Week 2

Machine Learning and Big Data - DATA622

CUNY School of Professional Studies

Review

Review concepts from last week

CUNY School of Professional Studies

Types of Machine Learning* (Predictive AI)

ML does one of three things*

1. **Predict a value**
2. **Predict a class**
3. **Cluster data**

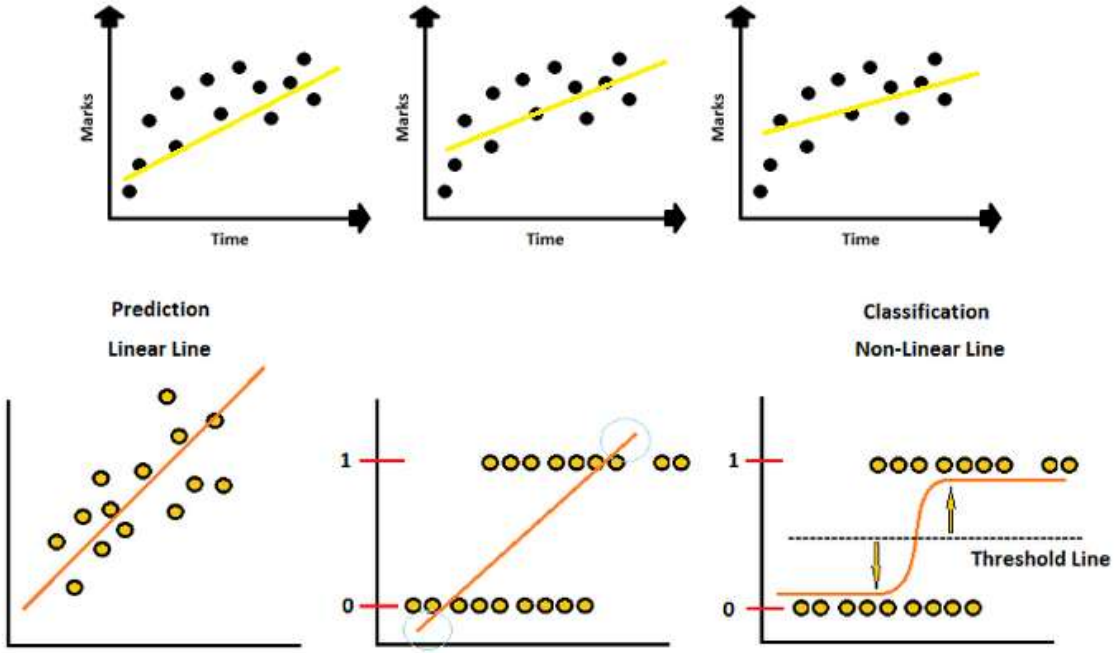
Supervised Machine Learning
(Requires labeled data)

Unsupervised Machine Learning
(Doesn't require labeled data)

*Note: Reinforcement Learning, semi-supervised, statistical & inductive models, stateful, temporal, knowledge graphs, etc. are a bit more complicated

