Week-04-L-02

Select a Statistical Model

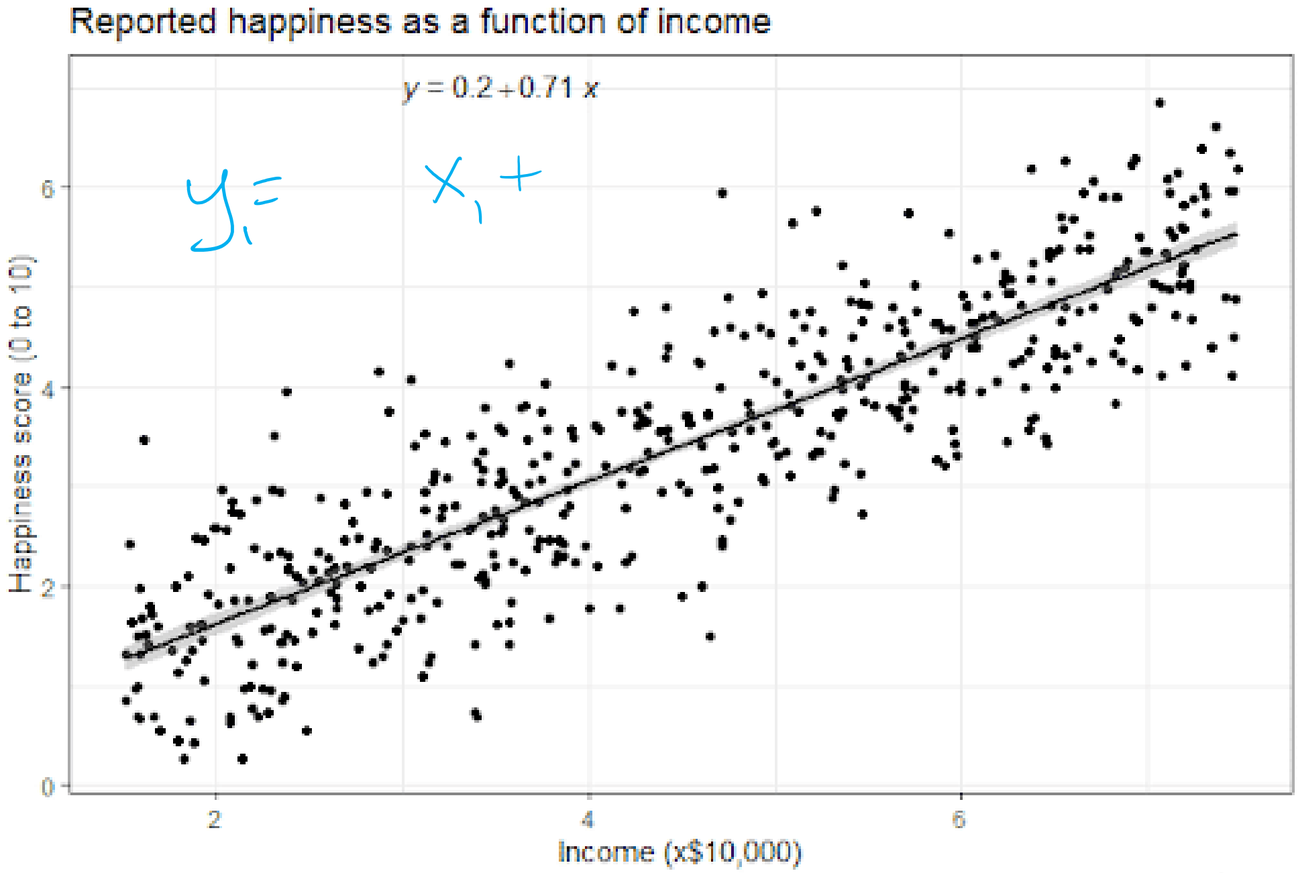
Statistical Modeling

Prof. J. Ramkumar
Department of ME & Design
Indian Institute of Technology Kanpur

Statistical Modeling

- The use of mathematical models and statistical assumptions to generate sample data and make predictions about the real world.
- A statistical model is a collection of probability distributions on a set of all possible outcomes of an experiment.

