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
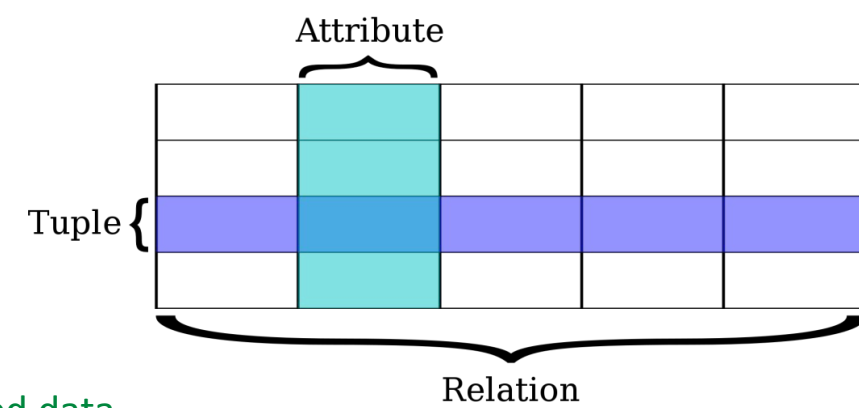
The idea of a Model

- 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

$$y = 3x + 4$$

- Mathematical
  - a representation of a system using mathematical concepts

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

- Statistical
  - a parameterized set of probability distributions

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 $\{(x_1, y_1), (x_2, y_2), ..., (x_i, y_i), ...(x_n, y_n)\}$
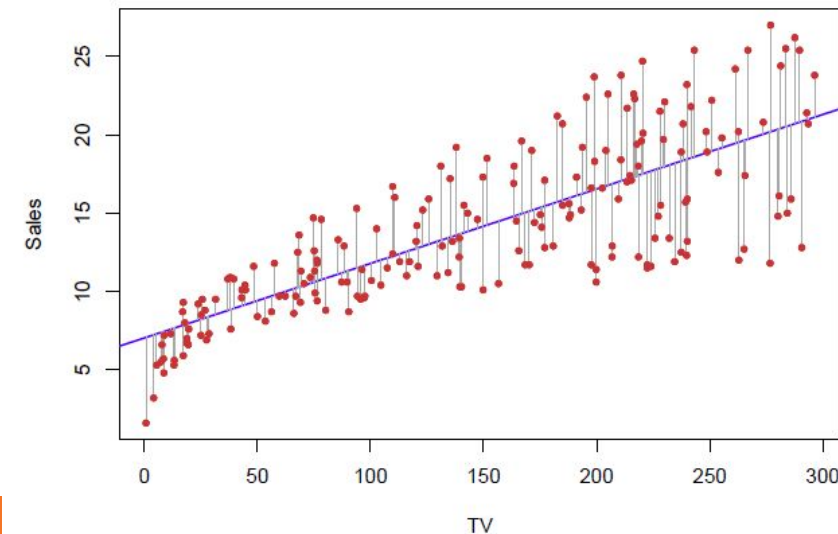
$$y = 3x + 4$$
$$y = x^2$$
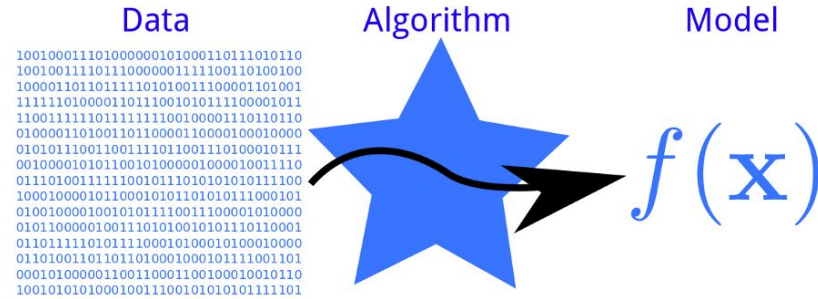$$y = e^x$$
$$y = \log(x)$$
$$y = \sin(x)$$