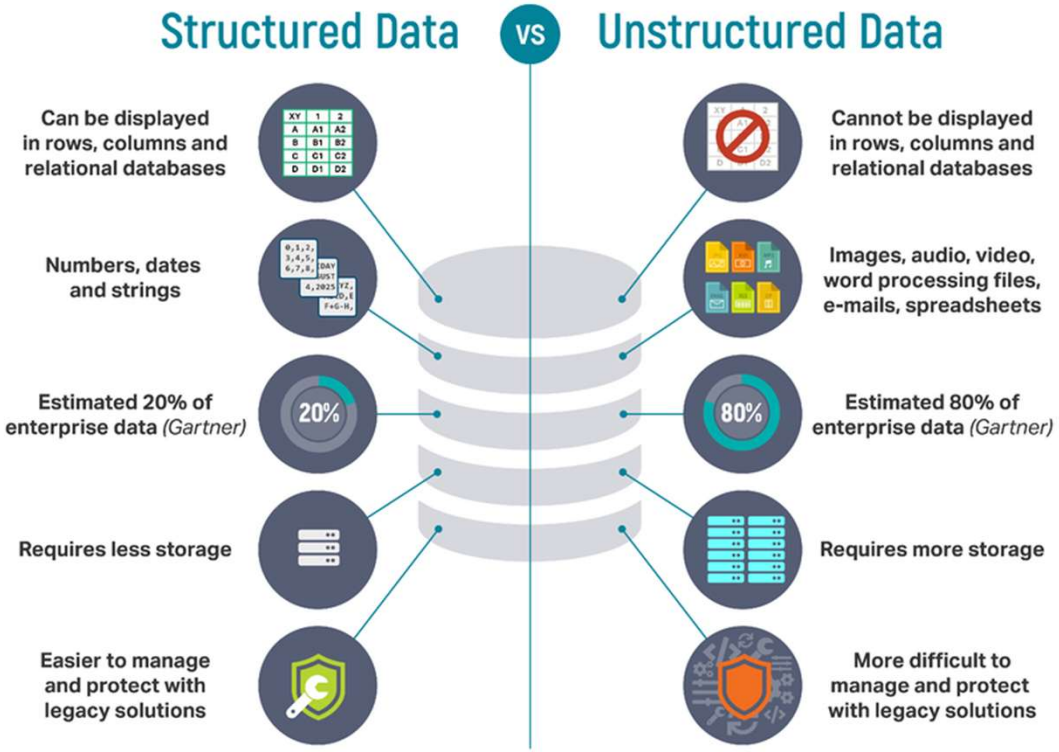
CUNY School of Professional Studies

Linear vs Logistics Regression



4

Types of Data



Source: igneous.io

CUNY School of Professional Studies

The topological view of Machine Learning

Let's think about what "learning" meaning by considering data geometry (data "shapes")

CUNY School of Professional Studies

Topological vs Mathematical

Let's Solve: $\begin{cases} 3x + y = 5 \\ 2x - y = 0 \end{cases}$

Algebra

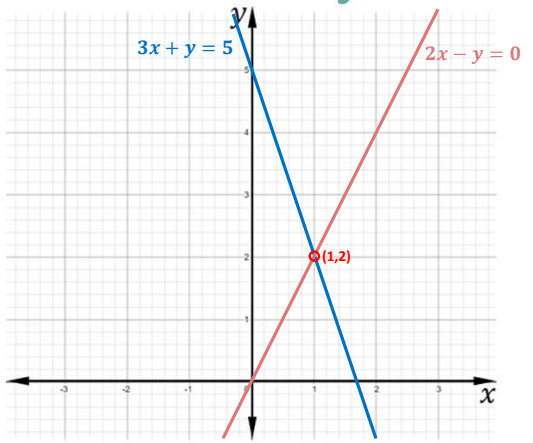
$$\begin{bmatrix} 3 & 1 \\ 2 & -1 \end{bmatrix} \cdot \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 5 \\ 0 \end{bmatrix}$$

$$C^{-1} = \frac{1}{3 \cdot -1 - 1 \cdot 2} \begin{bmatrix} -1 & -1 \\ -2 & 3 \end{bmatrix} = -\frac{1}{5} \begin{bmatrix} -1 & -1 \\ -2 & 3 \end{bmatrix}$$

$$-\frac{1}{5} \begin{bmatrix} -1 & -1 \\ -2 & 3 \end{bmatrix} \begin{bmatrix} 3 & 1 \\ 2 & -1 \end{bmatrix} \cdot \begin{bmatrix} x \\ y \end{bmatrix} = -\frac{1}{5} \begin{bmatrix} -1 & -1 \\ -2 & 3 \end{bmatrix} \begin{bmatrix} 5 \\ 0 \end{bmatrix}$$

$$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$$

Geometry



...they both give you the same answer but...

Which is easier to understand conceptually?

