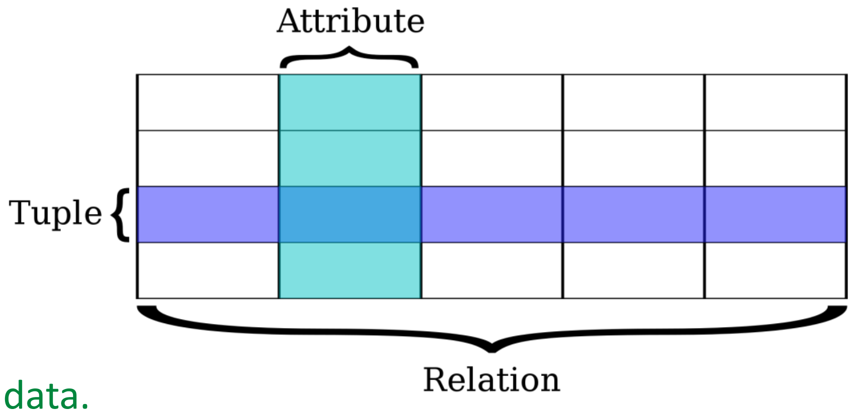
Supervised Learning

Modelling

~~Un~~Supervised Learning



- Unsupervised Learning
  - Given X
  - … the task of inferring a function to describe hidden structure from unlabeled data.
  - Distribution / Density, Summary statistics, Clustering, Association Rules, Dimensionality Reduction

- Supervised Learning
  - Given X & y (a **particular** random variable)
  - Find what is the relation between the particular random variable and other random variables
    - What if we are only interested in identifying customers who bought Milk?
  - Find how the value of the dependent variable depends on the value of others
  - Find how the outcome is related to the features
  - Key Variations: Type of outcome / dependent r.v.
    - Numeric (Discrete, Continuous, [0,1])
    - Categorical : Nominal, Ordinal

The idea of a Model

- Physical
  - a physical copy of an object such as a globe
- Computer
  - a simulation to reproduce behavior of a system
- Scientific
  - a simplified & idealized understanding of physical systems
  - Newton’s Law model the physical universe

- Conceptual
  - a representation of a system using general rules & concepts
- Mathematical
  - a representation of a system using mathematical concepts
- Statistical
  - a parameterized set of probability distributions

$$y = 3x + 4$$
$$y = x^2$$
$$y = e^x$$
$$y = \log(x)$$
$$y = \sin(x)$$

All models are false. Some models are useful.

The idea of a Statistical / ML Model

- Model
  - A function relates two (or more) variables
  - Captures the relation between x and y
  - For every value of x, there must be a unique value of y
  - Data looks like $\{(x_1, y_1), (x_2, y_2), ..., (x_i, y_i), ...(x_n, y_n)\}$

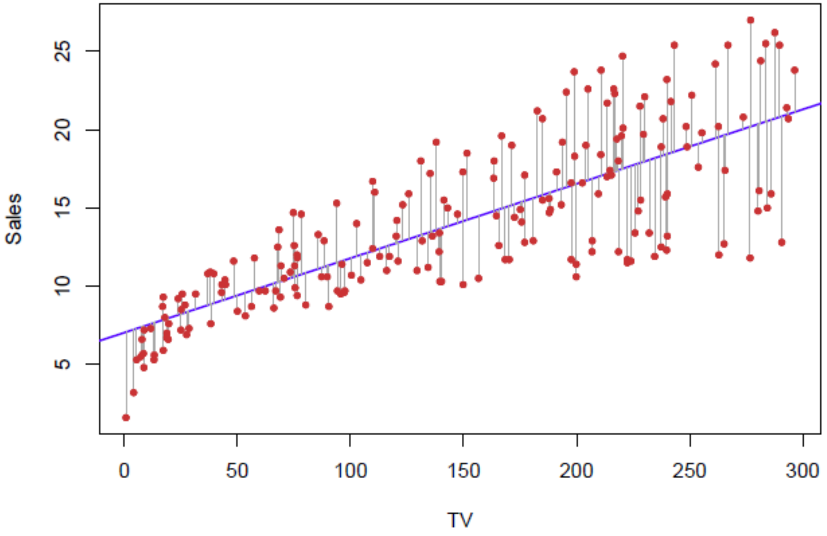
$$y = 3x + 4$$
$$y = x^2$$
$$y = e^x$$
$$y = \log(x)$$
$$y = \sin(x)$$

