Machine Learning– 40 Hours

Overview

Machine learning is becoming huge business in the days of big data and huge web sites. This course will introduce the participants to data science and machine learning in particular reviewing and training using the most popular algorithms. The course will teach machine learning using python and will emphasize the stengths and weaknesses of the most popular algorithms.

Notes

- This course does not include python basics which is a pre-requisite to this course.
- This course does not include deep learning which should be taken after this course.

Exercises

Using Python: numpy, scipy, pandas, scikit-learn.

Intended Audience / Who should attend

- Programmers or Exact Sciences people who wish to understand and utilize machine learning.
- Life science or social science people who has an affinity to computing and want to utilize machine learning in their research.
- Web site and mobile application developers who would like to use their data to do various predictions and analysis.

Prerequisites

- Background in mathematics or math affinity.
- Background in computing.
- Python programming at a basic level.
- Stastical affinity is a plus but is not required.

After attennding the course attendees should be able to

- Use Python to do data analysis and visualization.
- Understand the differneces between the various ML algorithms, their strengths and weaknesses and what algorithm to apply when confronment with a problem.
- Apply common machine learning algorithms to predict various results.
- Understand the machine learning life cycle of a real world project.

Syllabus

- **Data Science**
    - Introduction
    - What does "understanding data" mean?
    - Data Analytics introduction
    - Data concepts: Data bias, data holes, outliers, collection bias, accuracy bias
    - Basic statistics review

- **Practical data science**
    - Mathematical computing with python (numpy)
    - Scientific computing wiht python (scipy)
    - Data manipulating with python (pandas)
    - Data visualization with python (mathplotlib)

- **Machine learning overview**
    - Introduction
    - Theory
    - What cant and can't you learn?
    - Supervised learning vs non supervised learning
    - Classficatoin vs prediction
    - Overfitting, Underfitting
    - Size of data
    - Data feature extraction and creation
    - A real world machine learning cycle
    - Creating final deliverables
    - Machine learning with python (scikit-learn)

- **The algorithms**
  - Supervised learning
    - K-neighbours
    - Linear/Logistic regression
    - Decision trees
    - Naïve Bayes analysis
    - Linear Discriminant Analysis
    - Support Vector Machines
    - Stochastic Gradient Descent
    - Logistic Regression
    - Ensemble methods
      - Bagging.
      - Random Forrest.
  - Non-Supervised learning
    - K-means clustering
    - Association rules

- **Unstructured Data Analytics**
  - Sources of unstructured data.
  - Web scraping with BeautifulSoup
  - Natural language processing in python (scikit-learn)