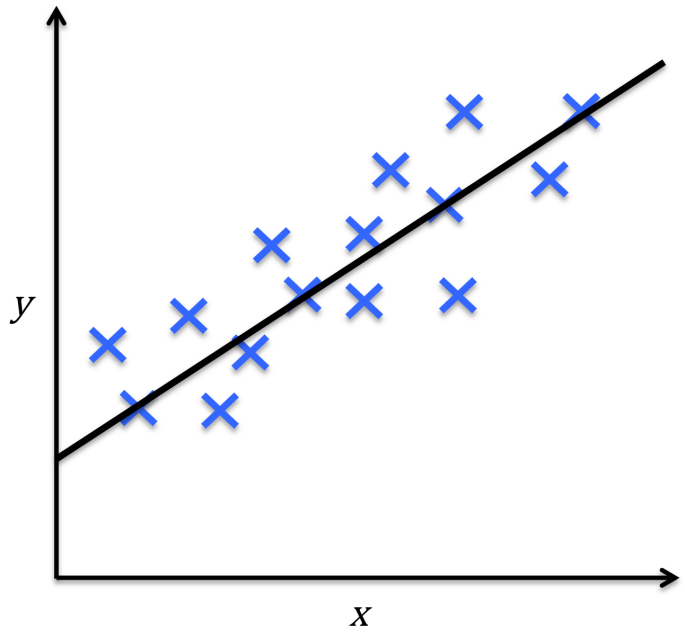
Classification

- Predict categorical class labels based on past observations
- Class labels are discrete unordered values
- Email spam classification example (binary)
- Handwritten digit classification example (multi-class)



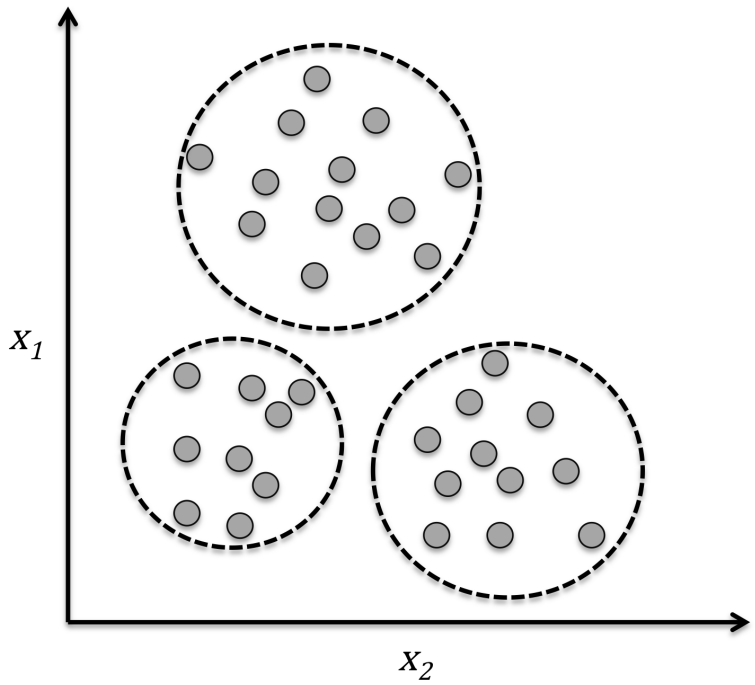
Regression

- Also a kind of supervised learning
- Prediction of continuous outcomes
- Predicting semester grades scores for students



Unsupervised learning

- Dealing with *unlabeled* data
- Cluster analysis
- Objects within a cluster share a degree of similarity



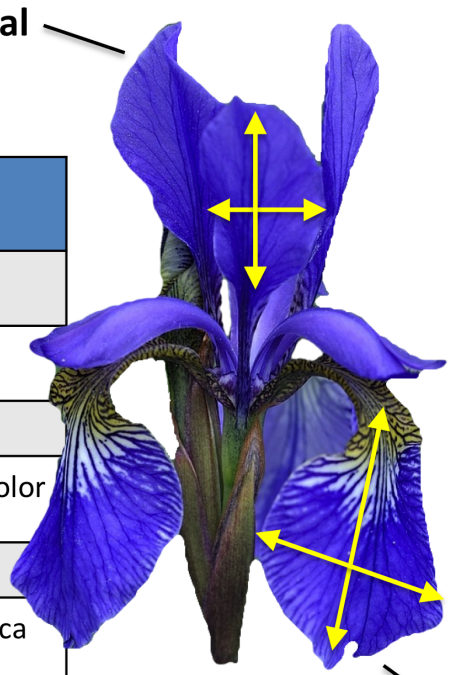
Unsupervised learing example

- Latent Dirichlet Allocation (LDA)
- Link to Wikipedia topics
- Wikipedia LDA entry
- Sara Palin topics

Iris dataset

Samples
(instances, observations)

	Sepal length	Sepal width	Petal length	Petal width	Class label
1	5.1	3.5	1.4	0.2	Setosa
2	4.9	3.0	1.4	0.2	Setosa
			...		
50	6.4	3.5	4.5	1.2	Versicolor
			...		
150	5.9	3.0	5.0	1.8	Virginica



