Machine Learning
1.1. Introduction

Teeradaj Racharak (เอ็กซ์)
r.teeradaj@gmail.com

Readings

Readings for these lecture notes:

- Bishop, C. (2006), *Pattern Recognition and Machine Learning*, Springer, Chapters 3, 4, 6, 7.
- Hastie, T., Timshirari, R., and Friedman, J. (2016), *Elements of Statistical Learning: Data Mining, Inference, and Prediction,* Springer, Chapters 2, 3, 4, 12.
- Bonnin, R. (2016), *Building Machine Learning Projects with TensorFlow*, Packt Publishing.
- Ng, A. (2017), *Machine Learning,* Stanford University.
- Dailey, M. (2018), *Machine Learning*, Asian Institute of Technology

These notes contains materials from Bishop (2016), Hastie *et al.* (2016), Ng (2017), and Dailey (2018)

Recap

THE MAP OF MATHEMATICS
FOUNDATIONS
FUNDAMENTAL RULES
MATHEMATICAL LOGIC
p ⇒ q
CONSISTENT SET OF AXIOMS?
NOPE
GÖDEL INCOMPLETENESS THEOREMS
SET THEORY
A ∩ B
CATEGORY THEORY
THEORY OF COMPUTATION
P≠NP?
COMPLEXITY THEORY
CARDINAL NUMBERS
ALEPH NULL
OCTONION
QUATERNION
a+bi+cj+dk
PI
EXPONENTIAL
COMPLEX NUMBERS
3, i, 4+3i, -4i
PRIME NUMBERS
3, 11, 47, 907
INFINITY
REAL NUMBERS
RATIONAL NUMBERS
NUMBER THEORY
INTEGERS
...-2, -1, 0, 1, 2...
NATURAL NUMBERS
1, 2, 3, 4, 5...
ARITHMETIC
COMBINATORICS
PARTITION THEORY
TREE
GROUP THEORY
GRAPH THEORY
FACTORS OF 30
ORDER THEORY
PERMUTATION GROUP
MEASURE THEORY
LINEAR ALGEBRA
NUMBER SYSTEMS
MATRICES
ALGEBRA
EQUATION
y=mx+c
VECTORS
STRUCTURES
PURE MATHEMATICS
MÖBIUS STRIP
TOPOLOGY
DIFFERENTIAL GEOMETRY
1 HOLE (GENUS 1)
COMPLEX ANALYSIS
FRACTAL GEOMETRY
MANDELBROT SET
DYNAMICAL SYSTEMS
PYTHAGORAS
TRIGONOMETRY
GEOMETRY
SPACES
CHANGES
CALCULUS
DIFFERENTIAL
GRADIENT = dy/dx
INTEGRAL
AREA = ∫ f(x) dx
DIFFERENTIAL EQUATIONS
VECTOR CALCULUS
FLUID FLOW
ECOSYSTEMS
BUTTERFLY EFFECT
CHAOS THEORY
ORIGINS
INDIA c.628
FIRST ZERO 0
ALGEBRA
c.1730 MATHEMATICAL NOTATION
NEGATIVE NUMBERS
-8 ||| CHINA 200 BCE
COUNTING
EGYPT FIRST EQUATION 3000 BCE
GREECE 600-300 BCE
CRYPTOGRAPHY
ENCRYPT
DECRYPT
PUBLIC KEY
PRIVATE KEY
COMPUTER SCIENCE
TURING MACHINE
MACHINE LEARNING
CAT
while awake:
do_science()
if self.tired():
awake = False
self.repair_brain()
PROBABILITY
BAYES' RULE
P(A|B) = P(B|A)P(A) / P(B)
STATISTICS
OPTIMIZATION
MATHEMATICAL FINANCE
GAME THEORY
ECONOMICS
APPLIED MATHEMATICS
NUMERICAL ANALYSIS
ENGINEERING
CONTROLLER
SYSTEM
SENSOR
CONTROL THEORY
MATHEMATICAL PHYSICS
MATHEMATICAL CHEMISTRY
THEORETICAL PHYSICS
BIOMATHEMATICS
BY DOMINIC WALLIMAN © 2017
YOUTUBE: THE MAP OF MATHEMATICS

Today Onward

Introduction

- Machine learning is now near the top of the list of skills U.S. companies want to see in the people they hire.
- Many tasks we want computers to do are difficult to program directly.
- A set of tools that let us specify the computer's behavior by giving examples of how it should respond in given situations, without specifying the computation necessary to formulate that response.
- **Essential idea:** we want to create a model from data that can later be queried when new situation arises.