$$y = f(x)$$

- 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
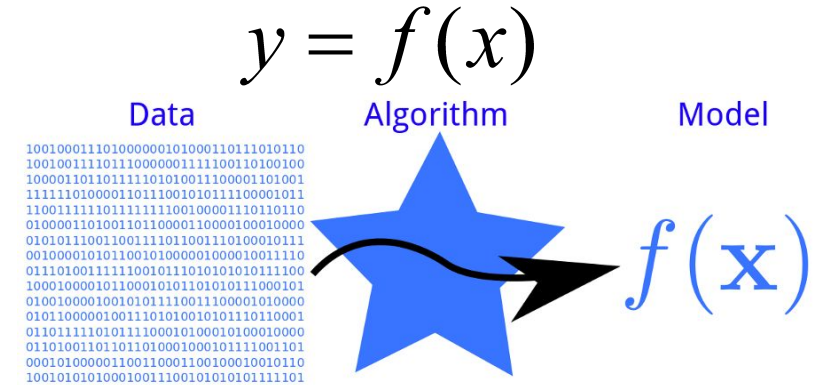 - Violations of Assumption is an indication of systemic errors

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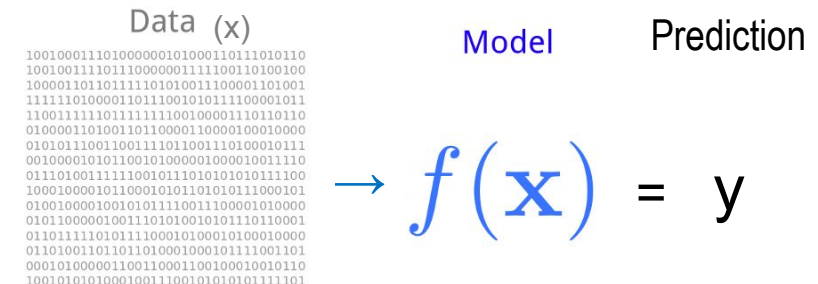$$\varepsilon \sim N(0, \sigma)$$

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

Un/Supervised **Learning**

Unsupervised (left panel)

Data	Algorithm	Model
10010001110100000010100011011101101001001110111000000111110011010010010000110110111110101001011100001011111111010000110111001010111100001011110011111101011111110010000111011011001000011010011011000010000100010000010101011100110011011010011011001000101100010001010110010100000100010011110011110100111111001011110101010101111001000100001011000010101101010101110001010100100000100101011110011100001010000010101100000010011010011001011011010001001000100000101101111101011110001010001010001000001011011110110110110010010000100010111100110100010100000110010001100010010010110100101010101000100111001010101010111101		$f(\mathbf{x})$

- 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 (right panel)

$$y = f(x)$$

Data	Algorithm	Model
1001000111010000001010001101110101100100100111101110000001111100110100100100001101101111101010010011100001011110011111101000011011100101011110000101111100111110101111111001000011101101100100001101001101100001000010001000001010101110011001101101001101100100010110001000101011001010000010001001111001111010011111101011111110010000111011011001000011010010110110100010001000001011011111010111100010100011000101100101001100010010010110100101010101010111000101		$f(\mathbf{x})$

- 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**
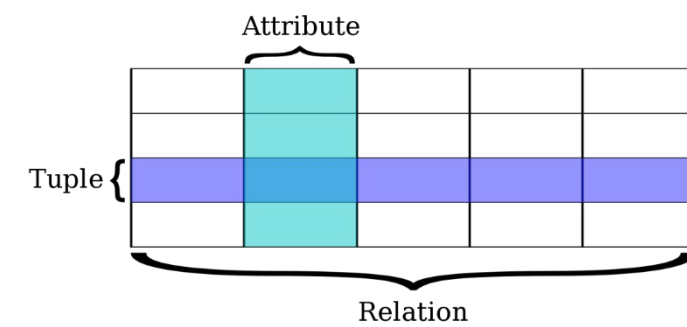 - Generalize : Make **predictions** about new data

Data (x)	Model	Prediction
100100011101000000101000110111010110100100111101101100000011111001101001001000011011011111010100101110000101111001111110100001101110010101111000010111110011111010111111100100001110110110010000110100110110000100001000100000101010111001100110110100110110010001011000100010101100101000001000100111100111101001111110101111111001000011101101100100001101001101100001000010001000001011011111010111100010100011000101100101001100010010010110100101010101010111000101	$\rightarrow f(\mathbf{x})$ = y	

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 $\{(x_1, y_1), (x_1, y_2), ..., (x_n, y_n)\}$
  - Regression vs. Classification

