

Introduction to Machine Learning

Ryerson Zone Learning Analytics School and
Business Career Hub

Overview

- Introduction
- Linear Regression
- Classification
- Next Steps

Introduction

- Why Machine Learning?
 - Utilize Big Data
 - Hardware Improvements
- Popular/Common examples
 - Recommendation Systems
 - Image Classification
 - Nature Language Processing
- Intuitions
 - Mapping relationships between Inputs (features) and Outputs (target) of the data
 - Allow algorithms to find relationships within the data

A Few Definitions

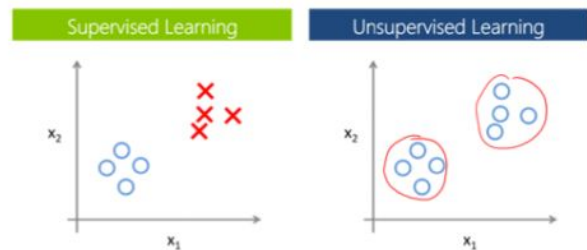
- Supervised vs. Unsupervised Learning ***

- Labelled data vs unlabelled data
- A data points label is its class
- We will only be using supervised learning

- Python

- High-level programming language used for a many data-science and machine learning applications
- Module
 - Open source code that may be imported to a workspace to avoid redundancy of rewriting code that is used often

Supervised vs Unsupervised Learning



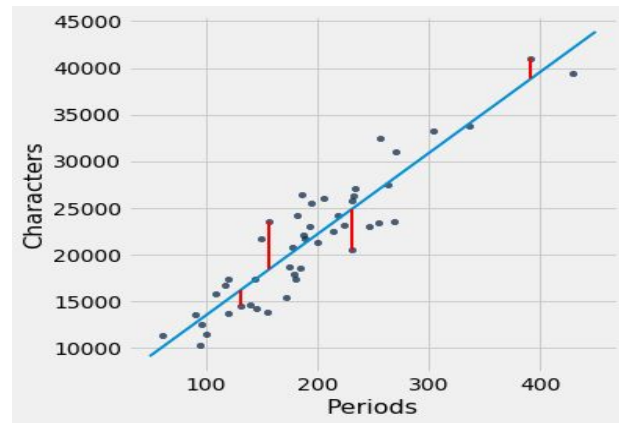
A Few Definitions

- Model Evaluation

- Determine how well we're doing at predicting values
- Mean Square Error used for Regression
 - Sum of all deviations from our predictions
- Cross entropy used for Classification
 - Math will not be covered in the tutorial

- Training/Validation/Test Sets

- Training Set
 - Fed to our model to learn the relationships between the features in the data
- Validation Set
 - Used for model selection. Trained models generate predictions that are tested against the validation set to gauge their effectiveness.
- Test Set
 - Once the best model is found, it is tested again to ensure it is robust and not overfit.



Linear Regression

- Common method of predictive analysis
- Can be evaluated a few different ways:
 - R squared
 - **Machine Learning (LMS - Least Mean Squares)**
- Used to predict target values that are continuous
 - E.g. Housing Prices, stock prices, Temperature approximation.
- Goal: Find the linear equation that bests fits the data points
 - Equation will take the form of $y = Wx + b$
 - Line that fits the data best will have the lowest MSE
- Start coding in collaboratory !
 - <https://colab.research.google.com/drive/1fS9Ok6fAkopLrhsvwaP20WNZoAlnlxYQ>

Break

- Ask Questions!

Classification

- Used for classifying data into discrete classes
 - E.g: Dog Breeds, Plant types, etc.
- Goal: Predict the label of a data point from its features
- Multiple different classification algorithms to choose from for classification purposes
 - Logistic Regression (Binary Classification)
 - Linear Discriminant Analysis (LDA)
 - K - Nearest Neighbors (KNN)
 - Decision Tree Classifier
 - Gaussian Naive Bayes
 - Support Vector Machines (SVM)
- Back to collaboratory!
 - <https://colab.research.google.com/drive/1fS9Ok6fAkopLrhsvwaP20WNZoAlnIxYQ>

Additional Resources

- Massive Open Online Courses (MOOC)
 - Coursera
 - “Machine Learning” and “Deep Learning” by Andrew Ng of Stanford University
 - Fast.ai
 - EDX
 - CS229 - Stanford U.
- Research Papers
 - Arxiv.org
 - jmlr.org