COIS 1010H - Special Topic

Introduction to Machine Learning

Wenying Feng

1

Image recognition

- Teaching a child the difference between dogs and cats by using flash-cards: you show a card, the child makes a choice.

2

Learning by examples

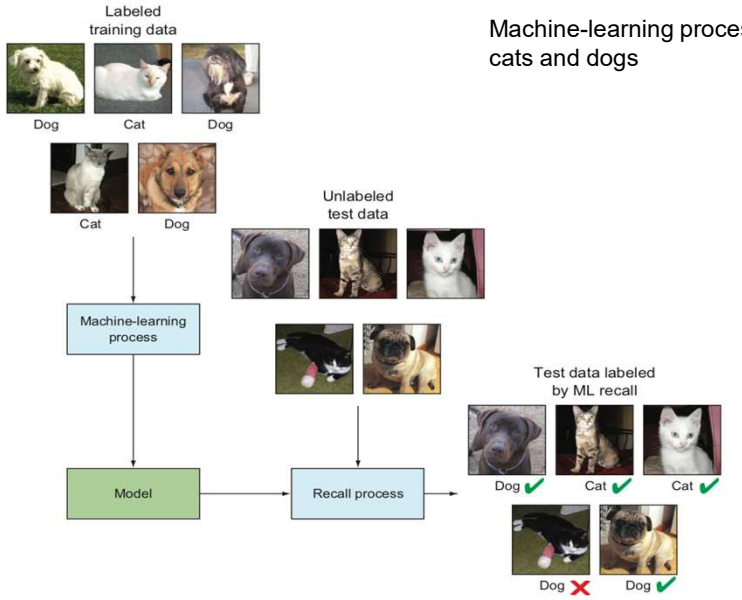
- As the child practices, his performance improves.
- It isn't necessary to first teach the child techniques for cat and dog recognition.
- Human cognition has built-in classification mechanisms. All that is needed are examples.

3

Generalization

- A key characteristic of learning for both human and machine learning is to apply knowledge gained through training to new unseen examples.

4



Machine-learning process for cats and dogs

Henrik, Richards and Fetherolf, Real-world Mahcine Learning, Manning publications, 2017

5

Program competition – cat or dog?

- Program for classifying whether images depicted a dog or cat?
- Provided 25,000 example images for training.
- After all the competitors had trained their algorithms, they were tested on their ability to classify 12,500 unlabeled test images.
- The winner classified 98.914% of the unseen test images correctly.
- Human error rate is around 7%.
- https://www.Kaggle.com/c/dogs-vs-cats

6

Supervised ML

- Machine learning in which, given examples for which the output value is known, the training process infers a function that relates input values to the output.
- Training data: the set of instances with a known target to be used to fit an ML model.
- Unsupervised machine learning: Machine-learning techniques that don't rely on labeled examples, but rather try to find hidden structure in unlabeled data.

7

Use cases for supervised ML

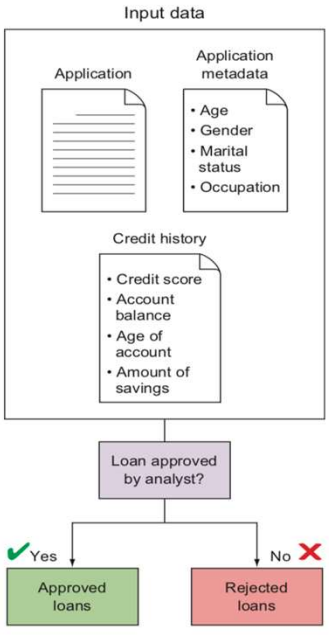
Problem	Description	Example use cases
Classification	Determine the discrete class to which each individual belongs, based on input data	Spam filtering, sentiment analysis, fraud detection, customer ad targeting, churn prediction, support case flagging, content personalization, detection of manufacturing defects, customer segmentation, event discovery, genomics, drug efficacy
Regression	Predict the real-valued output for each individual, based on input data	Stock-market prediction, demand forecasting, price estimation, ad bid optimization, risk management, asset management, weather forecasting, sports prediction
Recommendation	Predict which alternatives a user would prefer	Product recommendation, job recruiting, Netflix Prize, online dating, content recommendation
Imputation	Infer the values of missing input data	Incomplete patient medical records, missing customer data, census data

8

A real-world business problem

- A microlending company that provides loans to individuals who want to start small businesses.
- As the company continues to gain popularity, the number of applications begins to increase.

