MPSSDEGB
कौशल से रोजगार, समृद्धि अपार

FITT
Foundation For Innovation And Technology Transfer

SANKALP
Ministry of Skill Development & Entrepreneurship

DAY 2 - Machine Learning Fundamentals

Jan, 2024

Machine Learning Fundamentals

- Overview of Machine Learning
  - Types of learning: Supervised and Unsupervised.
- Supervised Learning Deep Dive
  - Introduction to classification and regression.
- Hands-on activity with supervised learning algorithms.
- Q&A and interactive discussion.

- Homework:
  - Complete a small supervised learning exercise.

Jan, 2024

MACHINE LEARNING FUNDAMENTALS B1M2L1T1

Overview of Machine Learning

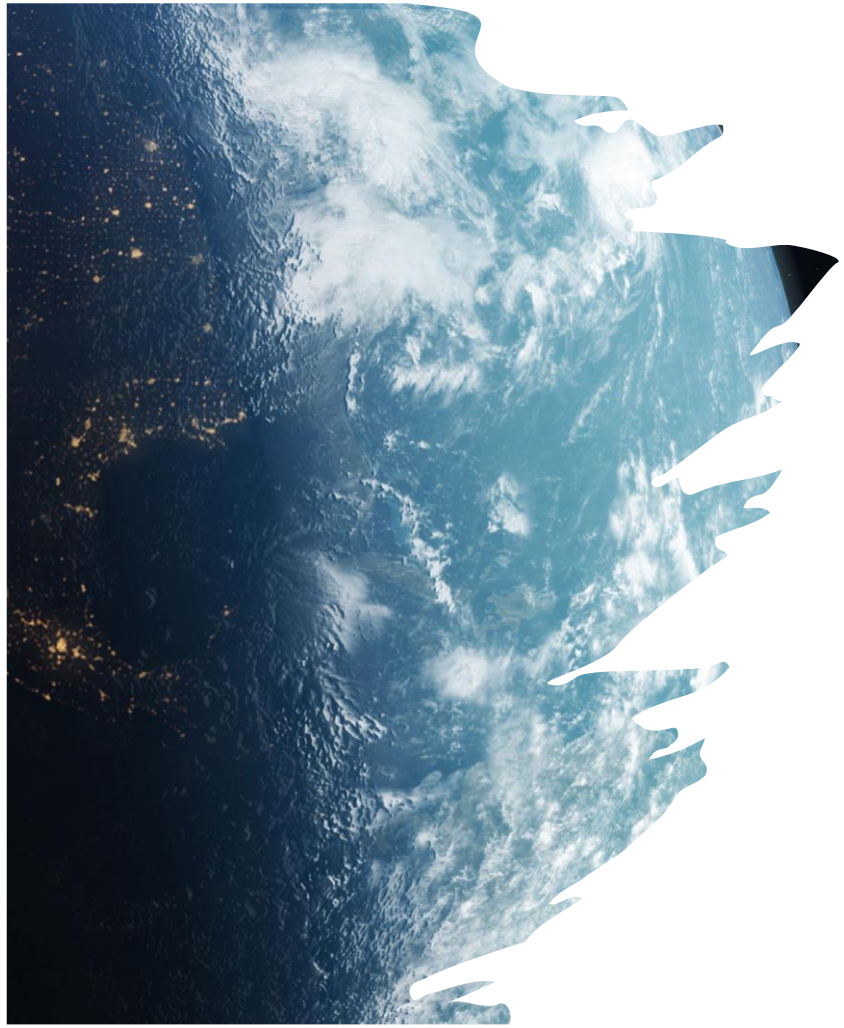
01 **Machine Learning Introduction**

02 **Supervised Learning**

03 **Unsupervised Learning**

Machine Learning (ML)

- ML is a branch of artificial intelligence:
  - Uses computing based systems to make sense out of data
    - Extracting patterns, fitting data to functions, classifying data, etc
  - ML systems can learn and improve
    - With historical data, time and experience
  - Bridges theoretical computer science and real noise data.



PREAMBLE

- We are in the midst of what is popularly called the information revolution—a revolution which was born shortly after the end of World War II.
- The new world was the world of machine intelligence and automated reasoning
- It was widely believed that there were no limits to what machines could do
- The era of thinking machines has arrived

INTERNET
WORLD WIDE WEB
WIRELESS TELEPHONY
FAX
DIGITAL LIBRARIES
DATA MINING
MANUFACTURING
INFORMATION RETRIEVAL
SMART SEARCH ENGINES

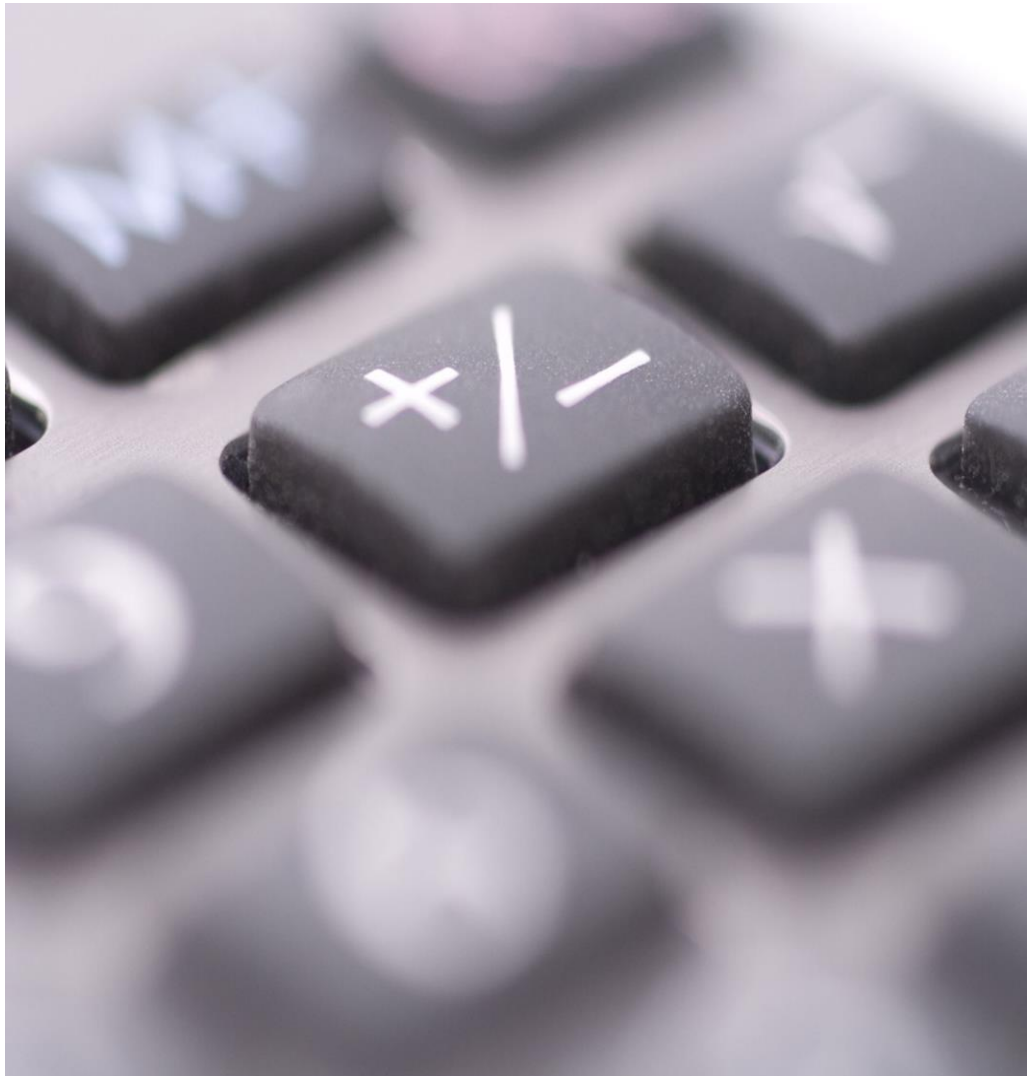
SMART CAMERAS
SMART APPLIANCES
SMART CARS
SMART ELEVATORS
SMART ROBOTS
INTELLIGENT

EXPERT SYSTEMS
SMART QUALITY CONTROL

INFORMATION REVOLUTION

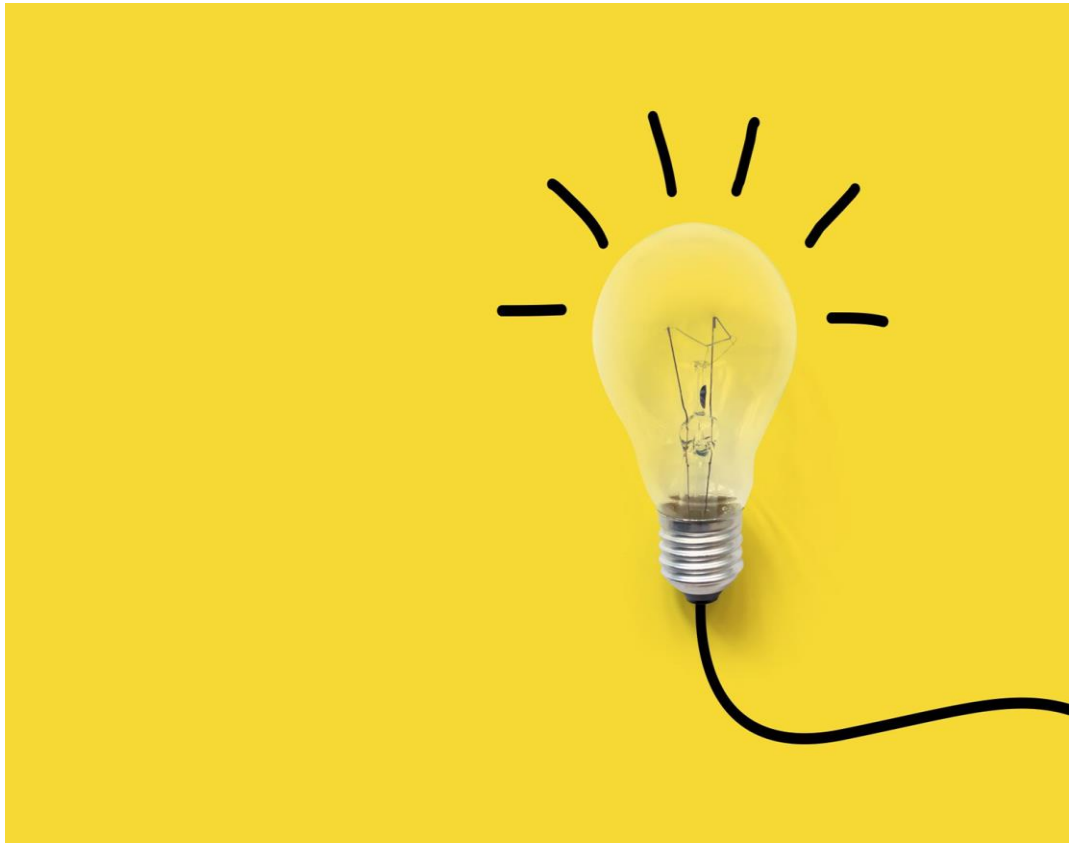
Measure of intelligence: MIQ (Machine Intelligence Quotient)