$$y = f(x)$$

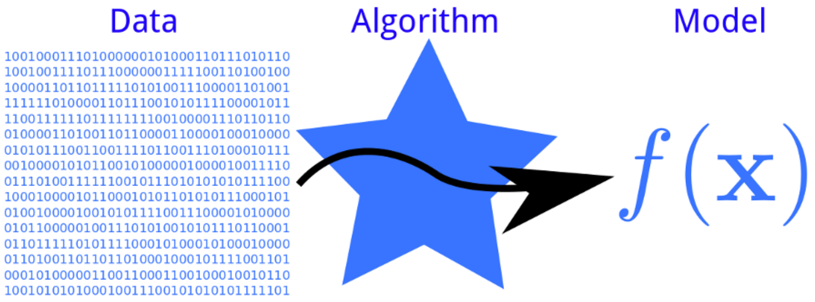
- Statistical Model
  - Real world data looks like $\{(x_1, y_1), (x_1, y_2), ..., (x_n, y_n)\}$
  - Multiple values of y for a single value of x
  - In expectation (on average), "model" captures the relationship between variables
  - Effects due to unobserved variables / Errors in measurements : capture by ε
  - Randomness / Stochasticity / Noise : Zero-mean; Normal distribution
  - Violations of Assumption is an indication of systemic errors

$$y = f(x) + \varepsilon$$
$$\varepsilon \sim \mathrm{N}(0, \sigma)$$

$$\hat{y} = \hat{f}(x) + 0$$
$$\mathrm{P}(y \mid x)$$

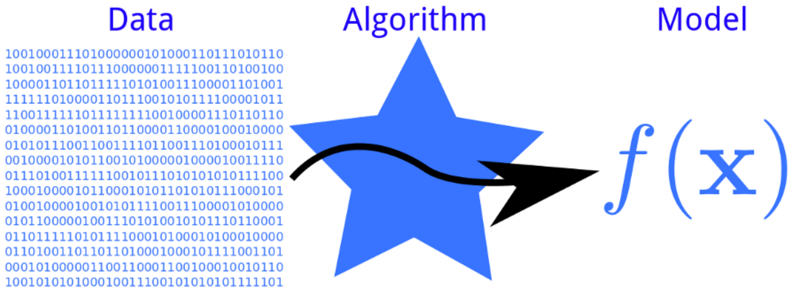


Un/Supervised Learning

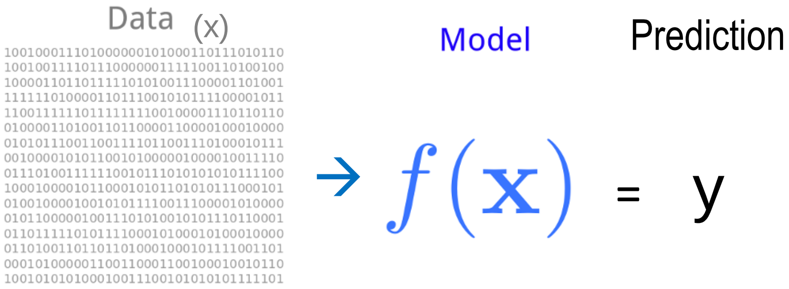


- Given X
  - … the task of inferring a function to describe hidden structure from unlabeled data.
  - Distribution / Density, Summary statistics, Clustering, Association Rules, Dimensionality Reduction

$$y = f(x)$$



- Given X & y (a **particular** random variable)
  - Find what is the **relation** between the particular random variable and other random variables
  - Find how the value of the **dependent (particular)** variable depends on the value of others
  - Find how the outcome is related to the **features**
  - Generalize : Make **predictions** about new data