9

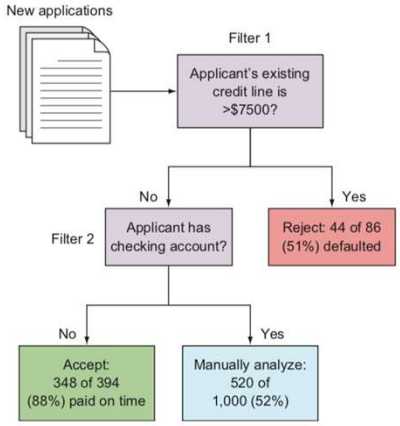


- Using data to make decisions
  - Traditional approaches

1) Hire one more analyst
2) Hire two more analysts
3) Training programs
4) Employ business rules by looking at trends between the applicant data and incidence of loan repayment

10

Filtering new applications through two business rules



- Most borrowers with a credit line of more than $7,500 defaulted on their loan.
- Most borrowers who had no checking account repaid their loan on time.
- Reduce manual analysis to only 52% of the incoming applications

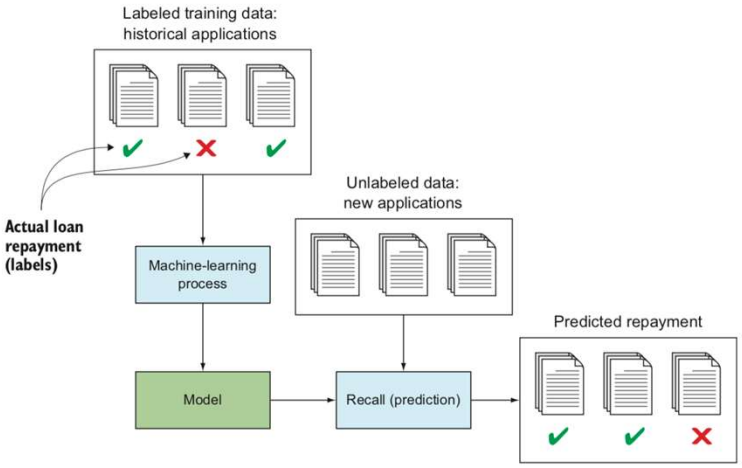
11

Drawbacks

- Effective filters becomes harder and harder-grows in complexity.
- The business rules become so complicated.
- The construction of your rules has no statistical rigor - better "rules" can be found by better exploration of the data?
- Patterns of loan repayment change over time - the system doesn't adapt to those changes.

12

The Machine Learning Approach



13

A linear model

$$y = \beta_0 + \beta_1 * (credit\_line) + \beta_2 * (Education\_level) + \beta_3 * Age$$

$\beta_0, \beta_1, \beta_2, \beta_3$ are parameters.

14

Advantages of ML

- *Accurate:* more data, the accuracy can increase automatically.
- *Automated:* learn new patterns automatically.
- *Fast:* generate answers in a matter of milliseconds, allowing systems to react in real time.
- *Customizable:* ML models are custom built from your own data, and can be configured to optimize whatever metric drives your business.
- *Scalable: a*s business grows, ML easily scales to handle increased data rates. Some ML algorithms can scale to handle large amounts of data on many machines in the cloud.

15

Challenges

- Acquiring data in a usable form
- Formulating the problem so that machine learning can be applied, and will yield a result that's actionable and measurable
- When the outcome you're trying to predict is complicated, choosing the algorithm and how to apply it may be an enormous effort in itself
- …

16

ML and data-driven problems

- Machine learning can solve a great variety of problems, some much more easily than others.
- The value of the solution isn't always proportional to the effort required.
- It is the perfect choice for many real-world, data-driven problems.