INTELLIGENT SYSTEMS REVOLUTION



Machine Learning: A Definition

- **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.**



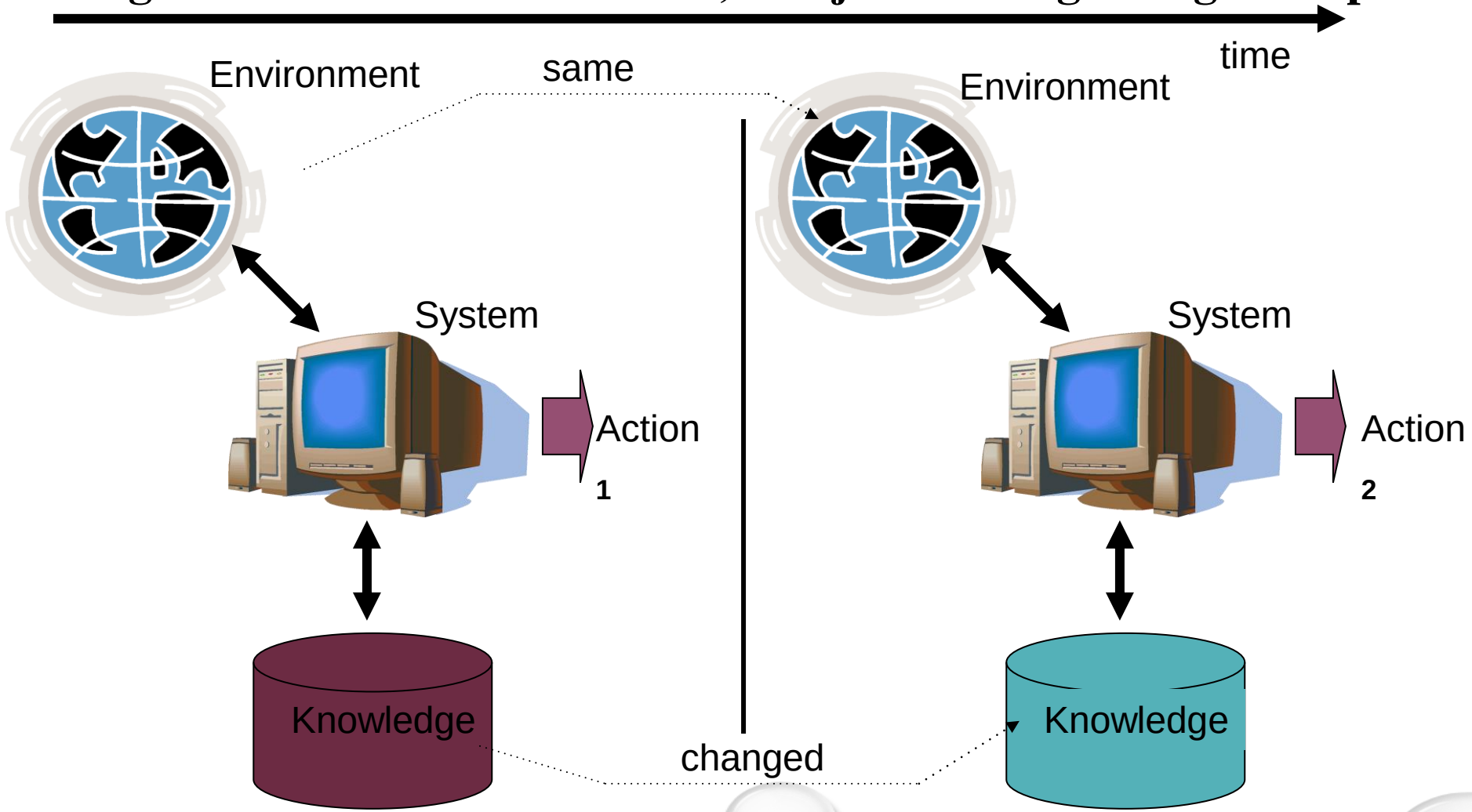
Machine Learning: A Definition

- 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.
- **Idea:** Synthesize computer programs by learning from representative examples of input (and output) data
- **Rationale:**

1. For many problems, there is no known method for computing the desired output from a set of inputs.
2. For other problems, computation according to the known correct method may be too expensive.

What Is Machine Learning?

“Logic is not the end of wisdom, it is just the beginning” --- Spock



What is Machine Learning?

1. **It is very hard to write programs that solve problems like recognizing a face.**
   - **We don't know what program to write because we don't know how our brain does it.**
   - **Even if we had a good idea about how to do it, the program might be horrendously complicated.**
2. **Instead of writing a program by hand, we collect lots of examples that specify the correct output for a given input.**
3. **A machine learning algorithm then takes these examples and produces a program that does the job.**
   - **The program produced by the learning algorithm may look very different from a typical hand-written program. It may contain millions of numbers.**
   - **If we do it right, the program works for new cases as well as the ones we trained it on.**

Some more examples of tasks that are best solved by using a learning algorithm

1. **Recognizing patterns:**
   - Facial identities or facial expressions
   - Handwritten or spoken words
   - Medical images
2. **Generating patterns:**
   - Generating images or motion sequences
3. **Recognizing anomalies:**
   - Unusual sequences of credit card transactions
   - Unusual patterns of sensor readings in a nuclear power plant or unusual sound in your car engine.
4. **Prediction:**
   - Future stock prices or currency exchange rates

Some web-based examples of machine learning

1. **The web contains a lot of data. Tasks with very big datasets often use machine learning**
   - especially if the data is noisy or non-stationary.
2. **Spam filtering, fraud detection:**
   - The enemy adapts so we must adapt too.
3. **Recommendation systems:**
   - Lots of noisy data. Million dollar prize!
4. **Information retrieval:**
   - Find documents or images with similar content.
5. **Data Visualization:**
   - Display a huge database in a revealing way

Why is Machine Learning Important?

- Some tasks cannot be defined well, except by examples (e.g., recognizing people).
- Relationships and correlations can be hidden within large amounts of data. Machine Learning/Data Mining may be able to find these relationships.
- Human designers often produce machines that do not work as well as desired in the environments in which they are used.
- The amount of knowledge available about certain tasks might be too large for explicit encoding by humans (e.g., medical diagnostic).
- Environments change over time.
- New knowledge about tasks is constantly being discovered by humans. It may be difficult to continuously re-design systems “by hand”.

Areas of Influence for Machine Learning

1. **Statistics:** How best to use samples drawn from unknown probability distributions to help decide from which distribution some new sample is drawn?
2. **Brain Models:** Non-linear elements with weighted inputs (Artificial Neural Networks) have been suggested as simple models of biological neurons.
3. **Adaptive Control Theory:** How to control a process with unknown parameters that must be estimated during operation?
4. **Psychology:** How to model human performance on various learning tasks?
5. **Artificial Intelligence:** How to write algorithms to acquire the knowledge humans can acquire, at least, as well as humans?
6. **Evolutionary Models:** How to model certain aspects of biological evolution to improve the performance of computer programs?

Designing a Learning System

- In designing a learning system, we have to deal with (at least) the following issues:

  1. Training experience
  2. Target function
  3. Learned function
  4. Learning algorithm

Training Experience

Issues concerning the training experience:

1. Direct or indirect evidence (supervised or unsupervised).
2. Controlled or uncontrolled sequence of training examples.
3. Representatively of training data in relation to test data.

Target Function and Learned Function

- The problem of improving performance can often be reduced to the problem of learning some particular target function.
- In many cases we can only hope to acquire some approximation to the ideal target function.

Learning Algorithm

In order to learn the (approximated) target function we require:

1. A set of training examples (input arguments)
2. A rule for estimating the value corresponding to each training example (if this is not directly available)
3. An algorithm for choosing the function that best fits the training data

Types of Machine Learning

- **Rote learning** – One-to-one mapping from inputs to stored representation. “Learning by memorization.” Association-based storage and retrieval.
- **Induction** – Use specific examples to reach general conclusions
- **Clustering** – Unsupervised identification of natural groups in data
- **Analogy** – Determine correspondence between two different representations
- **Discovery** – Unsupervised, specific goal not given
- **Genetic algorithms** – “Evolutionary” search techniques, based on an analogy to “survival of the fittest”
- **Reinforcement** – Feedback (positive or negative reward) given at the end of a sequence of steps

In other words the **Types of Machine Learning** may be as follows:

- **Supervised Learning**
  – Classification(pattern recognition)
  – Regression
- **Unsupervised Learning**
- **Reinforcement Learning**

