

Hands-on Machine Learning with Scikit-Learn and TensorFlow

CONCEPTS, TOOLS, AND TECHNIQUES TO BUILD INTELLIGENT SYSTEMS

History:

- ▶ Machine Learning has been around for decades.
- ▶ Optical Character Recognition (OCR).
- ▶ Spam Filter-First application (1990s)

Road Map

- ▶ Supervised v/s unsupervised learning
- ▶ Online v/s batch learning
- ▶ Instance based v/s model based learning.
- ▶ Work flow of the typical ML project.
- ▶ Main challenges you may face
- ▶ Evaluate and fine-tune a ML system.

Book link:

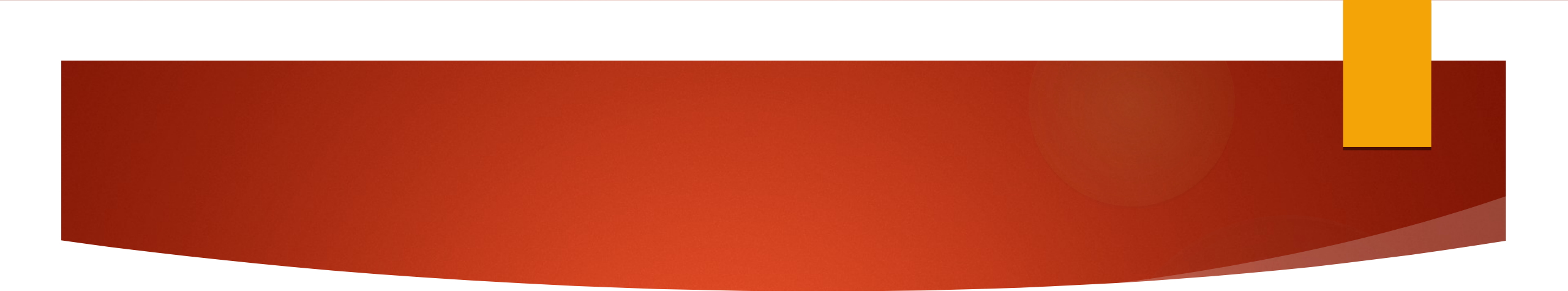
What is Machine Learning:

- ▶ Science of programming computers so they can learn from data
- ▶ Field of study that gives computers the ability to learn without being explicitly programmed. [Arthur Samuel, 1959]

Engineering-oriented one:

A computer program is said to learn from experience E , with respect to some task T and some performance measure P , if its performance on T , as measured by P , improves with experience E .

[Tom Mitchell, 1997]

- 
- ▶ Training set: Examples that are use to learn.
 - ▶ Training instance: Each training example.
 - ▶ Accuracy: performance measure.
 - ▶ Testing set: ?

Why use Machine Learning:

- ▶ Traditional programming will likely become a long list of complex rules.
- ▶ Pretty hard to maintain.
- ▶ Automatically learns.
- ▶ Program is much shorter, easier to maintain and most likely more accurate.

Traditional Programming v/s ML

Machine Learning is great for :

- ▶ To simplify code and perform better.
- ▶ Complex problems where there is no good solution.
- ▶ Fluctuating environment.
- ▶ Getting insight about complex problems and large amount of data.

Types of Machine Learning Systems

- ▶ Machine learning systems can be categorized based on :
- ▶ Trained with human supervision.
- ▶ They can learn incrementally on the fly.
- ▶ Detect pattern in the training data and build a predictive model.

Supervised/Unsupervised Learning:

- ▶ Machine Learning systems can be classified according to the amount and type of supervision.
- ▶ Supervised Learning
- ▶ Unsupervised Learning
- ▶ Semi supervised Learning
- ▶ Reinforcement Learning

Supervised Learning

- ▶ The training data you feed to the algorithm includes the desired solutions, called labels
- ▶ Supervised learning tasks:
 - ▶ Classification (Spam filter)
 - ▶ Regression – to predict the target numeric value (price of house)
 - ▶ Logistic Regression – used for classification
 - ▶ Output a value that corresponds to the probability of belonging class.
 - ▶ Eg. 20% chance of being spam

Supervised Learning Algorithms:

- ▶ K-Nearest Neighbors
- ▶ Linear Regression
- ▶ Logistic Regression
- ▶ Support Vector machines (SVMs)
- ▶ Decision Trees and Random Forests.
- ▶ Neural Networks (can be unsupervised or semisupervised)

Unsupervised Learning:

- ▶ The training data is unlabeled
- ▶ System tries to learn without a teacher

Unsupervised Learning Algorithms

- ▶ Clustering :
 - ▶ K-means
 - ▶ Hierarchical Cluster Analysis(HCA)
 - ▶ Expectation Maximization
- ▶ Visualization and dimensionality reduction
 - ▶ Principal Component Analysis
 - ▶ Kernel PCA
 - ▶ Locally –Linear Embedding(LLE)
- ▶ Associate rule Learning.....?
 - ▶ Apriori
 - ▶ Eclat

Semi supervised Learning

- ▶ Lots of unlabeled data and a little bit of labeled data
- ▶ Most algorithms are combination of unsupervised and supervised algorithms
- ▶ Google photo
 - ▶ Person A shows up in photo 1,2,5 and 7 (unsupervised)
 - ▶ Name of that person (supervised)

Reinforcement Learning:

- ▶ The learning system , called an agent in this context, can observe the environment , select and performs actions.
- ▶ Get rewards in return or penalties in the form of negative rewards.
- ▶ It learns by itself the best strategy, called policy.
- ▶ Example Robot
 - ▶ Deep Mind's AlphaGo Program
 - ▶ Beats the world champion Lee Sedol at the game of Go

Batch and Online Learning

- ▶ Whether or not a system can learn incrementally from a stream of incoming data
- ▶ Batch :
- ▶ System is incapable of learning incrementally.
- ▶ Take a lot of time and computing resources. So it is typically done offline