- Unsupervised
  - No clearly defined Dependent Variable
  - Find patterns in data
  - View the data as $\{(x_1), (x_2), ..., (x_n)\}$
  - 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
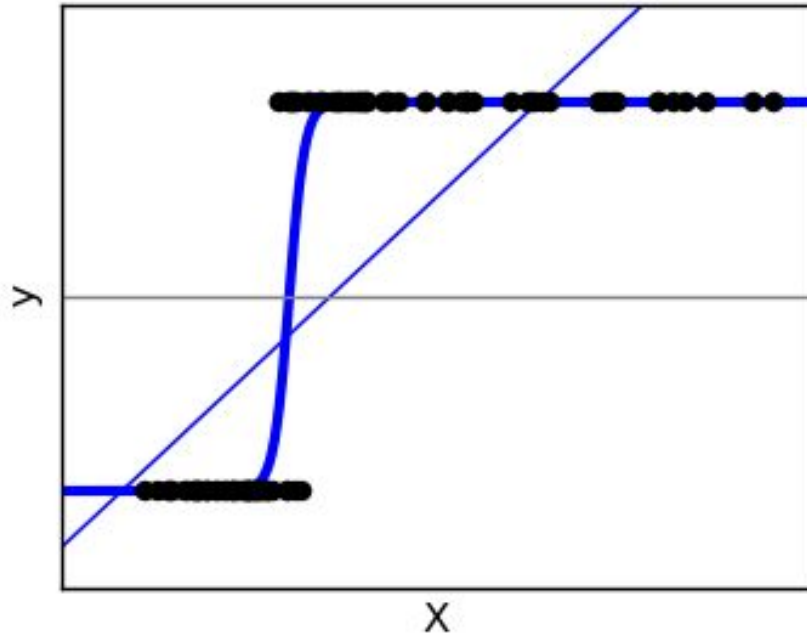
Supervised Learning : Linear Regression

