Machine Learning 2018
Introduction to Machine Learning

Kien C Nguyen

November 17, 2018

What is Machine Learning?

Definition

Machine Learning (ML) is a field of artificial intelligence that uses statistical techniques to give computer systems the ability to "learn" (e.g., progressively improve performance on a specific task) from data, without being explicitly programmed (Wikipedia).

https://en.wikipedia.org/wiki/Machine_learning

- Explores the study and construction of algorithms that can learn from and make predictions on data
- Such algorithms overcome following strictly static program instructions by making data-driven predictions or decisions, through building a model from sample inputs.
- Employed in a range of computing tasks where designing and programming explicit algorithms with good performance is difficult or infeasible
- Example applications include email filtering, detection of network intruders, NLP, and computer vision.

What is Machine Learning?

https://en.wikipedia.org/wiki/Machine_learning
Tom M. Mitchell (an American computer scientist) provided a widely quoted, more formal definition of the algorithms studied in the machine learning field:

Definition

"A computer program is said to learn from experience E with respect to some class of tasks T and performance measure P if its performance at tasks in T, as measured by P, improves with experience E."

What is Machine Learning?

In a course project at the VEF Academy's Machine Learning class, Hoang uses TCP dump data to separate an attack connection from a normal connection. He looks at features of a TCP connection such as duration, network service, number of bytes from source to destination, number of bytes from destination to source, to detect whether a connection is an attack or a normal one. He uses 70% instances from KDD Cup 1999 dataset to train and the remaining 30% to test his model. Feature values and the labels ('attack' or 'normal') in the train set are fed into his model. To evaluate the model, he employs a Bayesian approach on the test set where he assigns zero costs to True Negative and True Positive, and positive costs to False Negative and False Positive. Specify the following:

- The task T of the detection problem
- The experience E
- The performance measure P

Why Machine Learning now? Data

These days we have a humongous amount of new data everyday to make sense of. E.g. (from Murphy [1])

- There are about 1 trillion web pages
- one hour of video is uploaded to YouTube every second, amounting to 10 years of content every day
- Walmart handles more than 1M transactions per hour and has databases containing more than 2.5 petabytes (2.5×10^{15}) of information.
- Electronic & computer engineering advances help collect more data (IoT, sensors, cameras etc.)

Why Machine Learning now? Computational Power

We also have breakthroughs in ICs, memories, storages, microprocessors (GPUs), computer systems, and data centers.

- A CPU core is designed to support an extremely broad variety of tasks (e.g., render a webpage, drive word processors and enterprise software, manage peripherals) in addition to performing computations, whereas a GPU core is optimized exclusively for data computations.
- Because of this singular focus, a GPU core is simpler and has a smaller die area than a CPU, allowing many more GPU cores to be crammed onto a single chip.
- Consequently, ML applications, which perform large numbers of computations on a vast amount of data, can see huge (i.e., 5 to 10 times) performance improvements when running on a GPU versus a CPU.

(https://www.forbes.com/sites/forbestechcouncil/2017/12/01/for-machine-learning-its-all-about-gpus)

Why Machine Learning now? Research & Development

- A lot of recent Research & Development in Machine Learning, especially in Deep Learning.
- The development of software, especially open-source software, helps fuel the R & D in Machine Learning and the application of ML in every aspect of our lives.
- Think Python, R, scikit-learn, TensorFlow, PyTorch

Types of machine learning

- Machine learning is usually divided into two main types.
- In the predictive or supervised learning approach, the goal is to learn a mapping from inputs x to outputs y, given a labeled set of input-output pairs $D = \{(x_i, y_i)\}_{i=1}^{N}$. Here D is called the training set, and N is the number of training examples.
- The second main type of machine learning is the descriptive or unsupervised learning approach.
- Here we are only given inputs, $D = \{x_i\}_{i=1}^{N}$, and the goal is to find "interesting patterns" in the data. This is sometimes called knowledge discovery.
- We also have semi-supervised learning and reinforcement learning.

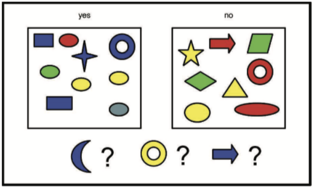
(from Murphy [1])

For supervised learning,

- when y_i is categorical, the problem is known as classification or pattern recognition. For example, the problem of classifying emails into 'spam' and 'not spam'.
- y_i is real-valued, the problem is known as regression. For example, the problem of predicting the income level.
- Another variant, known as ordinal regression, occurs where label space Y has some natural ordering, such as grades A–F.

(from Murphy [1])

An example from Murphy [1]



(a)

D features (attributes)

N cases

Color	Shape	Size (cm)		Label
Blue	Square	10		1
Red	Ellipse	2.4		1
Red	Ellipse	20.7		0

(b)

Figure 1.1 Left: Some labeled training examples of colored shapes, along with 3 unlabeled test cases. Right: Representing the training data as an $N \times D$ design matrix. Row i represents the feature vector $\mathbf{x}_i$. The last column is the label, $y_i \in \{0, 1\}$. Based on a figure by Leslie Kaelbling.

(from Murphy [1])

More examples

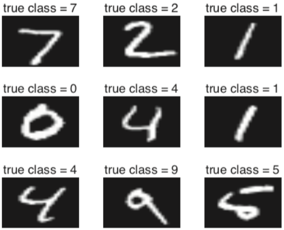
What types of Machine Learning do we use for the following problems

- Classify someone's emails into 'spam' and 'not spam'. You are given a train dataset with emails that have been labeled 'spam' or 'not spam'.
- Predict the house price using the features such as location, lot area, lot frontage, street, construction year etc. You are given a train dataset with features and prices of previous transactions.
- Using data of online customers from an e-commerce site, divide the customers into market segments so that you can concentrate your marketing efforts on relevant segments (target marketing).