Source: mathplanet

An Example

Toy Example: Iris data set

Iris Versicolor

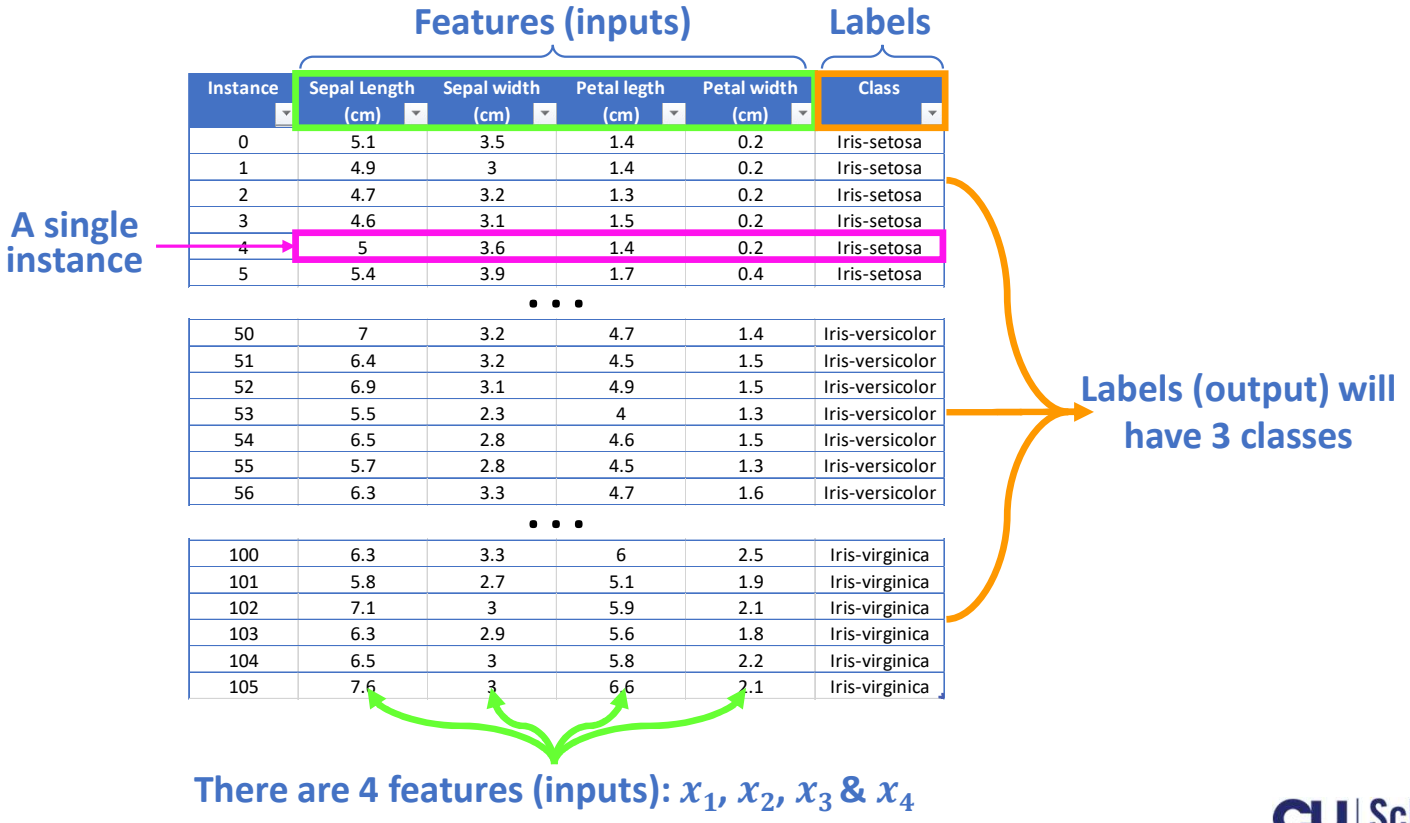
Iris Setosa

Iris Virginica

Source: Fisher (1936) https://onlinelibrary.wiley.com/doi/epdf/10.1111/j.1469-1809.1936.tb02137.x