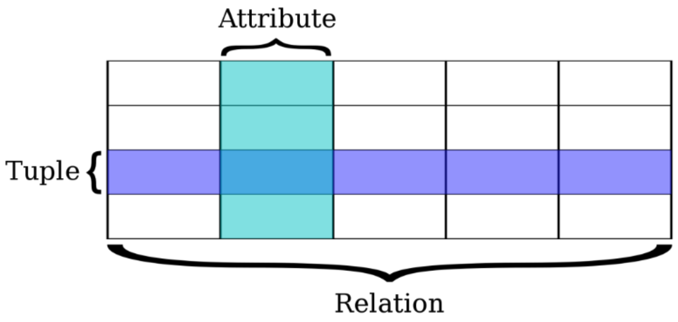
Supervised Learning

Variants

Un/Supervised Learning Models

- Supervised
  - Dependent vs. Independent Variables
  - Is there a variable of interest? Labelled data?
  - Do you know what you are looking for?
  - View the data as $\{(x_1, y_1), (x_1, y_2), \ldots, (x_n, y_n)\}$
  - Regression vs. Classification

- Unsupervised
  - No clearly defined Dependent Variable
  - Find patterns in data
  - View the data as $\{(x_1), (x_2), \ldots, (x_n)\}$
  - Often, a pre-processing step to Supervised