Learning Types

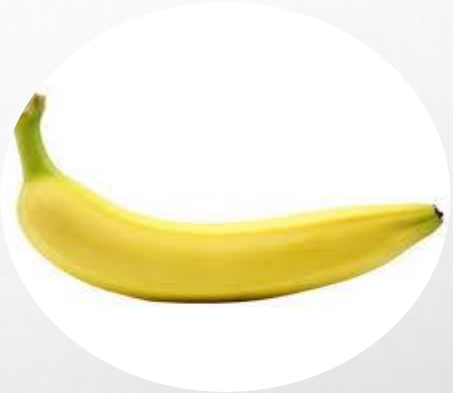
Supervised Learning

Unsupervised Learning

Semi Supervised Learning

Reinforcement Learning

Supervised Learning

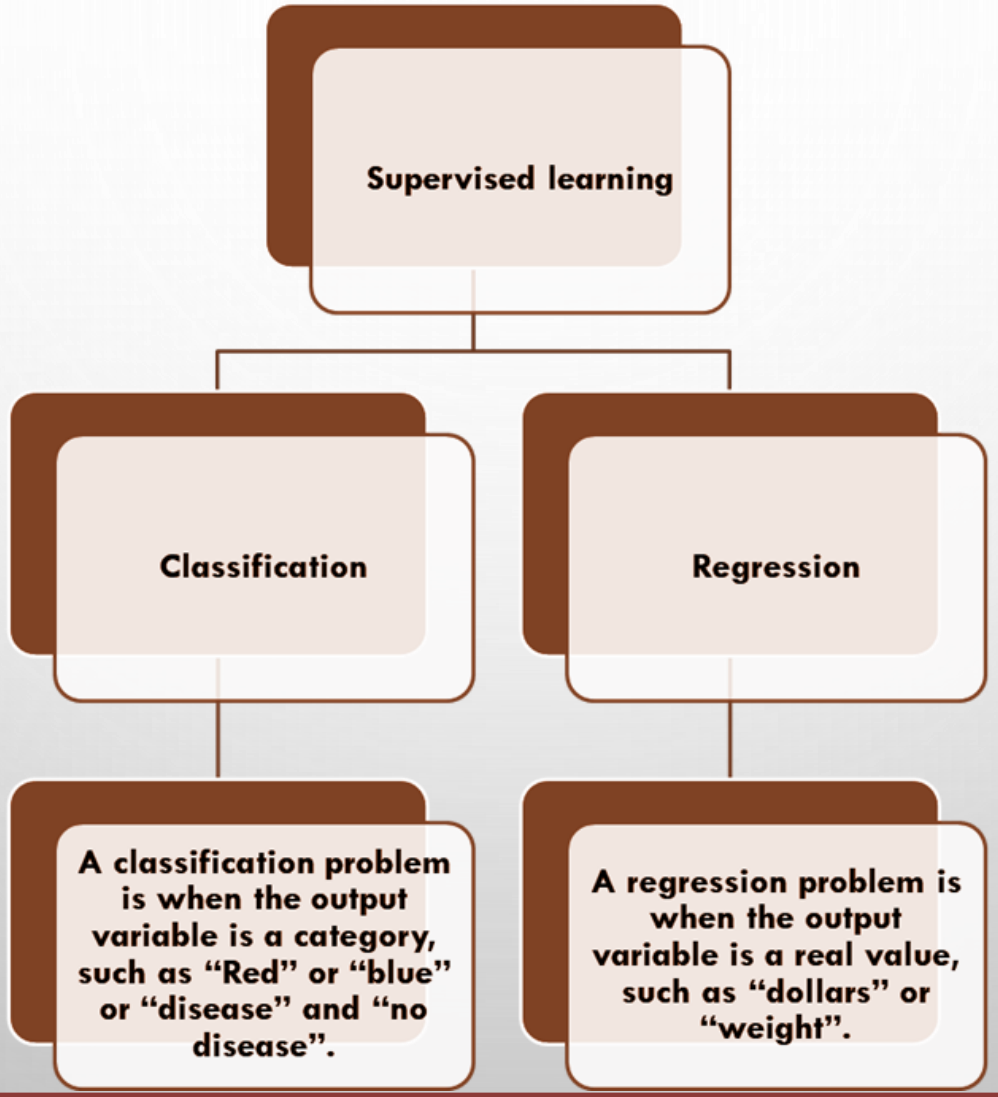


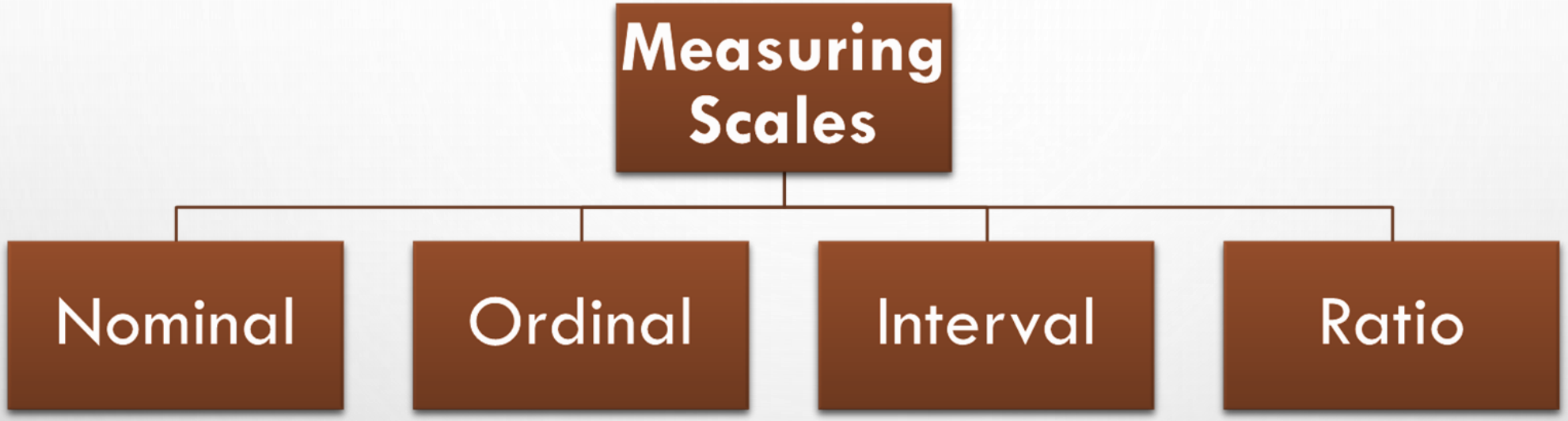
If shape of object is rounded and depression at top having color Red then it will be labelled as –Apple.

If shape of object is rounded and depression at top having color Green then it will be labelled as –Green Apple.

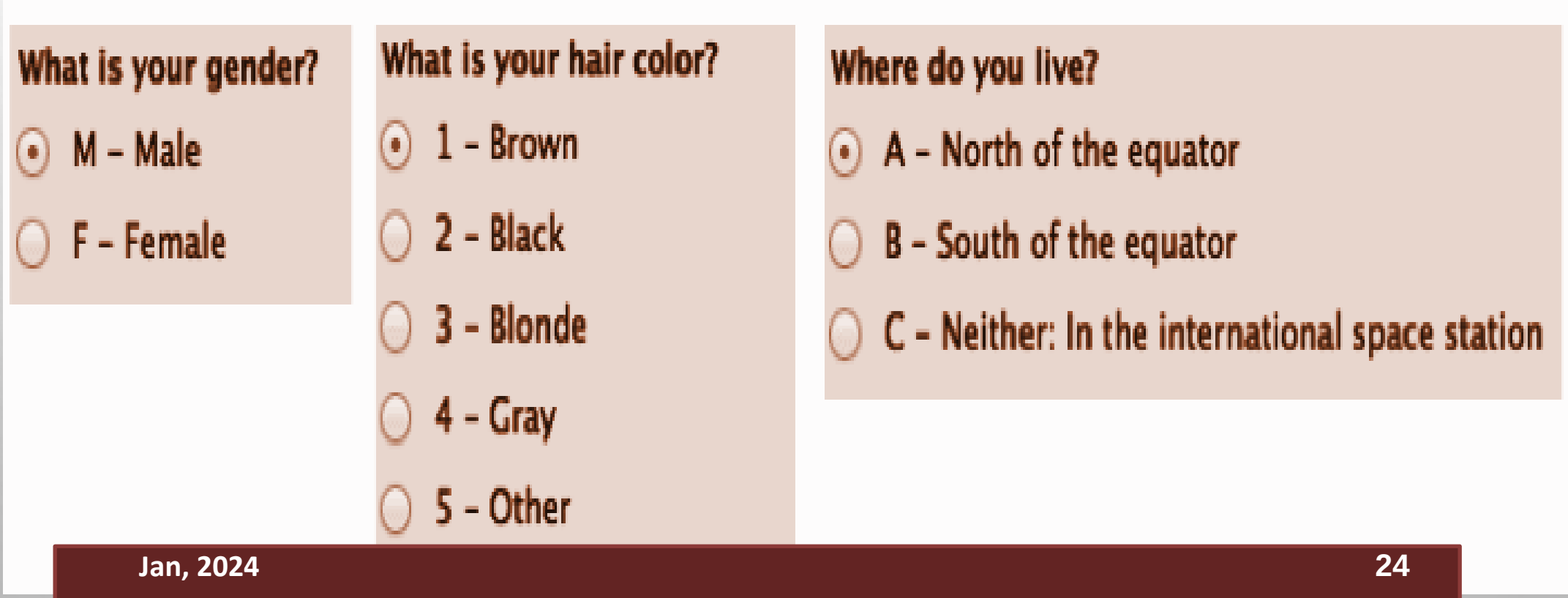
If shape of object is long curving cylinder having color Green-Yellow then it will be labelled as – Banana.

Supervised Learning

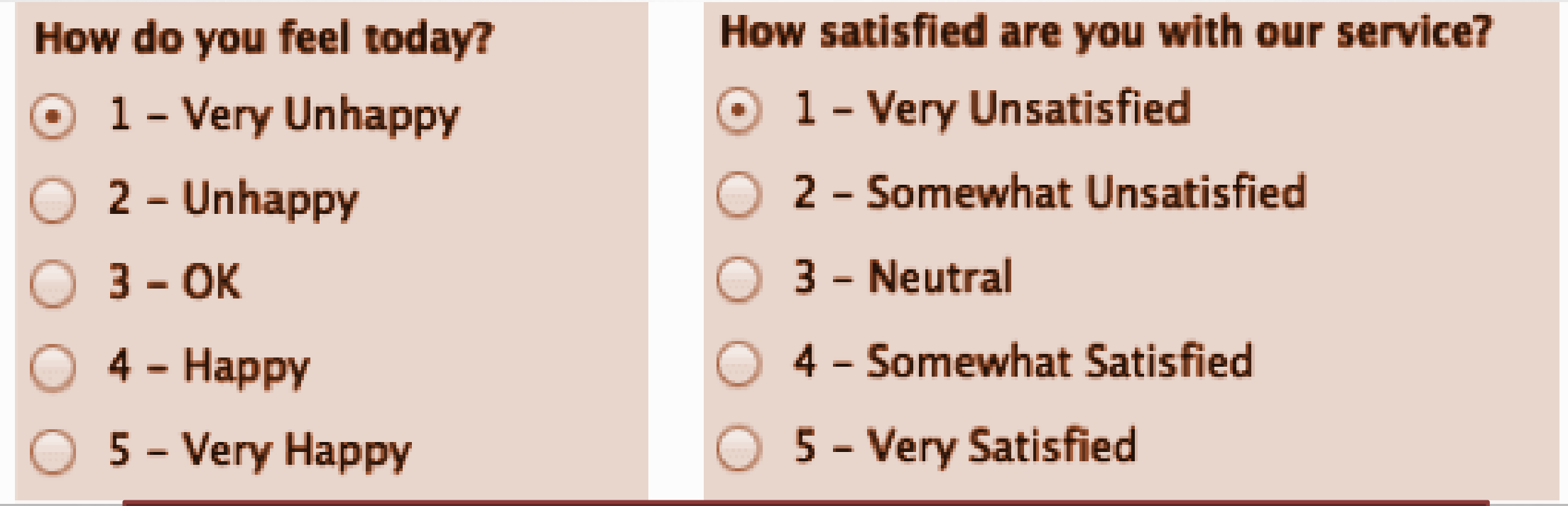



Measuring Scales
Nominal
Ordinal
Interval
Ratio

Nominal scale is a naming scale, where variables are simply “named” or labeled, with no specific order.