An example using Structured Data

Toy Example: Iris data set

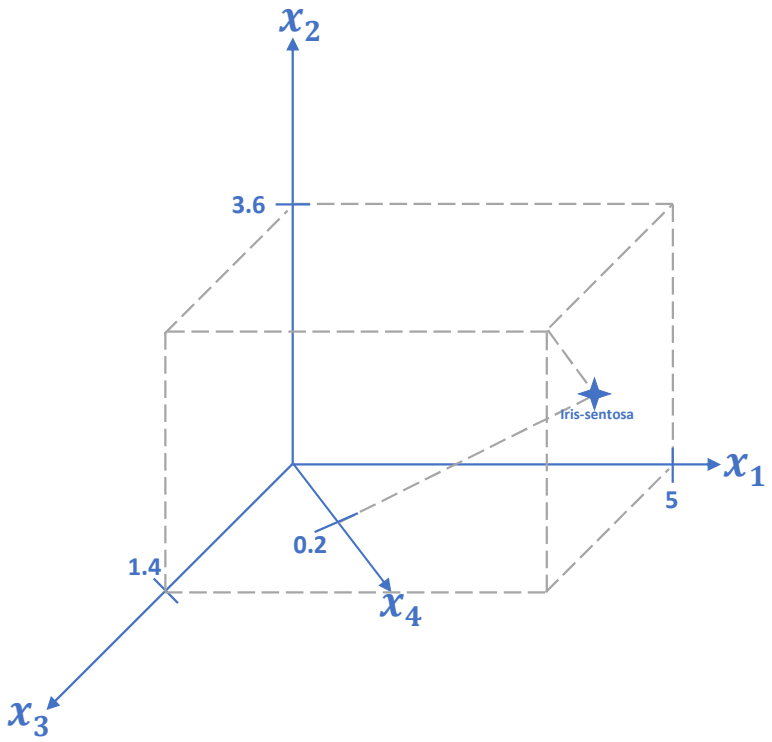


Instance	Sepal Length (cm)	Sepal width (cm)	Petal legth (cm)	Petal width (cm)	Class
0	5.1	3.5	1.4	0.2	Iris-setosa
1	4.9	3	1.4	0.2	Iris-setosa
2	4.7	3.2	1.3	0.2	Iris-setosa
3	4.6	3.1	1.5	0.2	Iris-setosa
4	5	3.6	1.4	0.2	Iris-setosa
5	5.4	3.9	1.7	0.4	Iris-setosa
• • •					
50	7	3.2	4.7	1.4	Iris-versicolor
51	6.4	3.2	4.5	1.5	Iris-versicolor
52	6.9	3.1	4.9	1.5	Iris-versicolor
53	5.5	2.3	4	1.3	Iris-versicolor
54	6.5	2.8	4.6	1.5	Iris-versicolor
55	5.7	2.8	4.5	1.3	Iris-versicolor
56	6.3	3.3	4.7	1.6	Iris-versicolor
• • •					
100	6.3	3.3	6	2.5	Iris-virginica
101	5.8	2.7	5.1	1.9	Iris-virginica
102	7.1	3	5.9	2.1	Iris-virginica
103	6.3	2.9	5.6	1.8	Iris-virginica
104	6.5	3	5.8	2.2	Iris-virginica
105	7.6	3	6.6	2.1	Iris-virginica

CUNY School of Professional Studies

Graphing the Data

- Every feature is a dimension
  4 features = 4 dimensions
- An instance is a single point in that 4-dimensional space
- All of the data forms a point-cloud in that 4-dimensional space