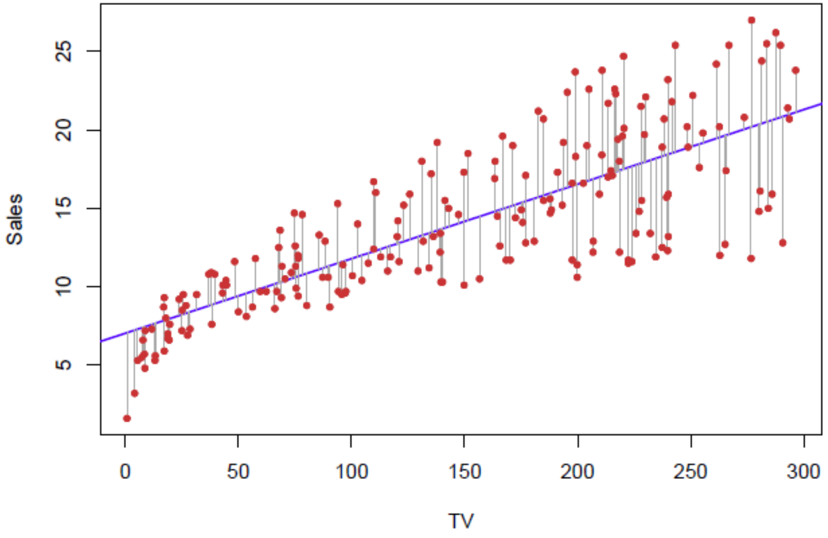


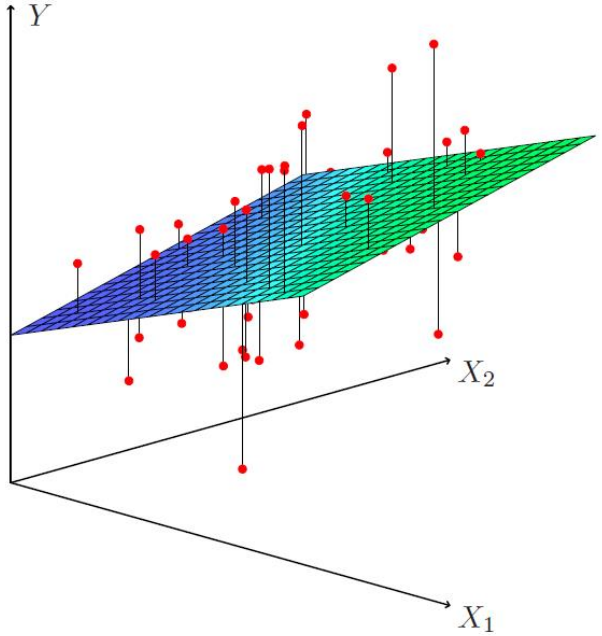
- Parameteric
  - Specify the "form" of f *(Specify model class)*
  - Learn exact f *(Learn model parameters)*
  - Restrictive but Interpretive
  - Less data required for learning

- Non-Parameteric
  - Learn model directly *(No restrictions on model class)*
  - Flexible but less Interpretive

- Model-Based vs. Model-Free
  - Models are not the only game in town
  - Model-Based: Linear Regression (What is the model?)
  - Model-Free: Nearest Neighbor, Collaborative Filtering