Ordinal scale has all its variables in a specific order, beyond just naming them.



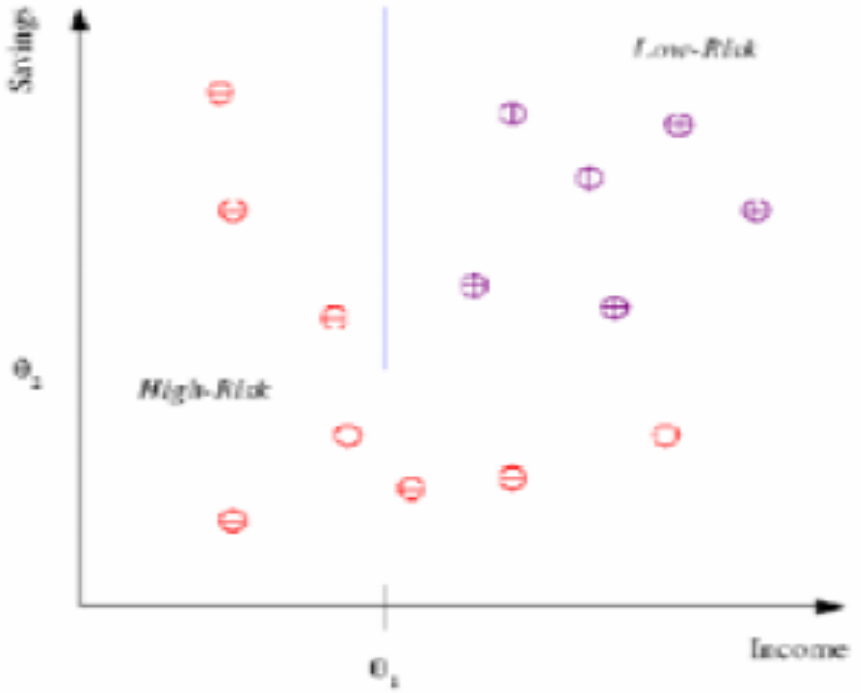
Interval scale	• Offers labels, order, as well as, a specific interval between each of its variable options.
Ratio scale	• Bears all the characteristics of an interval scale, in addition to that, it can also accommodate the value of “zero” on any of its variables.

Supervised Learning:

Classification

- Example: Credit scoring
  Differentiating between low-risk and high-risk customers from their *income* and *savings*

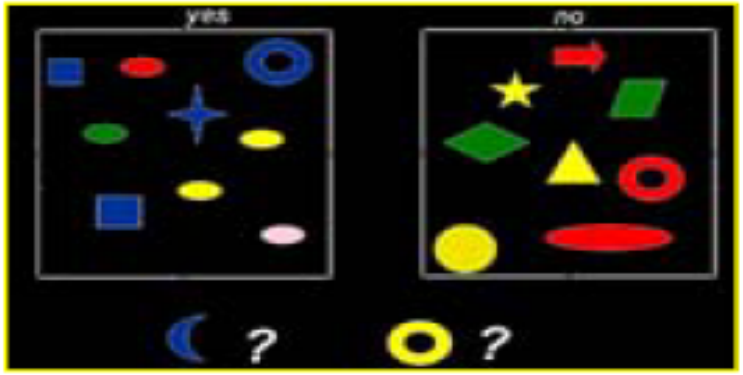
- Input data is two dimensional, output is binary



Discriminant:
IF *income*> θ1AND *savings*> θ2THEN low-risk
ELSE high-risk

Supervised Learning:

Classification



Training Set:

p features (attributes)

n cases

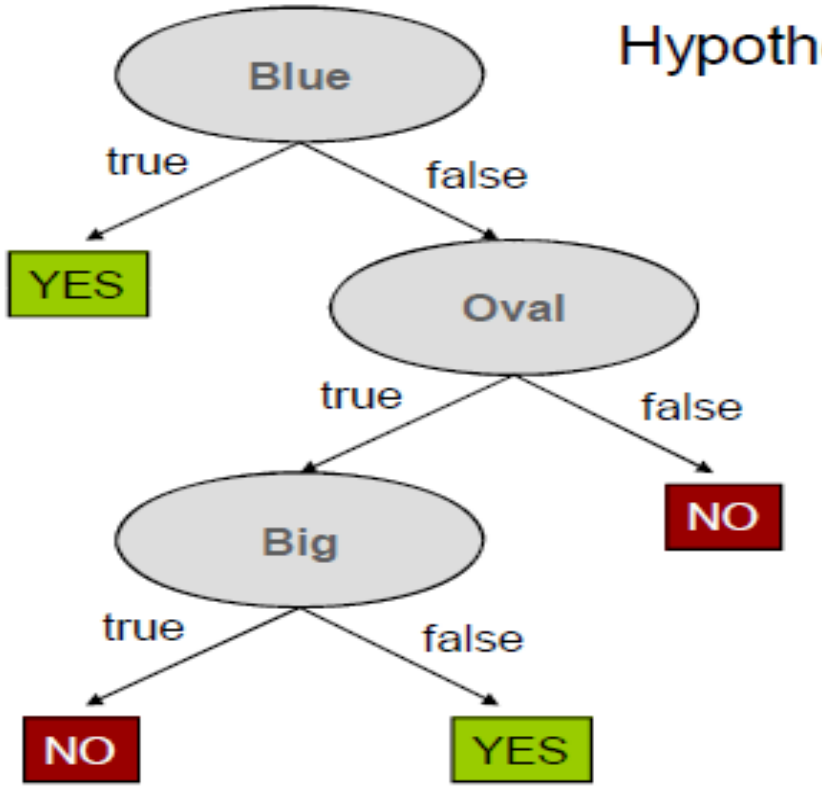
Color	Shape	Size	Label
Blue	Square	Small	Yes
Red	Ellipse	Small	Yes
Red	Ellipse	Large	No

Test Set:

Blue	Crescent	Small	?
Yellow	Ring	Small	?

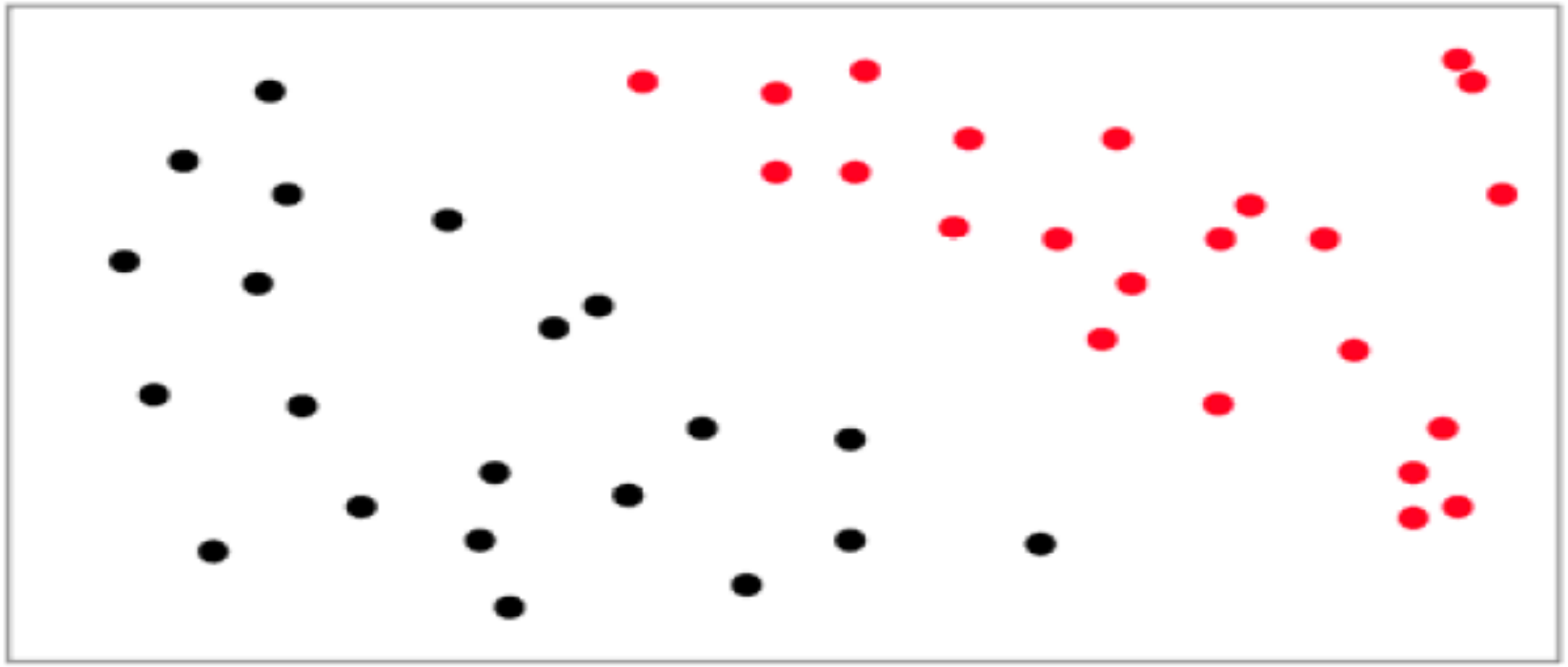
Supervised Learning:

Classification - Decision Tree



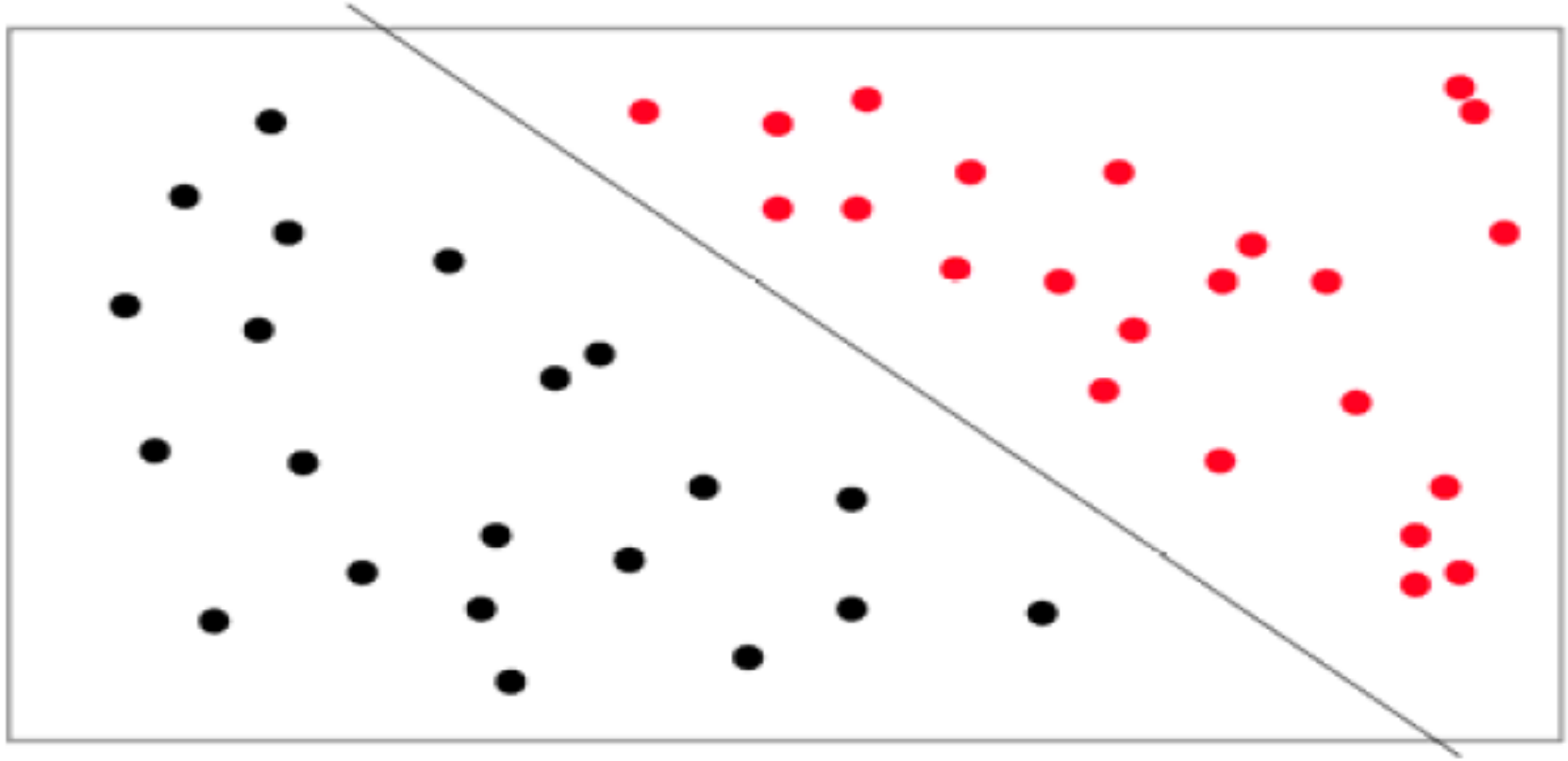
Supervised Learning:

What is the right Hypothesis?



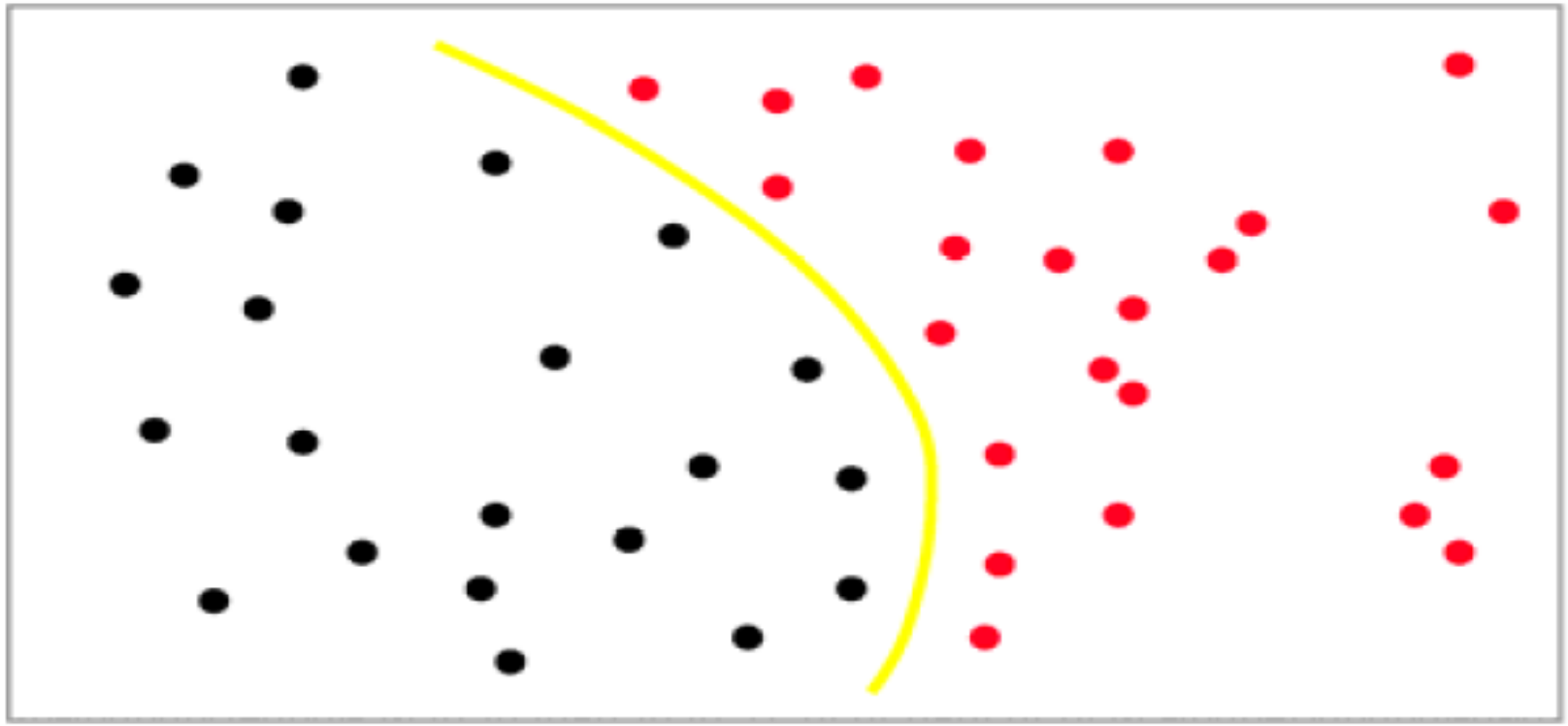
Supervised Learning:

Hypothesis – Linear Separation



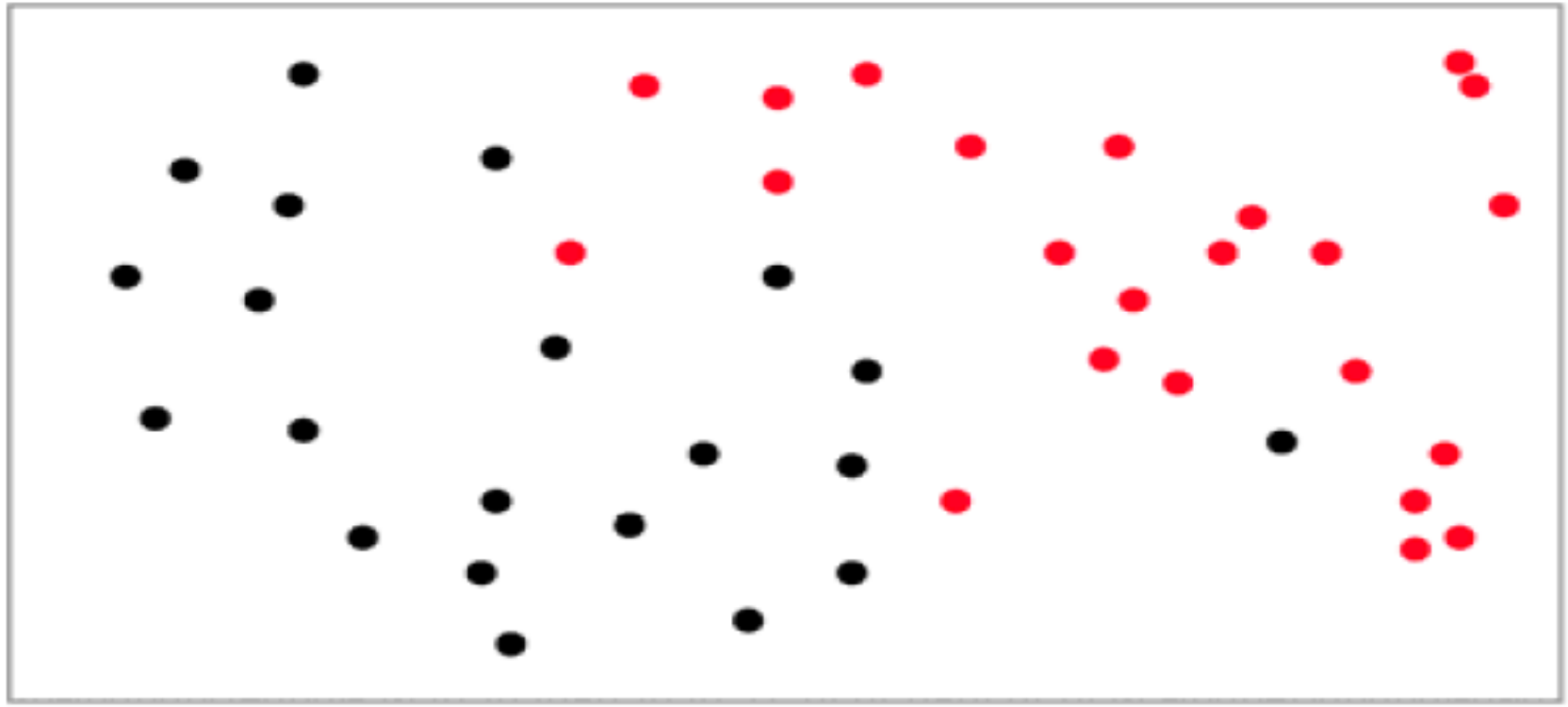
Supervised Learning:

Hypothesis – Quadratic Separation



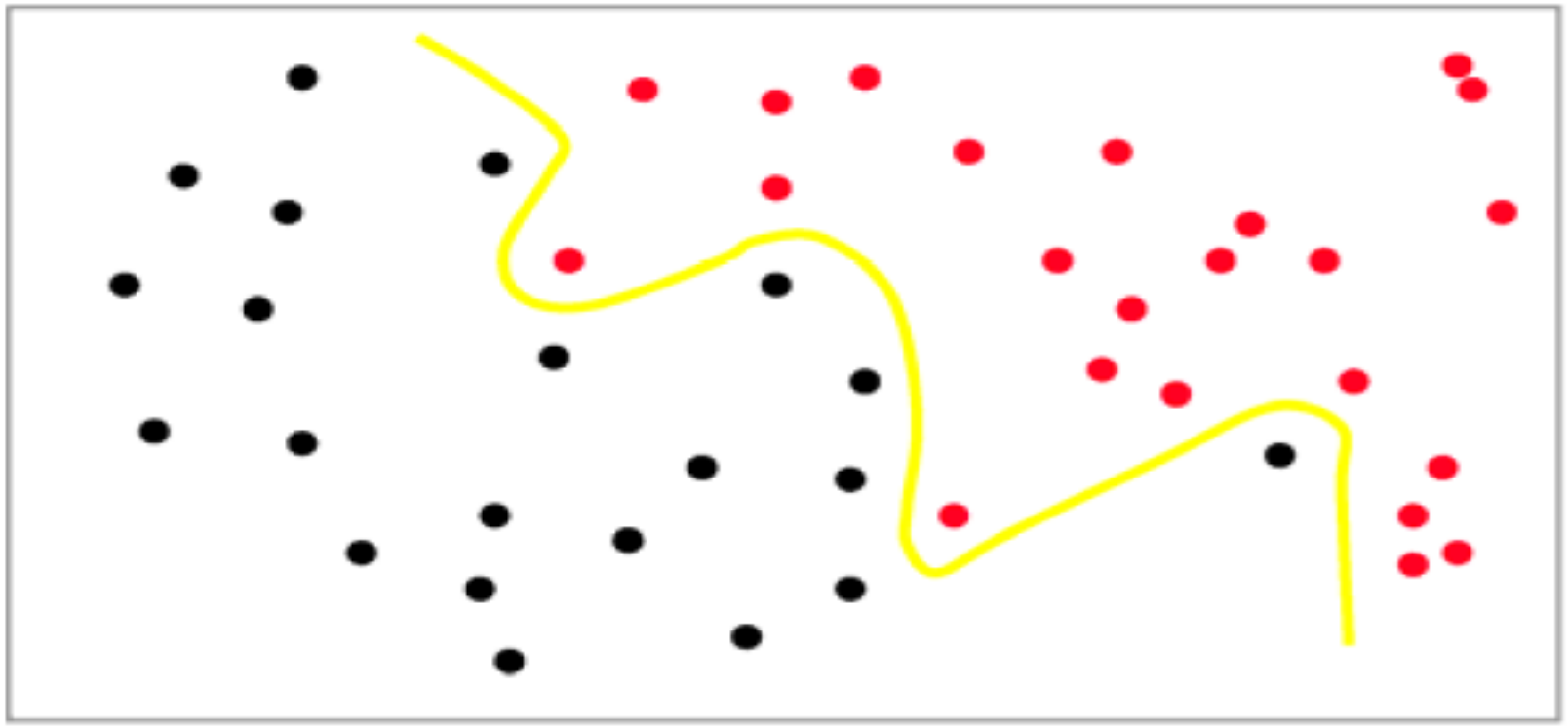
Supervised Learning:

Hypothesis – Noisy/Mislabeled Data



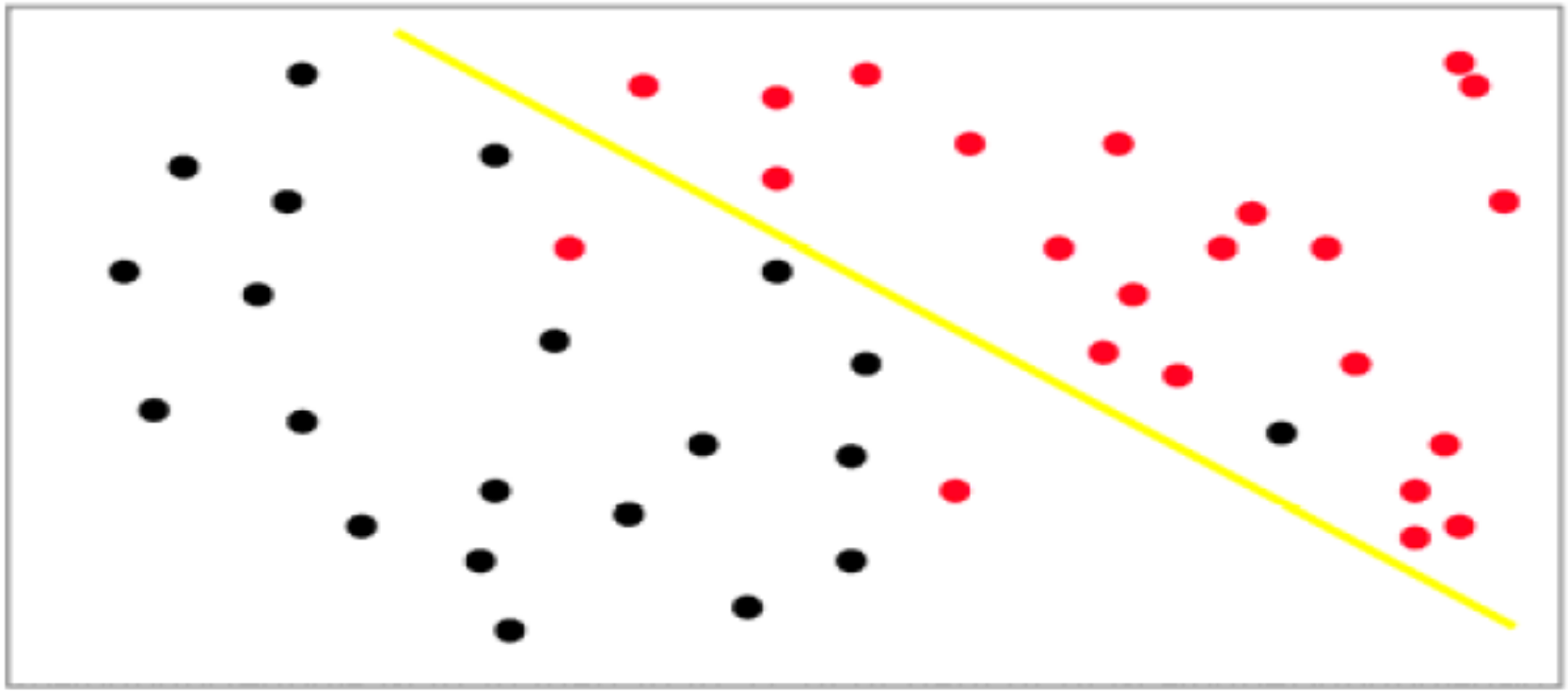
Supervised Learning:

Hypothesis – Overfitting



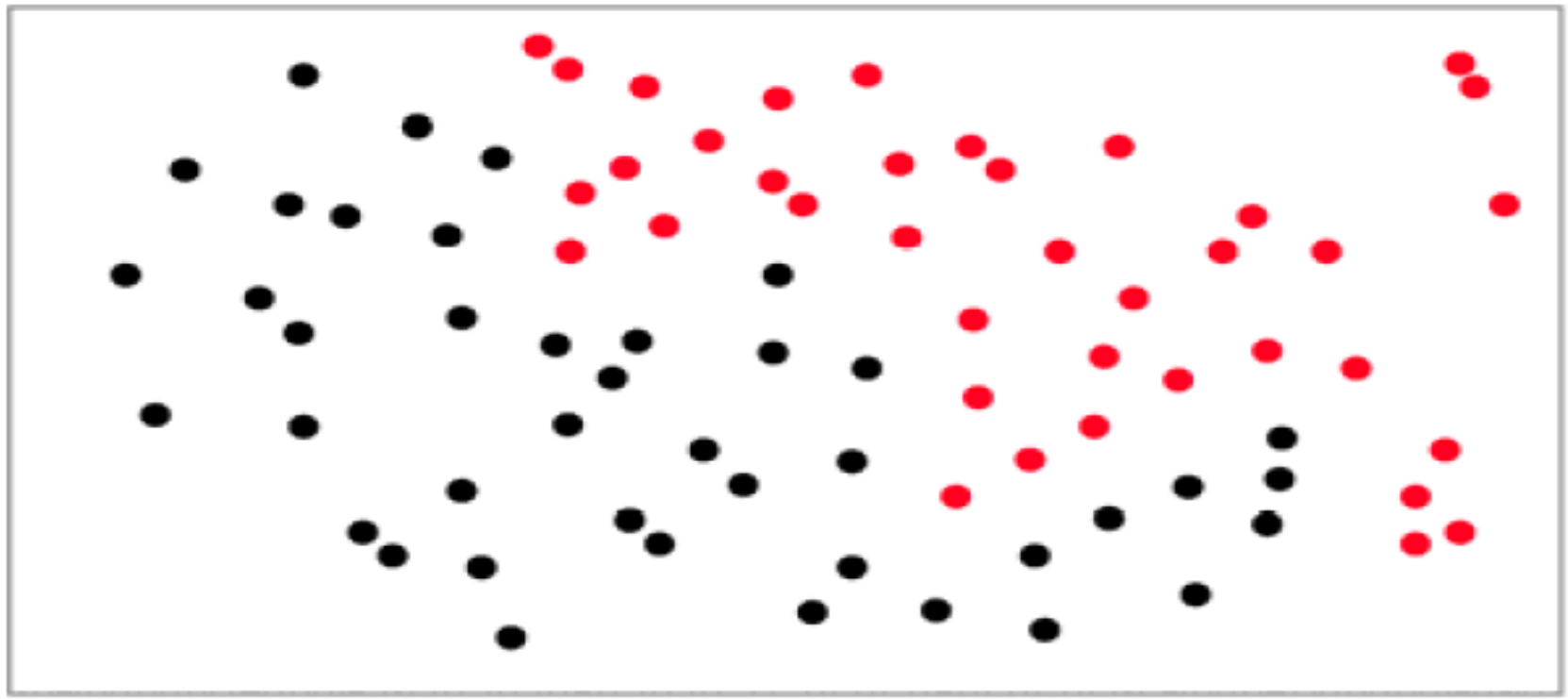
Supervised Learning:

Hypothesis – Underfitting?