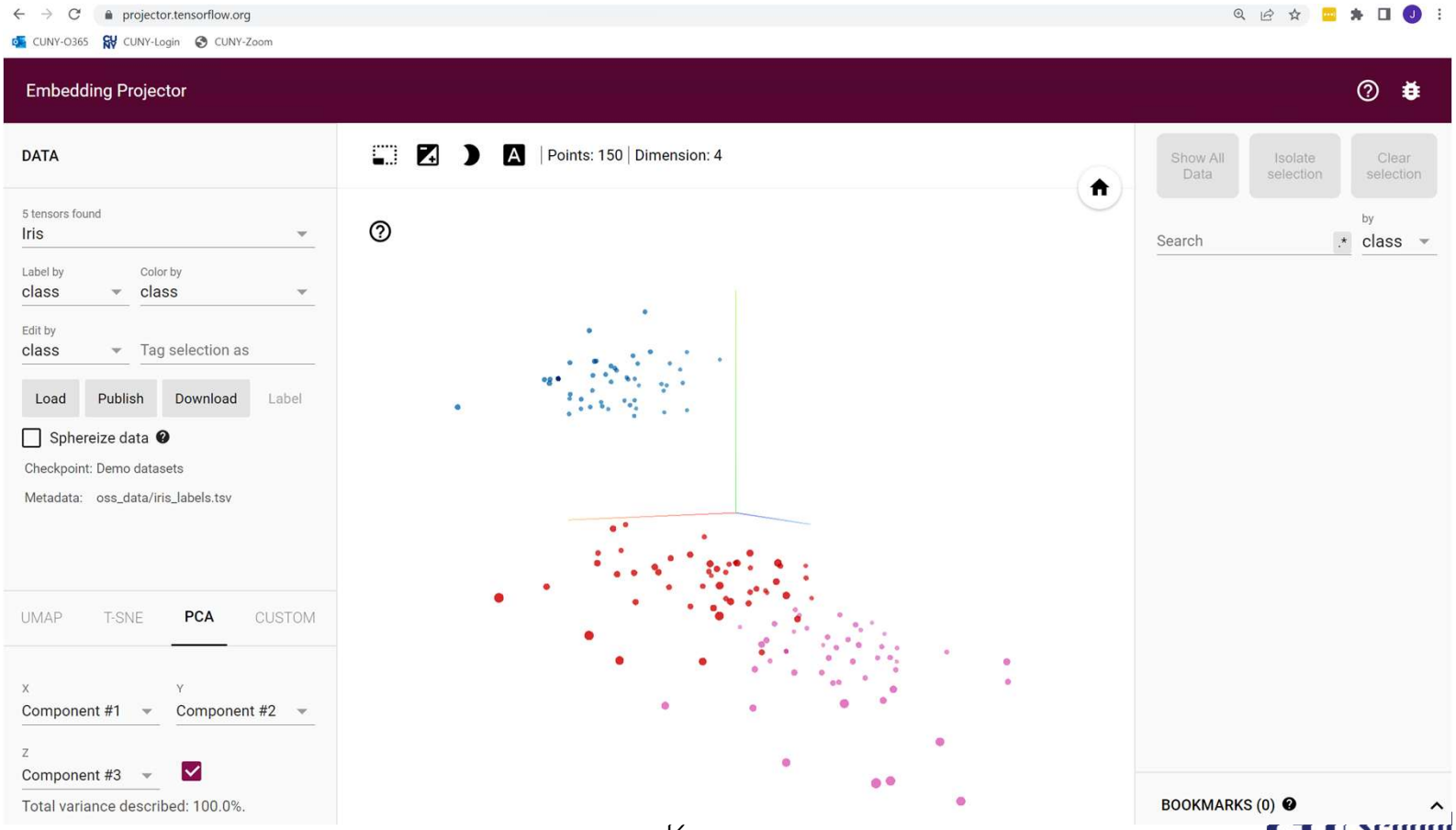
CUNY School of Professional Studies

Demo: Visualizing the data

projector.tensorflow.org

CUNY School of Professional Studies

Visualizing data



But what does the output data look like?

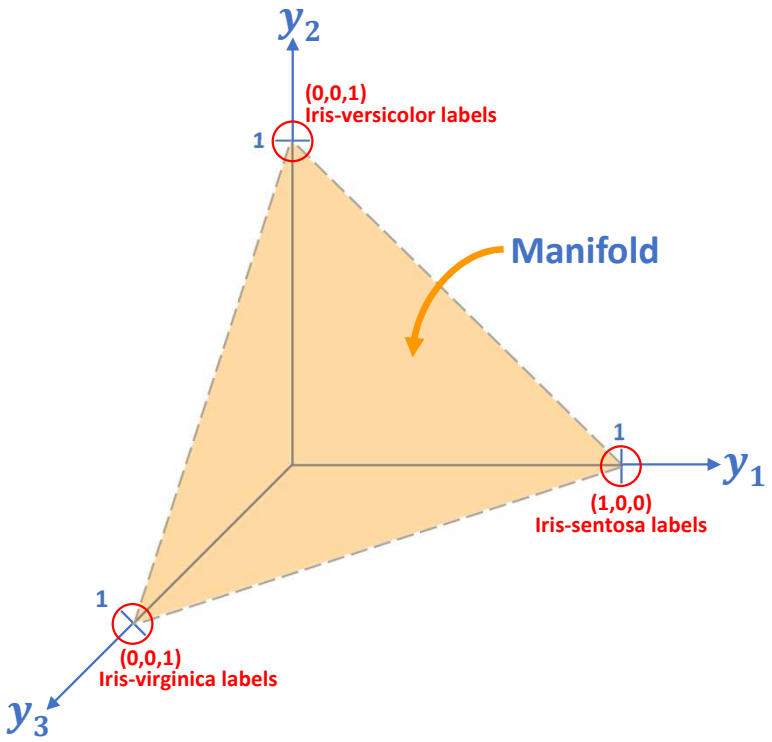
Let's consider what the solution space of the predictions looks like.

