CS 360/530: Foundations of Machine Learning

Spring 2020

Clint P. George

January 7, 2020

Course overview

This course will give a broad introduction to foundations of machine learning and statistical learning.

Prerequisites. Familiarity with the following is required.

- Basic computer programming—we use R/Python.
- Probability theory (CS 215, MA 605)
- Multivariable calculus and linear algebra (MA 105, MA 106, EE 611)

If you do not have the necessary background, please meet me before registering this course.

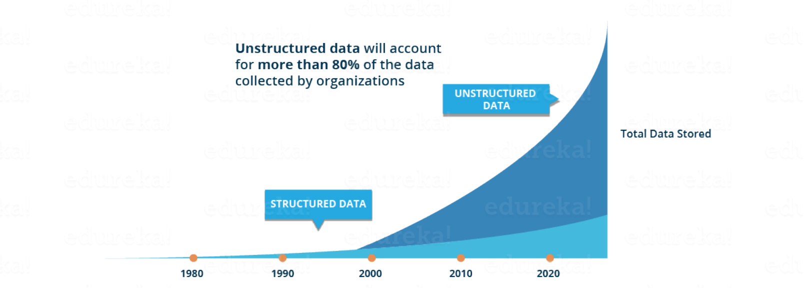
Course page and other resources

- `www.iitgoa.ac.in/~clint/courses/ml-spring-2020.html`
- We use Google Classroom
- Readings will be assigned for every lecture.

Data science

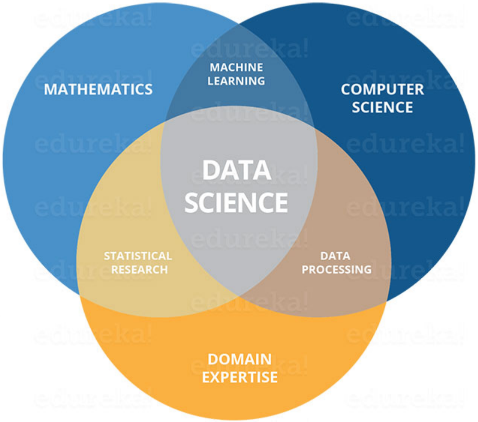
Data science—a field of study that aims to use a scientific approach to extract meaning and insights from data

Why do we need data science?—edureka.co



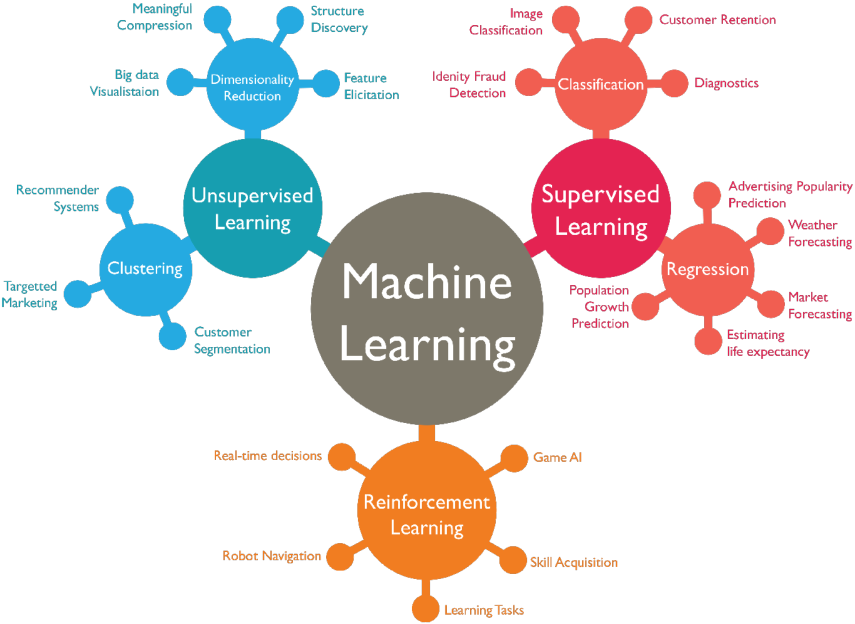
- sources: financial logs, text files, multimedia forms, sensors, and instruments
- Simple BI tools are incapable of processing the data

Skills required for data science



(edureka.co)

- Data science: focus is on the **data**
- Machine learning: focus is on **learning methods**


Machine Learning
Unsupervised Learning
Dimensionality Reduction
Meaningful Compression
Structure Discovery
Big data Visualistaion
Feature Elicitation
Clustering
Recommender Systems
Targetted Marketing
Customer Segmentation
Supervised Learning
Classification
Image Classification
Customer Retention
Idenity Fraud Detection
Diagnostics
Regression
Advertising Popularity Prediction
Weather Forecasting
Population Growth Prediction
Market Forecasting
Estimating life expectancy
Reinforcement Learning
Real-time decisions
Game AI
Robot Navigation
Skill Acquisition
Learning Tasks

Data science — outline

Learn how to load data into **R** (or Python), get it into the most useful structure, transform it, visualize it, and model it.

- Data visualization—e.g. via `ggplot2` in **R** and Exploratory data analysis
- Expected value, variance, the Central Limit Theorem
- Hypothesis Testing—a method that is used in making statistical decisions using experimental data. E.g. A/B testing
- High-dimensional space and problems, Dimensionality reduction—e.g. Principal Component Analysis (PCA)

Machine learning — outline

Supervised learning

- Linear regression, logistic regression, Perceptron (review)
- Generative learning algorithms, Gaussian discriminant analysis
- Maximum likelihood estimation (MLE)
- Support Vector Machines (SVMs)

Machine learning in practice

- Bias–Variance tradeoff and error analysis
- Regularization and model selection
- Experimental evaluation of learning algorithms, cross-validation
- Learning Theory, Generalization errors + model selection, VC dimension — if time permits.

Machine learning — outline

Unsupervised learning

- Mixture models and mixture of Gaussians
- The expectation maximization (EM) algorithm
- Probabilistic topic models

Deep learning — outline

- Multilayer neural networks
- Back-propagation algorithm
- Auto-encoders