p=1

p=2

p > 2 ?
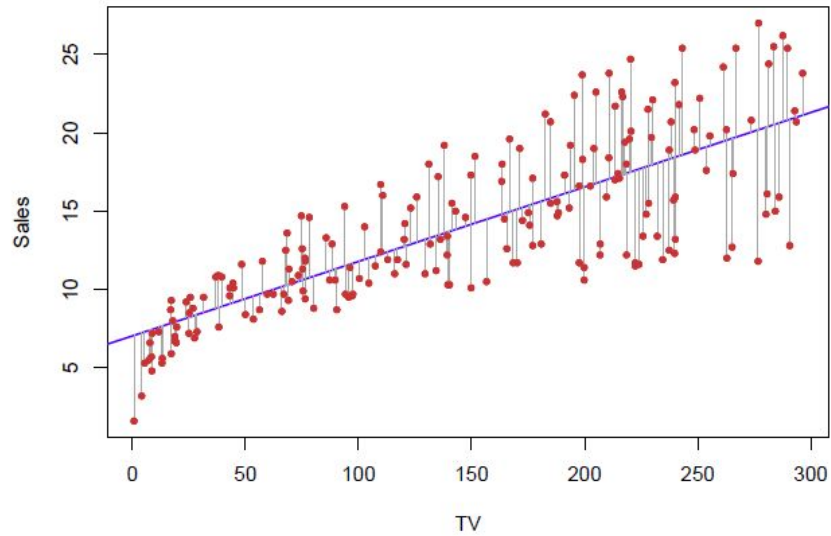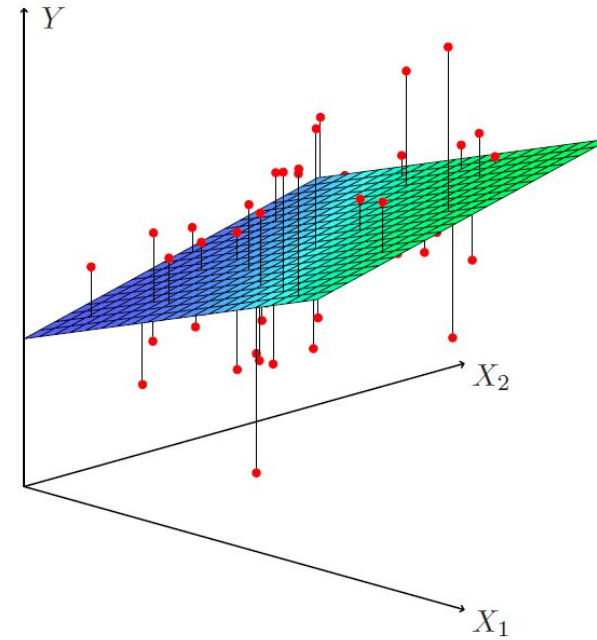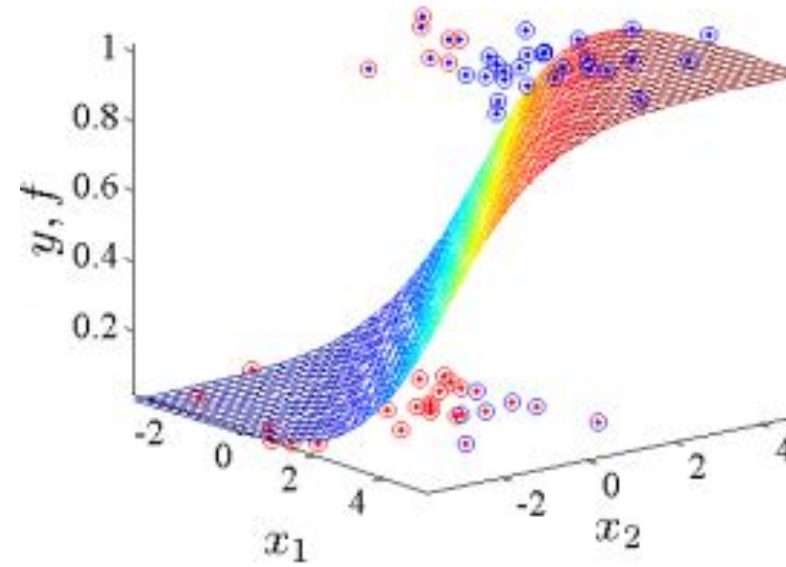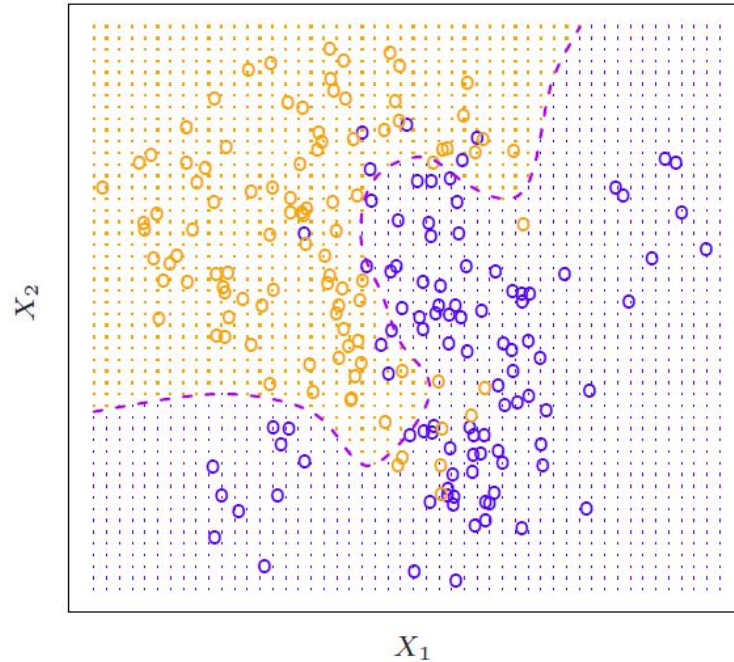
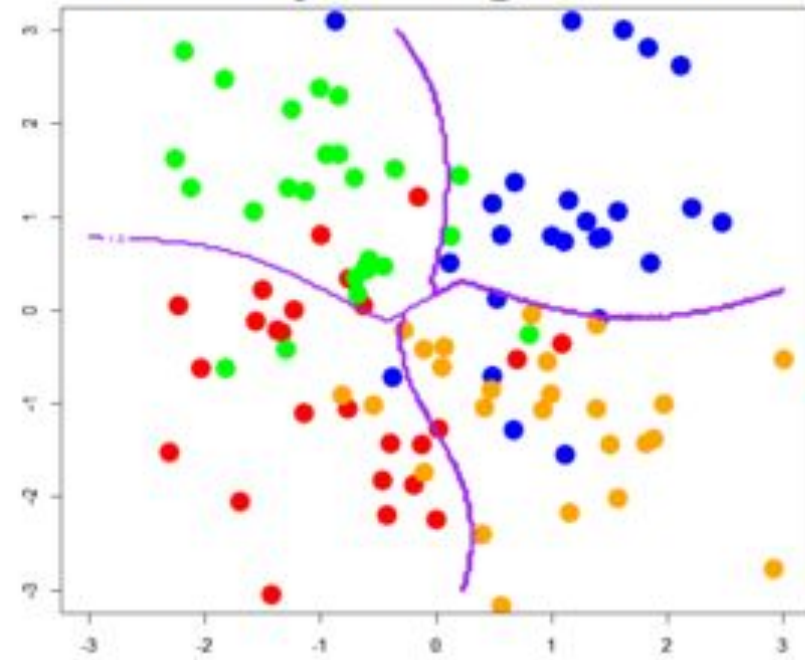# Supervised Learning : Binary classification