One-hot encoding

- ML requires numbers: labels must be converted to numbers
- Each class (type of label must be its own dimension)
- The value in each dimension conveys the probability it is of that class
- Training Data Labels always have a probability of 1 (100%)
  i.e. they are the “Ground Truth”

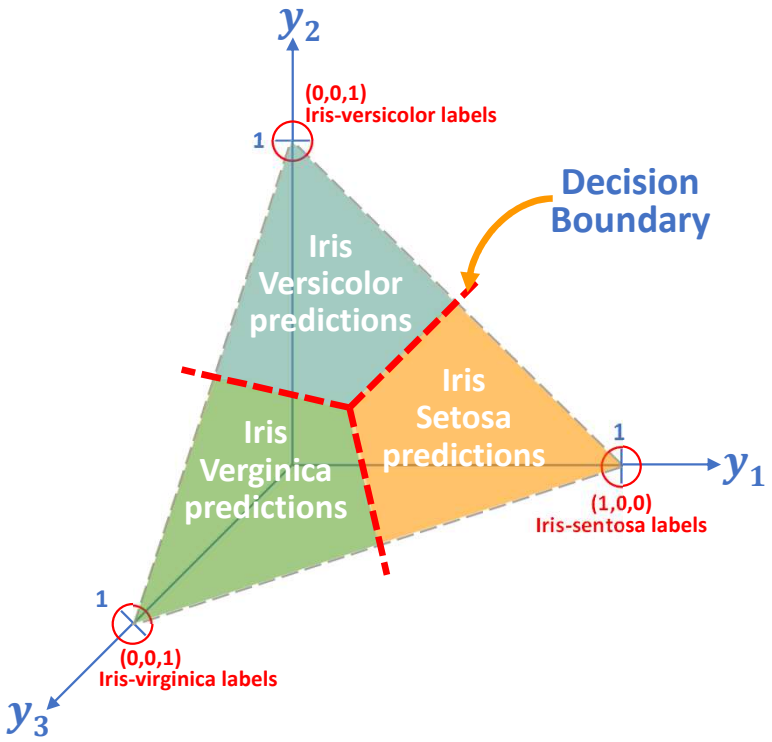
Solution Manifold

- The number of dimensions = number of classes. In this case 3 dimensions.
- A Label (or prediction) is one data-point in that 3-dimensional space
- Probabilities of all classes add up to 1 (100%) so points lie on a manifold
- Only labels have values of 1



Decision Boundary

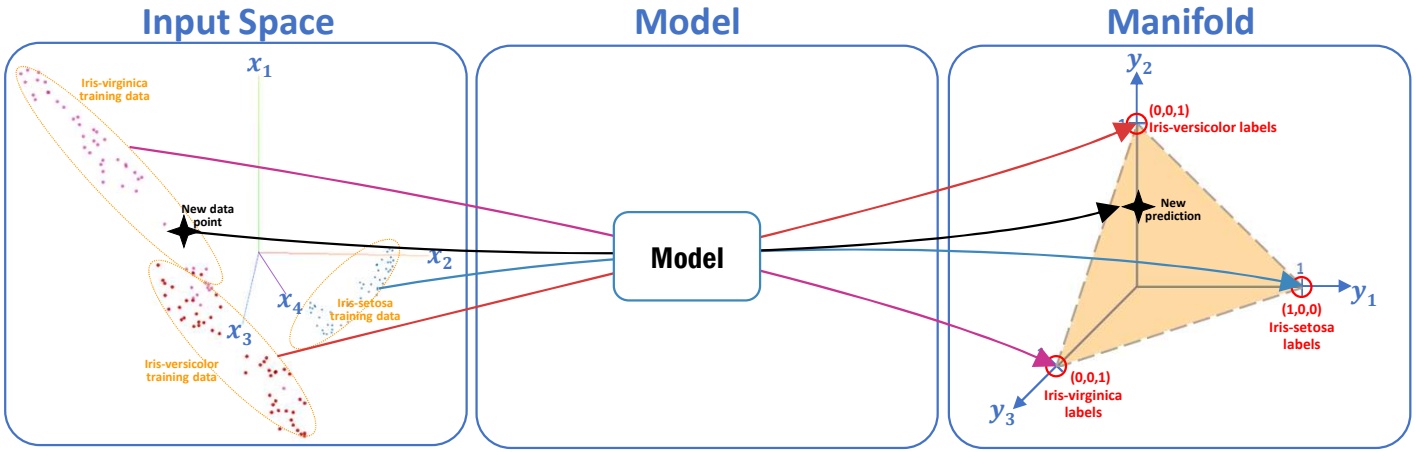
- A decision boundary separates the classes.
- For 2 classes the decision boundary is typically 0.5 (when probability of either class is 50%)
- It may be linear or non-linear



So What is Machine Learning?