Supervised Learning : Linear Regression



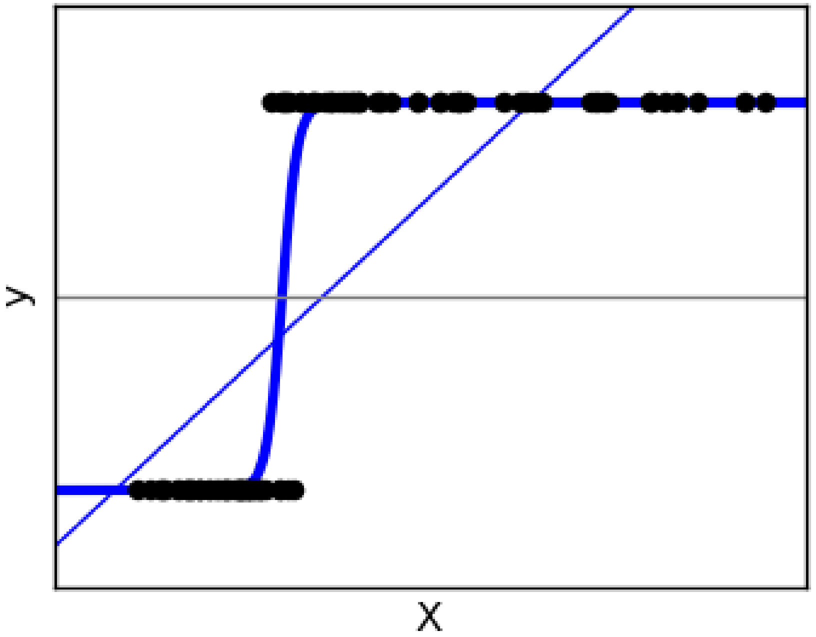


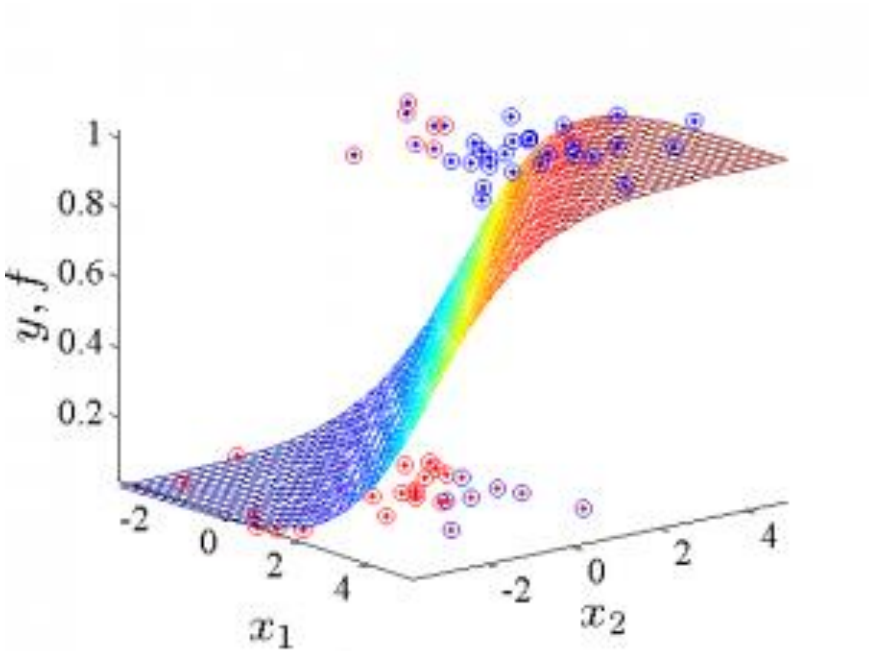
p=1

p=2

p > 2 ?