p=1

p=2

p > 2
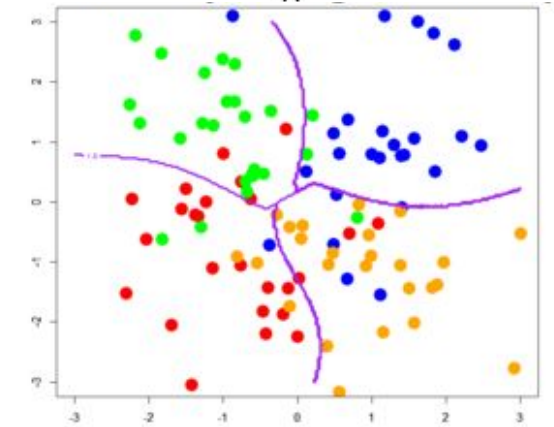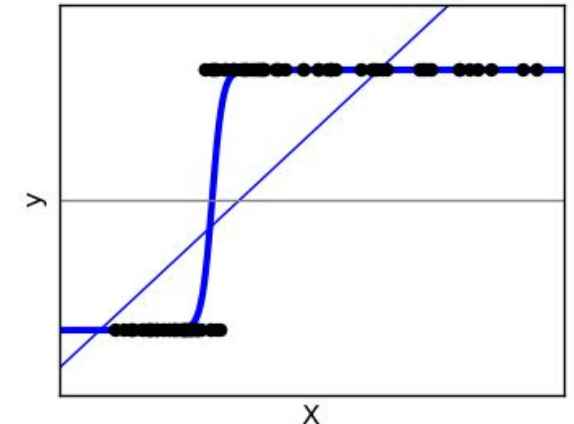
Supervised Learning : From Binary to Multi Class

p=2
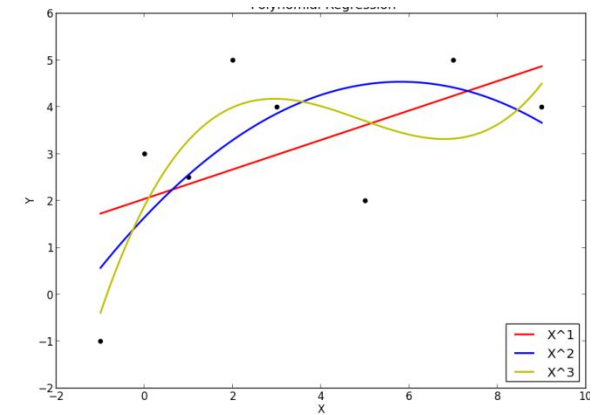

p=2

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

Machine Learning Framework

Data Summary / Exploratory Data Analysis / Feature Engineering → Training-Calibration-Test Data Split → Select ML Algorithm (Model Family) → Set Model Hyper-parameters → Learn Model parameters from **Training** data (Optimization) → Measure Performance on **Calibration** Data → Report performance on **Test** Data

ML model

New Data

Prediction

... → ... → Linear Regression → Polynomial Degree → Coefficients of independent vars. → ... → ...

... → ... → Decision Tree → Cost Complexity parameter → Splitting points (Regions in Feature Space) → ... → ...

Let's play

Thought Experiments

Data	Past credit card transactions of customers
Business Objective	Identify fraudulent transactions
Analytics	?

Data	Past purchases of customers
Business Objective	What is a customer likely to buy next?
Analytics	?

Data	Pricing and Sales data of a product portfolio
Business Objective	Determine price elasticity
Analytics	?

Thought Experiments (cont'd)

Data	?
Business Objective	How much should company spend on TV/radio/paper ads?
Analytics	?
Data	Past purchases of customers
Business Objective	Segment customers with similar purchase behavior
Analytics	?
Data	**A set of emails marked junk or not by a human**
Business Objective	Build a rules engine to determine emails as Junk or not?
Analytics	?