....From a topological perspective (for supervised training)

*The **transformation** of data from a **high dimensional** space to a **low-dimensional** manifold.*



Statistical Learning vs Machine Learning

Comparing statistical learning and broader machine learning

CUNY School of Professional Studies

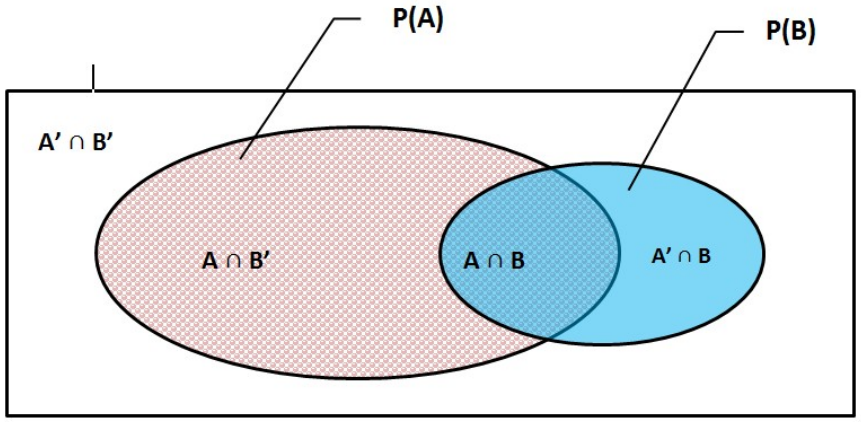
Bayes Theorem

$$P(A \mid B) = \frac{P(B \mid A)\, P(A)}{P(B)}$$

P(A)= The probability of A occurring P(B)= The probability of B occurring

P(A? B)=The probability of A given B P(B? A)= The probability of B given A

P(A? B))= The probability of both A and B occurring



19

Machine Learning vs Statistical Learning

MACHINE LEARNING	STATISTICAL LEARNING
Subfield of Artificial Intelligence	Subfield of mathematics
Uses algorithms	Uses equations
Requires minimum human effort; is automated	Requires a lot of human effort
Can learn from large data sets	Deals with smaller data sets
Has strong predictive abilities	Gives a best estimate: you gain some insights into one thing, but it's of little or no help with predictions
Makes predictions	Makes inferences
Learns from data and discovers patterns	Learns from samples, populations, and hypotheses

Source: *An introduction to Statistical Learning*

The 5 tribes of machines learning

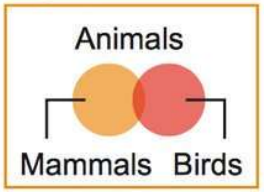
Different ways to train models

CUNY School of Professional Studies

Five Tribes of Machine Learning

What are the five tribes?

Symbolists



Use symbols, rules, and logic to represent knowledge and draw logical inference

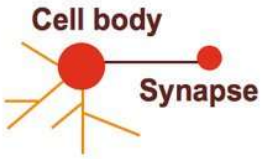
Favored algorithm
Rules and decision trees

Bayesians

Assess the likelihood of occurrence for probabilistic inference

Favored algorithm
Naive Bayes or Markov

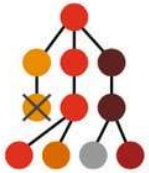
Connectionists



Recognize and generalize patterns dynamically with matrices of probabilistic, weighted neurons

Favored algorithm
Neural networks

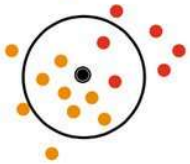
Evolutionaries



Generate variations and then assess the fitness of each for a given purpose

Favored algorithm
Genetic programs

Analogizers



Optimize a function in light of constraints ("going as high as you can while staying on the road")

Favored algorithm
Support vectors

Source: Pedro Domingos, *The Master Algorithm*, 2015

Which Algorithm to choose for a Model?

CUNY School of Professional Studies

Types of Machine Learning