Supervised Learning : Binary classification



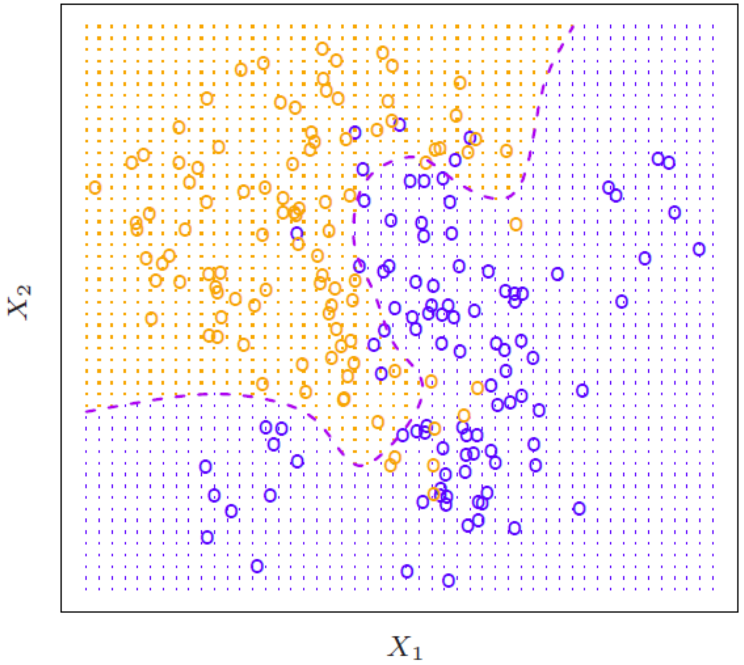


p=1

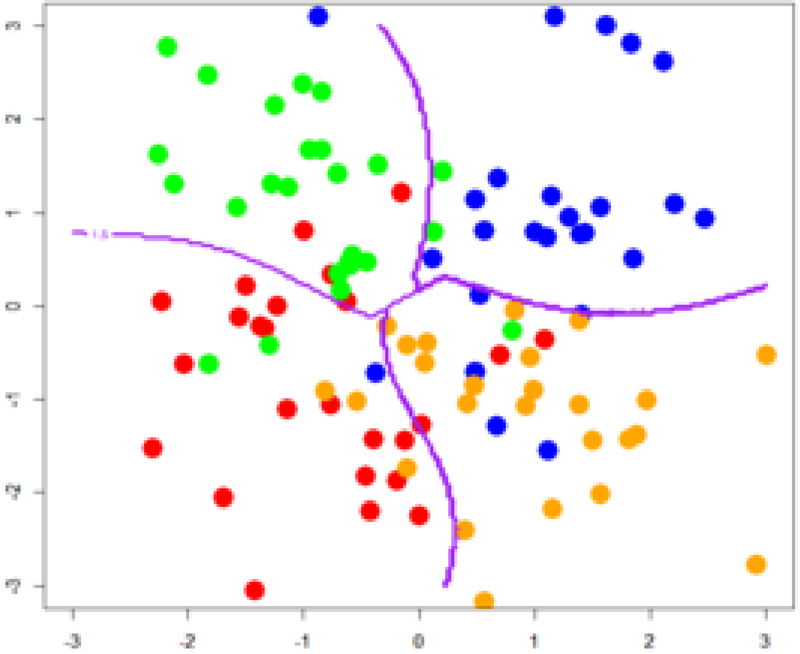
p=2

p > 2

Supervised Learning : From Binary to Multi Class



p=2

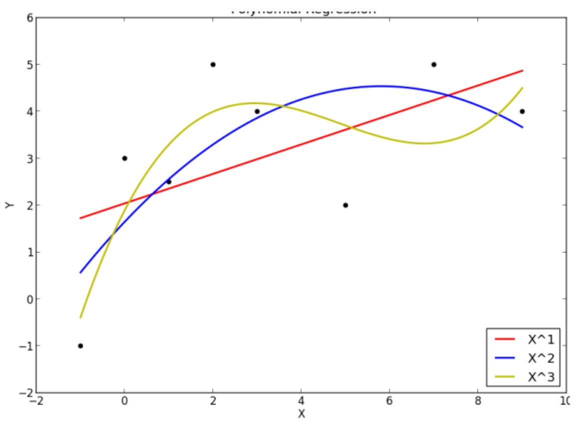


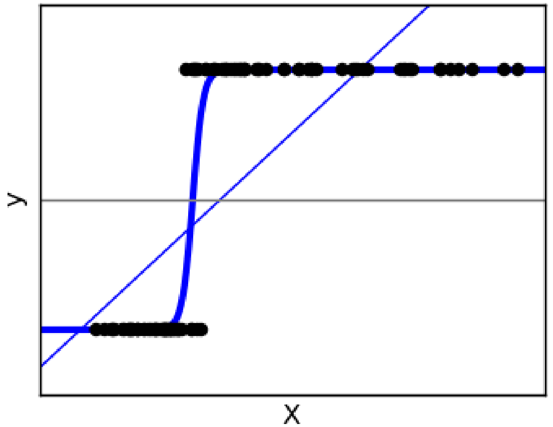
p=2

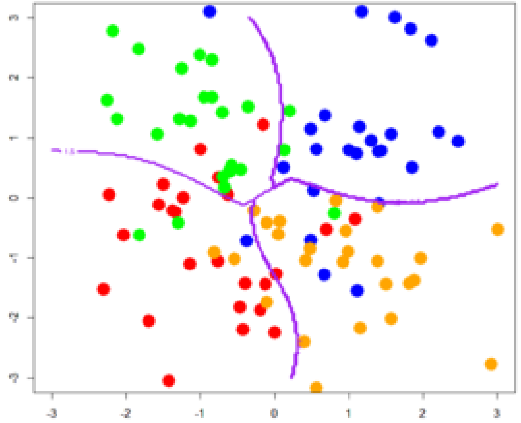
p > 2 ?

SL: Variant Summary

- Numeric y
    - Given input data x, f(x) is a numeric value
    - Regression: Linear, polynomial, lasso
    - Time Series : y = xt+1
- Numeric y in [0,1]
    - Given input data x, f(x) is a numeric value in between 0,1 (e.g. probability)
    - Regression: Logistic
- Categorical y
    - Given input data x, f(x) is a label / class / category (e.g. churn or not)
    - Classification: knn, logistic, decision tree, svm
- Ordinal y
    - Learn f(x) such that given input data x, f(x) is a rank (e.g. 1st, 2nd , …)
    - Ranking


Polynomial Regression
Y
X
X^1
X^2
X^3





3
2
1
0
-1
-2
-3
-3
-2
-1
0
1
2
3