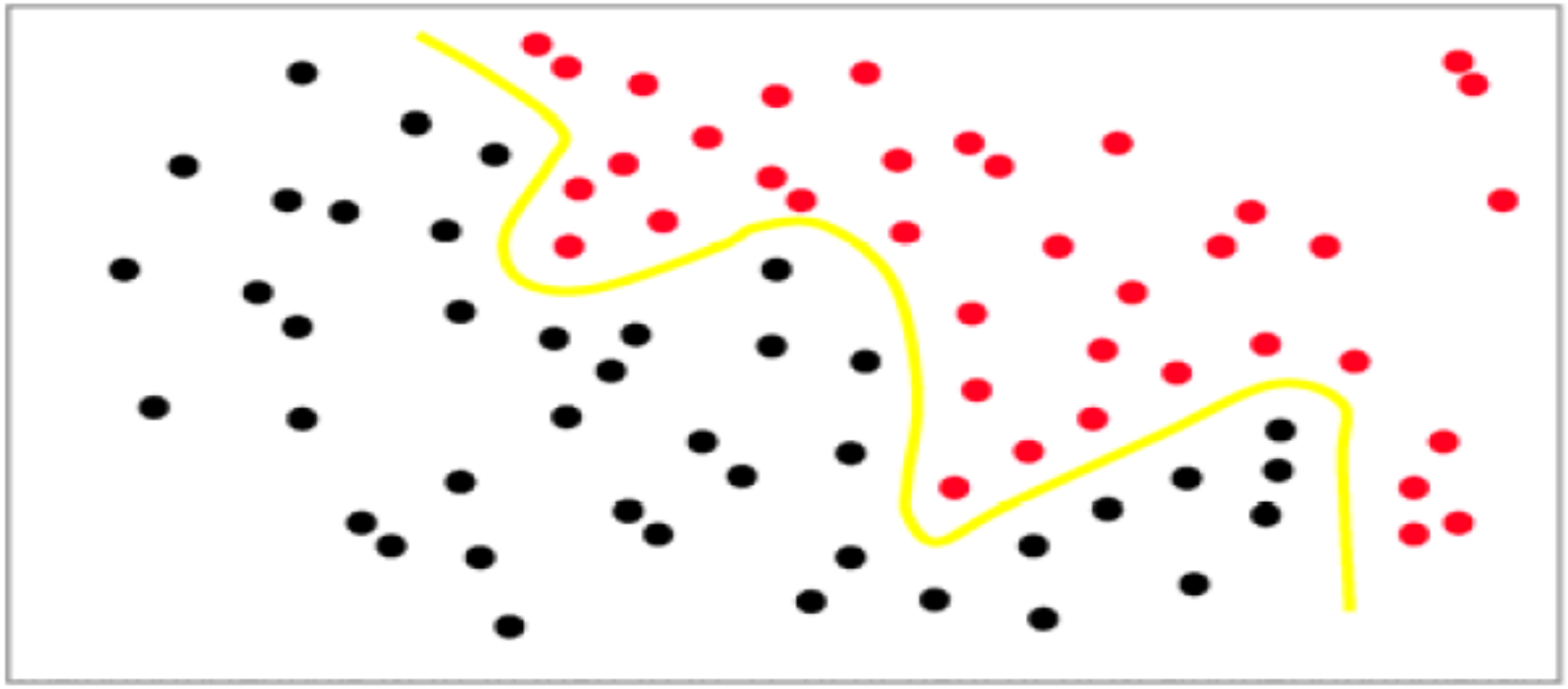
Supervised Learning:

Hypothesis – More data



Supervised Learning:

Hypothesis – More complex

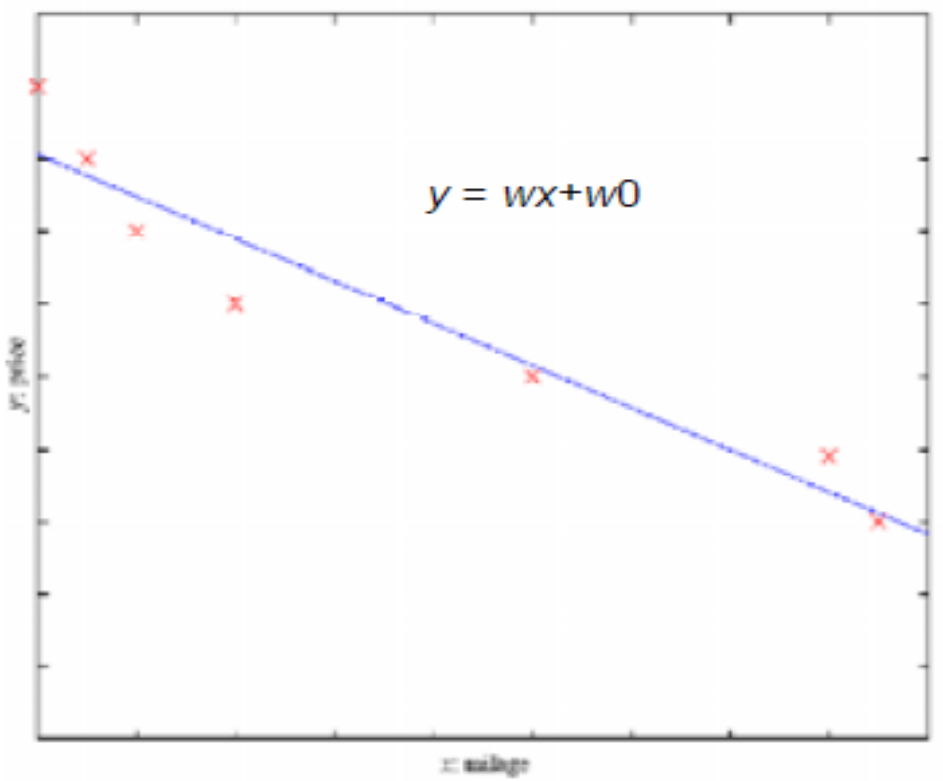


Supervised Learning:

Linear Regression

- Example:
  Price of a used car
  x : car attribute
  y : price

- $y = g\ (x \mid \theta)$
  model:
  $g\ (\ )$
  parameters:
  $\theta = (w, w_0)$

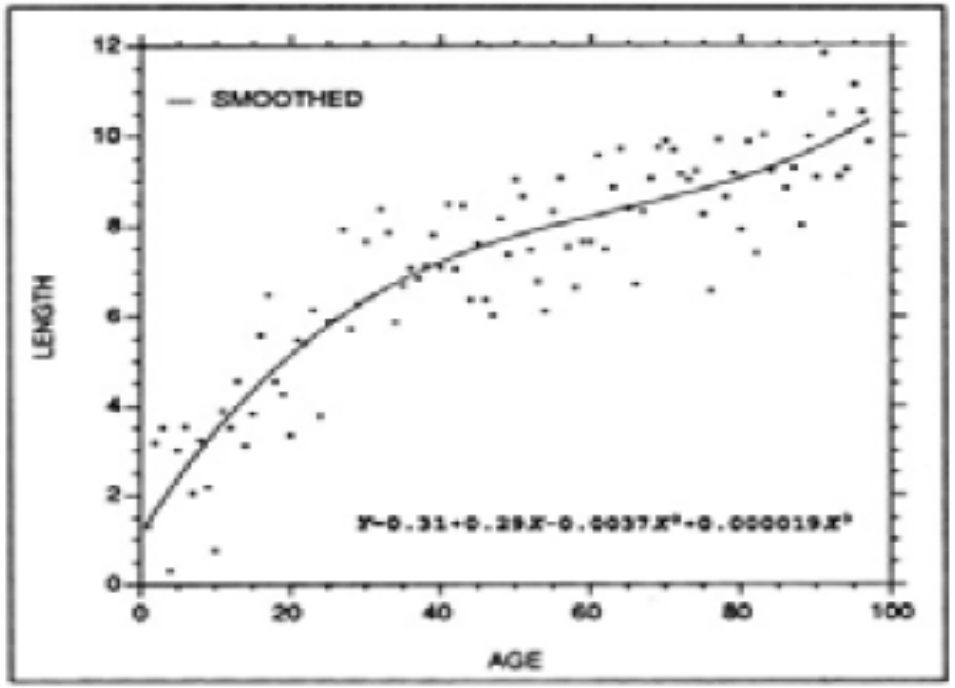


Supervised Learning:

Polynomial Regression

- Example:
  Growth of a species
  x : age
  y : length

- $y = g\,(x \mid \theta)$
  model:
  $g\,(\,)$
  parameters:
  $\theta = (w_3, w_2, w_1, w_0)$



Day 2 - Ungraded Quiz

https://forms.office.com/pages/responsepage.aspx?id=BfhDwM4AVk-HcLw0aPE62rzdPSuGcfdFii_74E63SrpUMlFSODQyWjZYSzBWVzVRRkVVR0owV0ZCTiQlQCN0PWcu

Introduction to simple linear regression

Relation between variables where changes in some variables may “explain” or possibly “cause” changes in other variables.

Explanatory variables are termed the **independent** variables and the variables to be explained are termed the **dependent** variables.