17

Artificial Intelligence (AI)

- AI is a far broader subject (robotics, language processing, computer vision systems …).
- Machine learning, as a type of AI, provides computers with the ability to learn without being explicitly programmed
- It focuses on the development of computer programs that can adjust themselves with new input data

18

Capability of an AI system

- An AI system must be capable of doing:

  (1) Store knowledge
  (2) Applying the knowledge stored to solve problems
  (3) Acquire new knowledge through experience

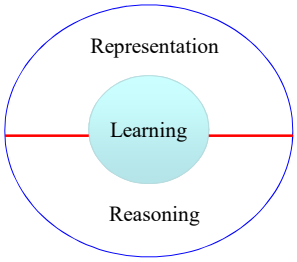
19

Components of an AI System

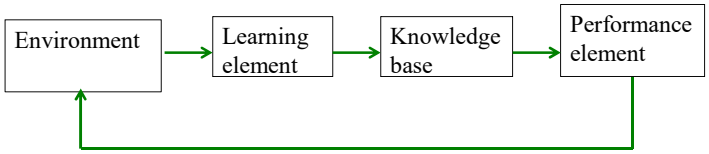
- Representation
  - use a language of symbol structure to represent knowledge of interest and the solution to the problem
- Reasoning
  - the ability to apply the knowledge to solve problems
  - control mechanism that determine which operation to apply to particular problem
- Learning - Acquire new knowledge

20

Key Components of an AI System



- The environment supplies some information to a learning element
- The learning element then uses this information to make improvements in a knowledge base
- The performance element uses the knowledge base to perform the task and provides feedback
- The feedback mechanism enables the machine to evaluate its hypotheses and revise them if necessary



21

Diagrams adopted from Neural Networks, Simon Haykin, Prentice Hall, 2ed edition, 1999.

In the Past 50 Years

Could Computer learn to play

+

Statistics

Theories of learning processes
Algorithms of learning procedures
Industry in Data Mining

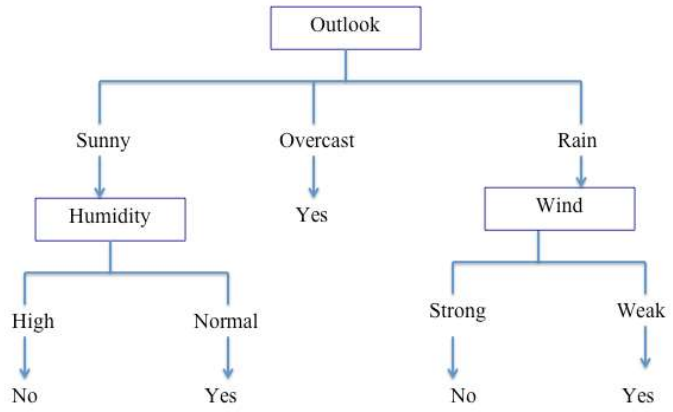
22

Example: Decision Tree Learning

- Decision tree learning is a method for approximating discrete-valued target functions
- The learned function is represented by a decision tree
- Commonly used in data mining
- Decision trees classify instances or examples by starting at the root of the tree and moving through a leaf node

23

A Decision Tree for the Concept of Play Tennis



24

Conditions to Play Tennis

$(\text{Outlook}=\text{Sunny} \wedge \text{Humidity}=\text{Normal}) \vee$

$(\text{Outlook}=\text{Overcast}) \vee$

$(\text{Outlook}=\text{Rain} \wedge \text{Wind}=\text{Weak})$

String Representation using binary digits

25

String Representation (5 binary digits)

Sunny Humidity (high, normal)

100 01 10

Overcast

010

Rain Wind (strong, weak)

001 01 10

- How many possible output? How to represent?

26

When Decision Trees can be Applied

- Instances that are represented by a fixed set of attributes and their values
- The target function has discrete output values. Decision tree methods can easily be extended to learn functions with more than two possible output values
- The training data may contain missing attribute values. Decision tree methods can be used even when some training examples have unknown values

27

Artificial Neural Networks – Motivation

- To construct a computer capable of “human-like thought”
- The human brain as a whole is far too complex to model
- At the most basic level, the human brain is composted primarily of neuron cells
- Artificial Neural Networks (ANN) attempt to simulate the behavior of these cells