Examples of ML in real life

We are using machine learning every time that we:

- use a credit card
- get a recommendation from Netflix or Amazon
- Ask Google for directions by voice
- Take a ride in our Tesla !

Let’s brainstorm about things closer to home that might be using machine learning already or might benefit from it in the near future.

How many ways to learn?

Types of Learning

1. **Supervised learning:** ‘right’ answers are given
2. **Unsupervised learning:** ‘right’ answers are ‘not’ given
3. **Reinforcement learning:** derives a policy that enables an agent to behave optimally in an uncertain environment using feedback on the goodness of the outcome over time.
4. **Miscellaneous (*e.g.* transfer learning):** storing knowledge gained solving one problem and applying it to a different but related problem.

ML's success in industry

Further readings

Qiang Yang

New Bright Chair Professor of Engineering, Hong Kong Univ. of Sci. and Tech.
Verified email at cse.ust.hk - Homepage

Artificial Intelligence Transfer Learning Machine Learning and Data…

TITLE

A survey on transfer learning
SJ Pan, Q Yang
IEEE Transactions on knowledge and data engineering 22 (10), 1345-1359

Top 10 algorithms in data mining
X Wu, V Kumar, JR Quinlan, J Ghosh, Q Yang, H Motoda, GJ McLachlan, ...
Knowledge and information systems 14 (1), 1-37

Graph embedding and extensions: A general framework for dimensionality reduction
S Yan, D Xu, B Zhang, HJ Zhang, Q Yang, S Lin
IEEE transactions on pattern analysis and machine intelligence 29 (1), 40-51

Boosting for Transfer Learning
W Dai, Q Yang, GR Xue, Y Yu
Proceedings of the 24th Annual International Conference on Machine Learning …

- Check publications of Prof. Qiang Yang.
- Blog of Sebastian Ruder, who is currently working on NLP with deep learning: http://ruder.io/transfer-learning/index.html

Definitions of ML

- Arthur Samuel (1959).
  - Machine learning (ML) is a field of study that ‘gives’ computers the ability to learn without being explicitly programmed.

- Tom Mitchell (1998).
  - A computer program is said to ‘learn’ from experience *E* w.r.t. some task *T* and some performance measure *P*, if its performance on *T*, as measured by *P*, improves with experience *E*.

Question

- Suppose your email program watches which emails you do or do not mark as spam, and based on that learns how to better filter spam. What is the task T in this setting?

  i. Classify emails as spam or not spam
  ii. Watching you label emails as spam or not spam
  iii. The number (or fraction) of emails correctly classified as spam / not spam
  iv. None of the above — this is not a machine learning problem

Question

- Suppose your email program watches which emails you do or do not mark as spam, and based on that learns how to better filter spam. What is the task *T* in this setting?

  i. Classify emails as spam or not spam **T**
  ii. Watching you label emails as spam or not spam **E**
  iii. The number (or fraction) of emails correctly classified as spam / not spam **P**
  iv. None of the above — this is not a machine learning problem

Now, let’s talk about the intuition of each basic problem in ML

1. Supervised Learning

Intuition (Regression)

Housing price prediction

Intuition (Regression)

Housing price prediction

Intuition (Regression)

Housing price prediction

Intuition (Regression)

Housing price prediction

Intuition (Regression)

Housing price prediction

Is this the only learning algorithm that we can use?

Intuition (Regression)

Housing price prediction

Is this the only learning algorithm that we can use?

Intuition (Regression)

Housing price prediction

Is this the only learning algorithm that we can use?

Intuition (Regression)

Housing price prediction

Is this the only learning algorithm that we can use?

<u>Supervised Learning</u>

"right answer" is given

Intuition (Regression)

Housing price prediction

Is this the only learning algorithm that we can use?