Plan → Significant → model.
expt



Source: www.heavy.ai

x_1 ⇒ water x_3 = atmosphere
x_2 ⇒ chem

Statistical Modeling

- Supervised

Supervised data modeling will learn the relationship between input and output through labelled training data:

1. Regression model:

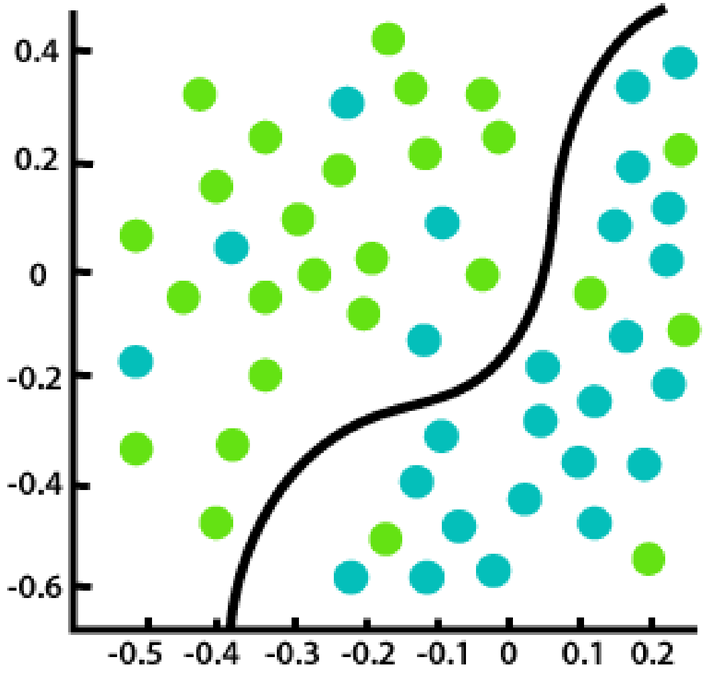
- A type of predictive statistical model that analyzes the relationship between a dependent and an independent variable.
- Common regression models include logistic, polynomial, and linear regression models.

2. Classification model:

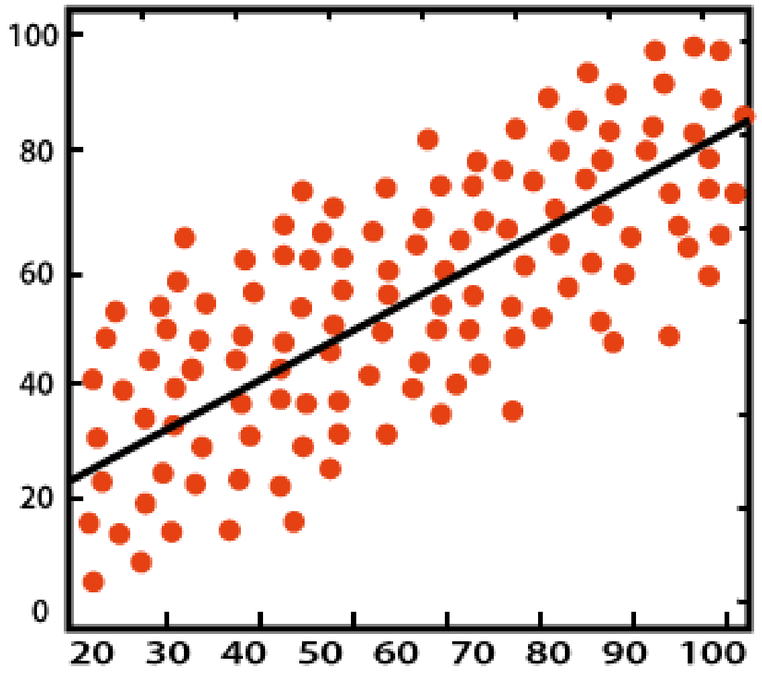
- A type of machine learning in which an algorithm analyzes an existing, large and complex set of known data points.
- This is for understanding and then appropriately classifying the data.

Statistical Modeling

- Supervised



Classification



Regression

Source: www.javapoint.com

Statistical Modeling

- Unsupervised

Unsupervised modeling (using unlabeled data) techniques include clustering algorithms and association rules:

1. K-means clustering:

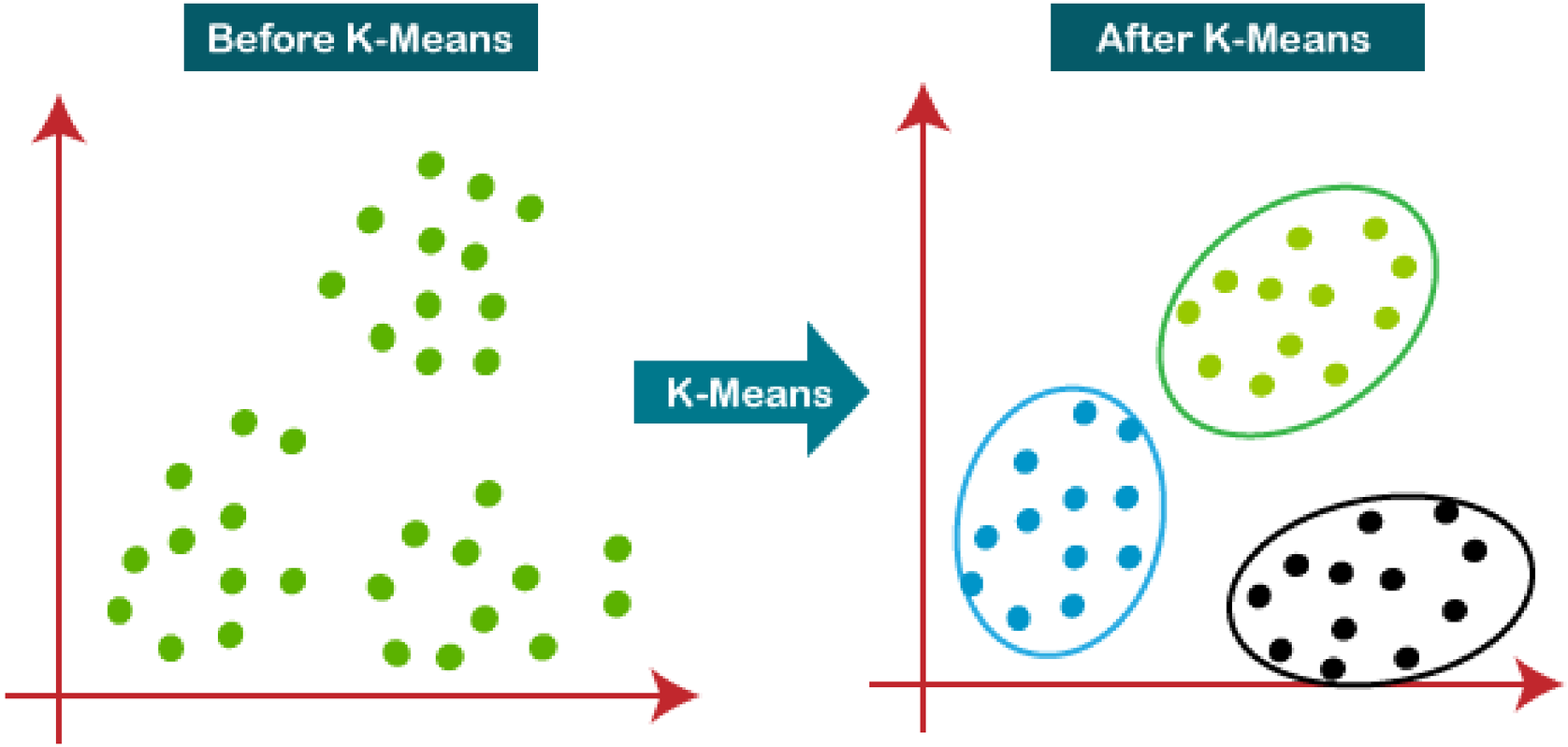
- Aggregates a specified number of data points into a specific number of groupings based on certain similarities.

2. Reinforcement learning:

- An area of deep learning that concerns models iterating over many attempts.
- It rewards the moves that produce favorable outcomes and penalizing steps that produce undesired outcomes.
- Thus training the algorithm to learn the optimal process.

Statistical Modeling

- Unsupervised



Source: www.javapoint.com

Statistical Modeling

- parametric vs nonparametric

There are three main types of statistical models:

1. Parametric: ⇐ Probability distributions

(1) A family of probability distributions that has a finite number of parameters.

2. Nonparametric:

(3) Models in which the number and nature of the parameters are flexible and not fixed in advance.

63 60 - 70

3. Semiparametric:

(2) The parameter has both a finite-dimensional component (parametric) and an infinite-dimensional component (nonparametric).

Statistical Modeling
- vs Mathematical Modeling

Similarity:

- Mathematical modeling also translates real-world problems into tractable mathematical formulations whose analysis provides insight, results and direction useful for the originating application.

Dissimilarity:

- Unlike statistical modeling, mathematical modeling involves static models that represent a real-world phenomenon in mathematical form.

- Once a mathematical model is formulated, it does not necessitate change.
- Statistical models are flexible and, with the aid of machine learning, can incorporate new, emerging patterns and trends, and will adjust with the introduction of new data.

Statistical Modeling - vs Machine Learning

- Machine learning is a subfield of computer science and artificial intelligence that involves building systems that can learn from data rather than explicitly programmed instructions.

- Machine learning models seek out patterns hidden in data independent of all assumptions, therefore predictive power is typically very strong.

- Machine learning requires little human input and does well with large numbers of attributes and observations.

Thank you