Document classification and email spam filtering

In document classification, the goal is to classify a document, such as a web page or email message, into one of C classes, that is, to compute $p(y = c|x, D)$, where x is some representation of the text. A special case of this is email spam filtering, where the classes are spam y = 1 or not spam y = 0. (from Murphy [1])

In the special case that the images consist of isolated handwritten letters and digits, for example, in a postal or ZIP code on a letter, we can use classification to perform handwriting recognition. A standard dataset used in this area is known as MNIST, which stands for "Modified National Institute of Standards".



(from Murphy [1])

Face detection and recognition

- A harder problem is to find objects within an image; this is called object detection or object localization.
- An important special case of this is face detection.
- One approach to this problem is to divide the image into many small overlapping patches at different locations, scales and orientations, and to classify each such patch based on whether it contains face-like texture or not.
- This is called a sliding window detector.
- The system then returns those locations where the probability of face is sufficiently high.
- Having found the faces, one can then proceed to perform face recognition, which means estimating the identity of the person

(from Murphy [1])

(a) (b)

Figure 1.6 Example of face detection. (a) Input image (Murphy family, photo taken 5 August 2010). Used with kind permission of Bernard Diedrich of Sherwood Studios. (b) Output of classifier, which detected 5 faces at different poses. This was produced using the online demo at `http://demo.pittpatt.com/`. The classifier was trained on 1000s of manually labeled images of faces and non-faces, and then was applied to a dense set of overlapping patches in the test image. Only the patches whose probability of containing a face was sufficiently high were returned. Used with kind permission of Pittpatt.com

(from Murphy [1])

Table 1: Students, gender (M: Male, F: Female), and the exam scores

Student	Gender	Midterm 1 Score	Midterm 2 Score	Final Score
Thanh	M	7	6	8
Hoai	F	9	8	8
Andy	M	10	8	9
Beatrix	F	6	7	7
Wen	M	5	6	7
Mikhail	M	6	6	6
Shichao	M	7	8	7
Natasha	F	8	9	8
Hung	M	9	8	8
Yen	F	10	9	9
Dietmar	M	8	5	7
John	M	5	6	7
Angelina	F	7	7	8
Jennifer	F	9	8	8
Nick	M	6	8	8
Tung	M	8	7	7
Olga	F	6	7	7
Gaurav	M	7	8	8
Bwatna	F	10	9	10
Celine	F	7	6	8

Use a linear regression model to regress the Final score against the Midterm 1 score and Midterm 2 score. What are the coefficients (including the intercept) that you get from the model? What is the coefficient of determination (R^2) of the model?

Regression

```python
# Importing the libraries
import numpy as np
import matplotlib.pyplot as plt
import pandas as pd
from sklearn.preprocessing import StandardScaler
from sklearn.linear_model import LinearRegression
from sklearn.metrics import r2_score

scores = pd.read_csv("screening_test_problem_5.csv")

# If we want to do feature scaling using StandardScaler
# X = StandardScaler().fit_transform(X)

regressor = LinearRegression()
regressor.fit(X, y)
```

```
# coefficients
print 'Coefficients: ', regressor.coef_

# intercept
print 'Intercept: ', regressor.intercept_

# r2
r2 = r2_score(y, regressor.predict(X))
print 'r2 = ', r2
```

Results without feature scaling:
Coefficients: [0.30 0.26]
Intercept: 3.58
R^2: 0.62
If we scale Midterm 1 scores and Midterm 2 scores using StandardScaler:
Coefficients: [0.47 0.30]
Intercept: 7.75

Unsupervised learning

- Clustering
- Principal Component Analysis

Unsupervised learning - Clustering

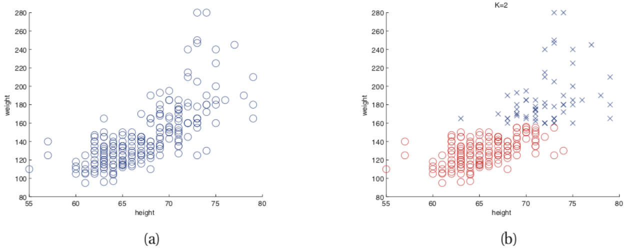
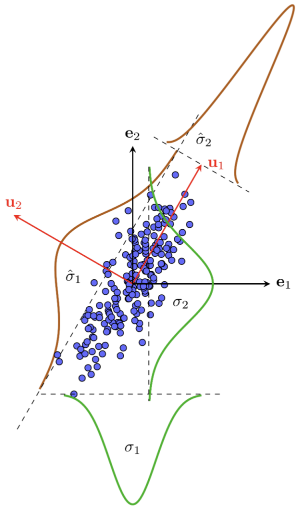


Figure 1.8 (a) The height and weight of some people. (b) A possible clustering using $K = 2$ clusters. Figure generated by `kmeansHeightWeight`.

(from Murphy [1])

Unsupervised learning – Principal Component Analysis



(Figure from https://machinelearningcoban.com/)

References

[1] K. P. Murpy – Machine Learning – A Probabilistic Perspective, MIT Press, 2012.
[2] Vu Huu Tiep – https://machinelearningcoban.com/