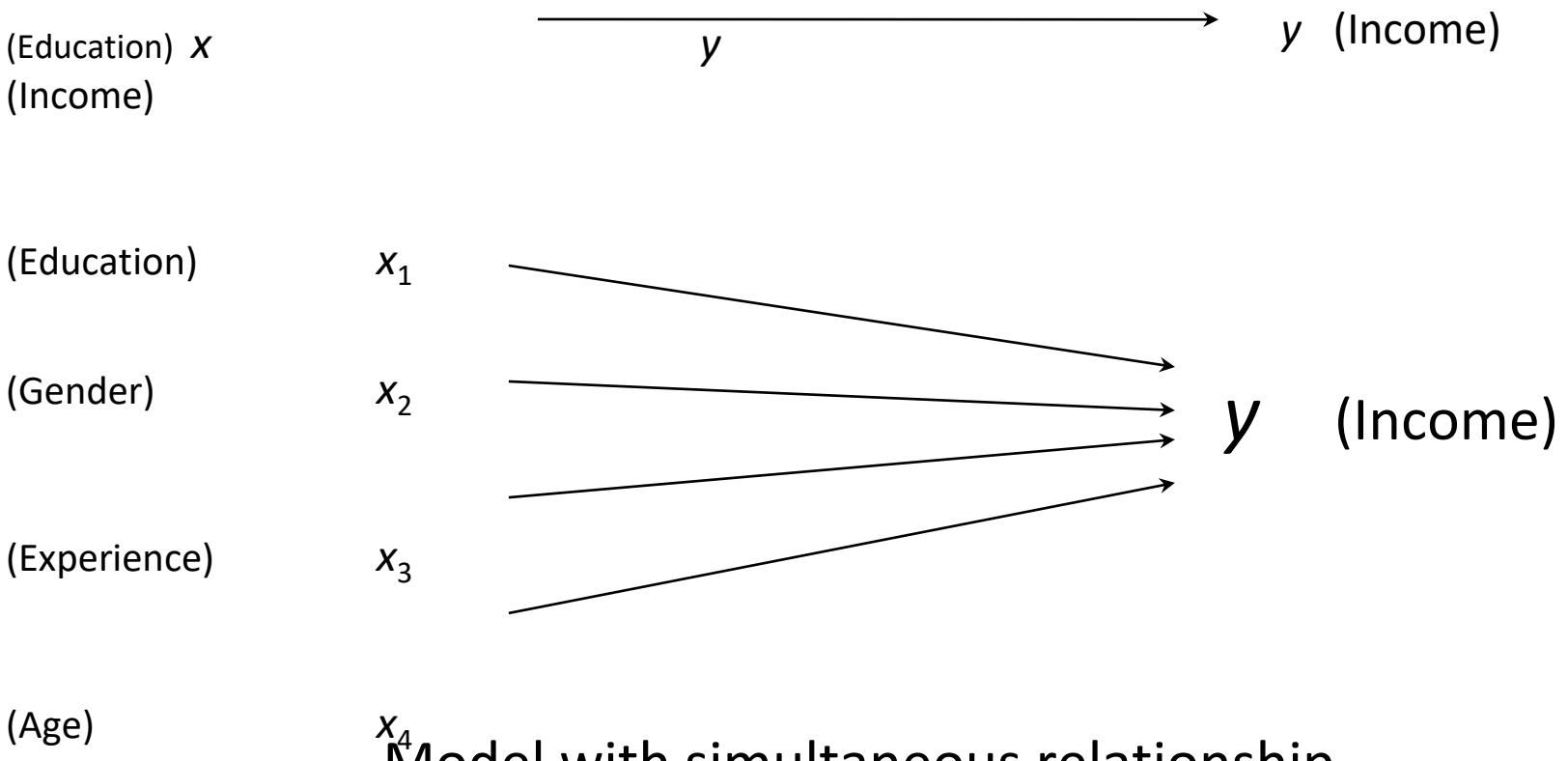
Regression model estimates the nature of the relationship between the independent and dependent variables.

- Change in dependent variables that results from changes in independent variables, ie. size of the relationship.
- Strength of the relationship.
- Statistical significance of the relationship.

EXAMPLES

Dependent variables	Independent variables
Retail price of petrol in Chennai	The price of crude oil.
Employment income	Hours of work, education, occupation, sex, age, region, years of experience, unionization status, etc
Quantity of product sold	Price

Bivariate or simple regression model



Model with simultaneous relationship

Price of wheat ⟷ Quantity of wheat produced

Bivariate or simple regression model

x is the independent variable
y is the dependent variable
The regression model is

$$y = \beta_0 + \beta_1 x + \varepsilon$$

The model has two variables, the independent or explanatory variable, *x,* and the dependent variable *y*, the variable whose variation is to be explained.
The relationship between *x* and *y* is a linear or straight line relationship.
Two parameters to estimate – the slope of the line β_1 and the *y*-intercept β_0 (where the line crosses the vertical axis).
ε is the unexplained, random, or error component. Much more on this later.

REGRESSION LINE

$$y = \beta_0 + \beta_1 x + \varepsilon$$

The regression model is

- Data about x and y are obtained from a sample.
- From the sample of values of x and y, estimates b_0 of β_0 and b_1 of β_1 are obtained using the least squares or another method.

The resulting estimate of the model is

- The symbol is termed "y hat" and refers to the predicted values of the dependent variable y that are associated with values of x, given the linear model.

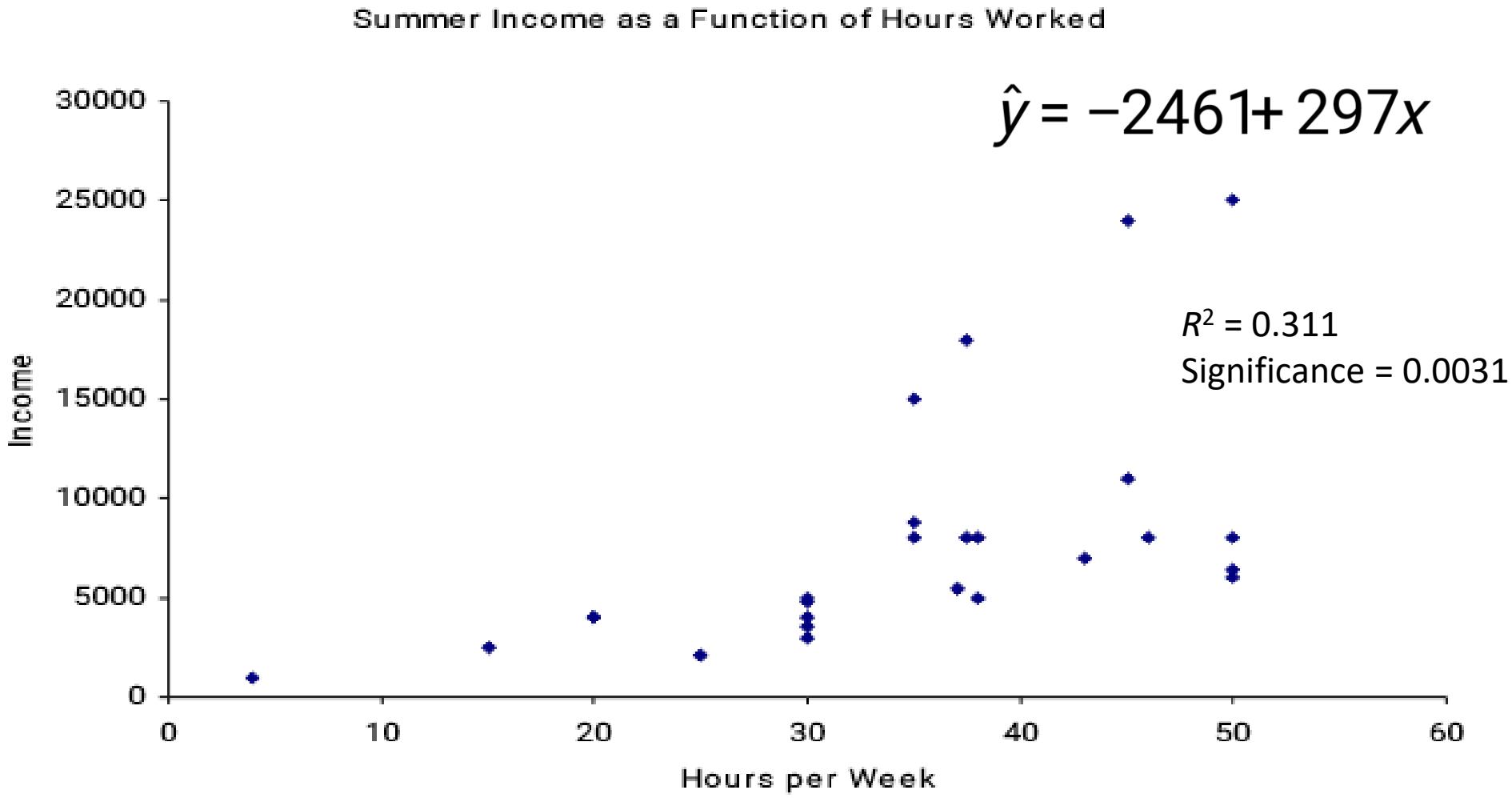
USES OF REGRESSION

- Amount of change in a dependent variable that results from changes in the independent variable(s) – can be used to estimate elasticities, returns on investment in human capital, etc.
- Attempt to determine causes of phenomena.
- Prediction and forecasting of sales, economic growth, etc.
- Support or negate theoretical model.
- Modify and improve theoretical models and explanations of phenomena.

REGRESSION EXAMPLE - DATASET

Income	hrs/week		Income	hrs/week
8000	38		8000	35
6400	50		18000	37.5
2500	15		5400	37
3000	30		15000	35
6000	50		3500	30
5000	38		24000	45
8000	50		1000	4
4000	20		8000	37.5
11000	45		2100	25
25000	50		8000	46
4000	20		4000	30
8800	35		1000	200
5000	30		2000	200
7000	43		4800	30

PLOT OF DATASET



Outliers

Rare, extreme values may distort the outcome.

Could be an error.

Could be a very important observation.

More than 3 standard deviations from the mean.

Dataset and Code

Simple linear regression & Logistic regression code:

https://colab.research.google.com/drive/1OtmxBWhCg7FRbdEvGTOLmKEI26A5R19O#scrollTo=YkaPT7YDquMj

Day 2 - Quiz and Discussion:

https://docs.google.com/presentation/d/1EMFNtric6c64CxXSx2DoS3TtTShxRdLj/edit?usp=share_link&ouid=113983923138295569548&rtpof=true&sd=true

Homework:
https://drive.google.com/file/d/1w6RHTn5fMMA6QFdkxXIAlmQPd8_xal1J/view?usp=sharing

THANK YOU