Features
(attributes, measurements, dimensions)

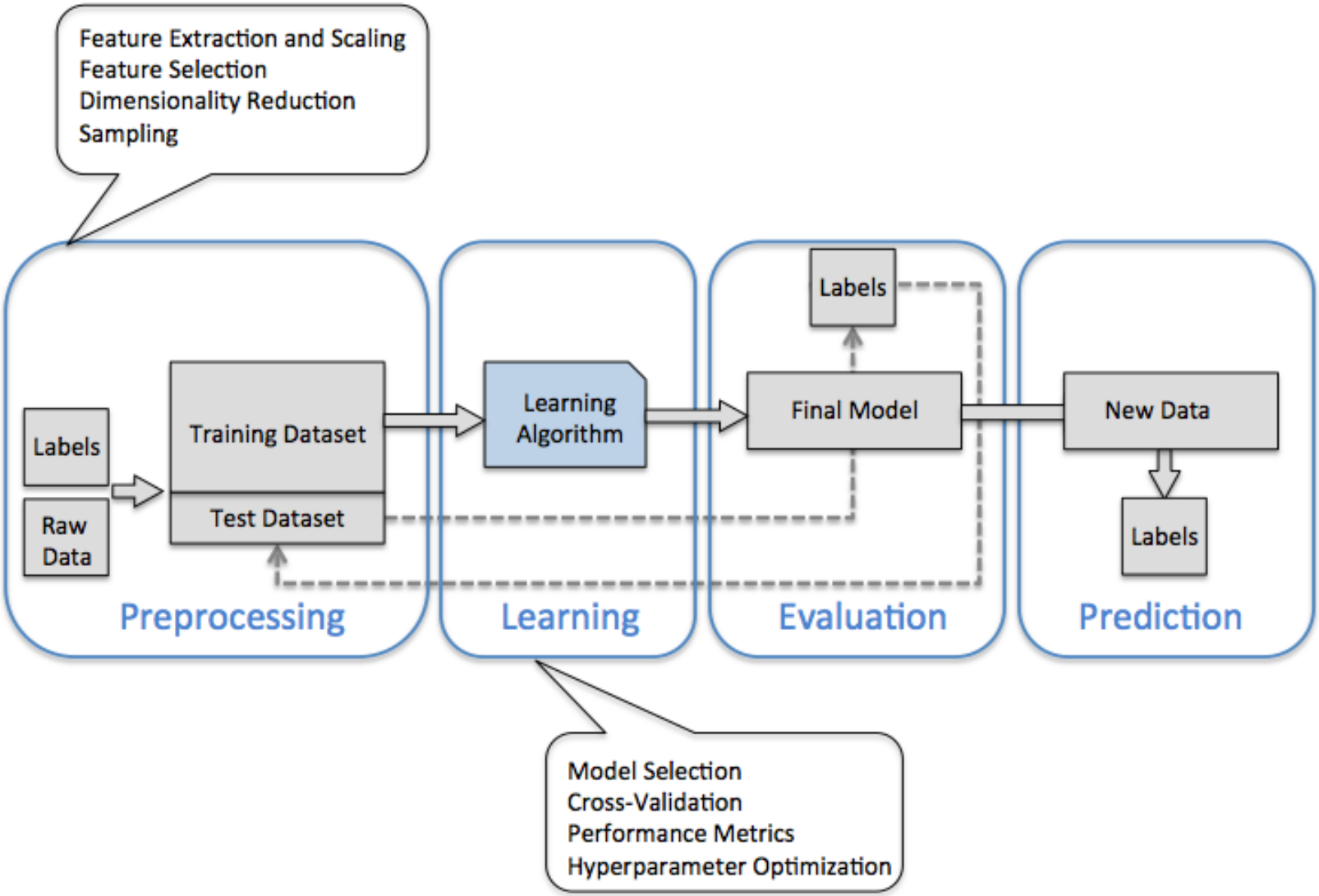
Class labels
(targets)

Basic terminology

- Measurements of 150 iris flowers (150 samples / 4 features)
- From 3 different species (Setosa, Versicolor, Virginica)
- Rows are samples and columns are features
- 150×4 matrix $\mathbf{X} \in \mathbb{R}^{150 \times 4}$:

$$\begin{bmatrix} x_1^{(1)} & x_2^{(1)} & x_3^{(1)} & \dots & x_4^{(1)} \\ x_1^{(2)} & x_2^{(2)} & x_3^{(2)} & \dots & x_4^{(2)} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ x_1^{(150)} & x_2^{(150)} & x_3^{(150)} & \dots & x_4^{(150)} \end{bmatrix}$$

A roadmap for building ML systems



Model selection

- No Free Lunch Theorems
- Each classification algorithm makes assumptions
- Often we empirically determine what works best
- But how do we know what works best?
  - Classification accuracy
  - Train+Dev+Test split
  - Hyperparameter optimization (knobs of the model)

Python for ML

- Libraries for scientific computing such as NumPy and SciPy
- Performance of interpreted languages is inferior
- But NumPy and SciPy build upon lower level C and Fortran subroutines
- Scikit-learn library
- See page 13 for installation instructions (or just google)
- NumPy Slides