<u>Supervised Learning</u>

"right answer" is given

<u>Regression</u>

Predict continuous valued output (price)

Intuition (Classification)

Breast cancer (malignant, benign)

Intuition (Classification)

Breast cancer (malignant, benign)

<u>Classification</u>

Discrete valued output (0 or 1)

Intuition (Classification)

Breast cancer (malignant, benign)

<u>Classification</u>

Discrete valued output (0 or 1)

Discrete value may be ≥ 2

e.g. 0, 1, 2, 3

Intuition (Classification)

Breast cancer (malignant, benign)

<u>Classification</u>

Discrete valued output (0 or 1)

Discrete value may be ≥ 2

e.g. 0, 1, 2, 3

- 0 : Benign
- 1 : Cancer type#1
- 2 : Cancer type#2
- 3 : Cancer type#3

Intuition (Classification)

Breast cancer (malignant, benign)

<u>Classification</u>

Discrete valued output (0 or 1)

Discrete value may be ≥ 2

e.g. 0, 1, 2, 3

- 0 : Benign
- 1 : Cancer type#1
- 2 : Cancer type#2
- 3 : Cancer type#3

Intuition (Classification)

Breast cancer (malignant, benign)

<u>Classification</u>

Discrete valued output (0 or 1)

Discrete value may be ≥ 2

e.g. 0, 1, 2, 3

- 0 : Benign
- 1 : Cancer type#1
- 2 : Cancer type#2
- 3 : Cancer type#3

1 feature (attribute)

Intuition (Classification)

Breast cancer (malignant, benign)

X : Malignant

X : Benign

Intuition (Classification)

Breast cancer (malignant, benign)

X : Malignant

X : Benign

Intuition (Classification)

Breast cancer (malignant, benign)

X : Malignant

X : Benign

Intuition (Classification)

Breast cancer (malignant, benign)

X : Malignant

X : Benign

Possible to have ≥ 2 features *e.g.*

- Clump Thickness
- Uniformity of Cell Size
- Uniformity of Cell Shape
- *etc.*

Intuition (Classification)

Breast cancer (malignant, benign)

X : Malignant

X : Benign

Possible to have ≥ 2 features *e.g.*

- Clump Thickness
- Uniformity of Cell Size
- Uniformity of Cell Shape
- *etc.*

In fact, real problems may involve many features.

Challenges:
How to deal with this situation if a computer is going to run out of memory?

Intuition (Classification)

Breast cancer (malignant, benign)

X : Malignant

X : Benign

Possible to have ≥ 2 features *e.g.*

- Clump Thickness
- Uniformity of Cell Size
- Uniformity of Cell Shape
- *etc.*

Answer:
Some mathematical tricks in support vector machine (SVM) allow us to deal with this !

Recap

- Supervised learning
  - "right answers" are given
- Regression
  - Predicts a continuous value output
- Classification
  - Predicts a discrete value output

Question

- You are running a company, and you want to employ learning algorithms to address each of the following problems.
  1. You have a large inventory of identical items. You want to predict how many of these items will sell over the next 3 months.
  2. You want to have a software to examine individual customer accounts; and for each account, you want to decide if it has been hacked / compromised.

  i. Treat both as classification problems
  ii. Treat problem#1 as a classification problem and problem#2 as a regression problem
  iii. Treat problem#1 as a regression problem and problem#2 as a classification problem
  iv. Treat both as regression problems

Question

- You are running a company, and you want to employ learning algorithms to address each of the following problems.
  1. You have a large inventory of identical items. You want to predict how many of these items will sell over the next 3 months.
  2. You want to have a software to examine individual customer accounts; and for each account, you want to decide if it has been hacked / compromised.

  i. Treat both as classification problems
  ii. Treat problem#1 as a classification problem and problem#2 as a regression problem
  iii. Treat problem#1 as a regression problem and problem#2 as a classification problem
  iv. Treat both as regression problems

2. Unsupervised Learning

Supervised Learning

Unsupervised Learning

Intuition (Clustering)

Intuition (Clustering)

Social Network Analysis

Market Segmentation Analysis

Question

- Of the following examples, which would you address using an unsupervised learning algorithm? (Circle all that apply)

  i. Given email labeled as spam / not spam, learn a spam filter.
  ii. Given a set of news articles found on the web, group them into sets of articles about the same stories.
  iii. Given a database of customer data, automatically discover market segments and group customers into different market segments.
  iv. Given a dataset of patients diagnosed as either having diabetes or not, learn to classify new patients as having diabetes or not.

Question

- Of the following examples, which would you address using an unsupervised learning algorithm? (Circle all that apply)

  i. Given email labeled as spam / not spam, learn a spam filter.
  ii. Given a set of news articles found on the web, group them into sets of articles about the same stories.
  iii. Given a database of customer data, automatically discover market segments and group customers into different market segments.
  iv. Given a dataset of patients diagnosed as either having diabetes or not, learn to classify new patients as having diabetes or not.