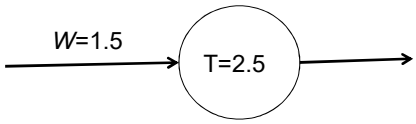
28

Problems commonly solved with ANN

- Classification
- Prediction
- Pattern recognition
- Optimization

29

Artificial neuron

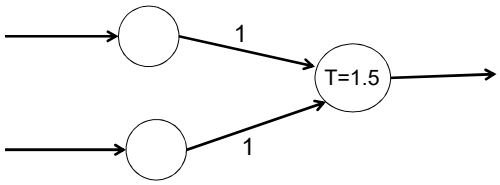


Threshold T=2.5
Weight W=1.5

- A perceptron is the simplest neural network possible: a computational model of a single neuron
- It consists of one or more inputs, a processor, and a single output
- An incoming signal will be amplified, or de-amplified by the weight.
- If the weighted input exceeds the threshold, the neuron will fire.

30

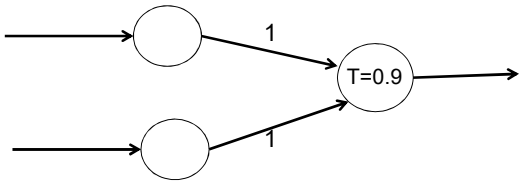
A neural network for the AND logical operation



A	B	A AND B
0	0	
0	1	
1	0	
1	1	

31

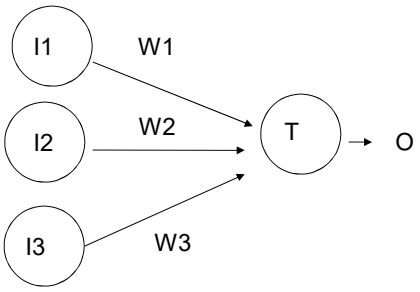
The OR logical operation



A	B	A OR B
0	0	
0	1	
1	0	
1	1	

32

More than two inputs

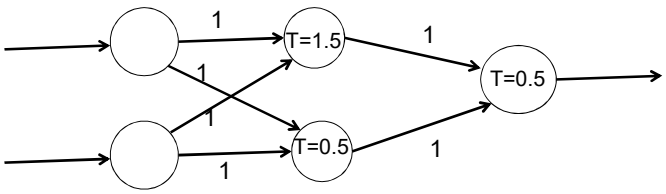


$$O = \left\{ \begin{array}{l} 1 : \sum_i w_i I_i > T \\ 0 : otherwise \end{array} \right\}$$

Constant Activation Function

33

XOR logical operation – more layers



A	B	A XOR B
0	0	
0	1	
1	0	
1	1	

34

Learning Procedure

- Randomly assign weights (between 0-1)
- Present inputs from training data
- Get output O, nudge weights to gives results toward our desired output
- Repeat; stop when no errors, or enough epochs completed

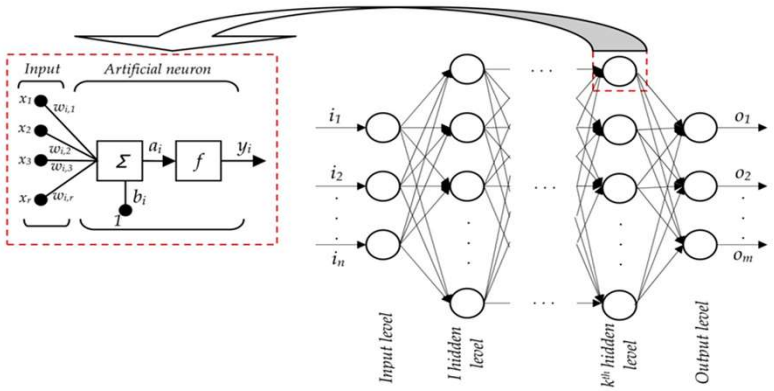
35

How is a perceptron network used

- generally used to learn how to make classifications
- diagnosis of patients with heart disease using collected data:
  - Age, Sex, Chest Pain Type, Resting BPS, Cholesterol, …, Diagnosis
  - `67,1,4,120,229,…, 1`
  - `37,1,3,130,250,… ,0`
  - `41,0,2,130,204,… ,0`
- Train network to predict heart disease of new patient

36

Machine Learning for Energy Forecast

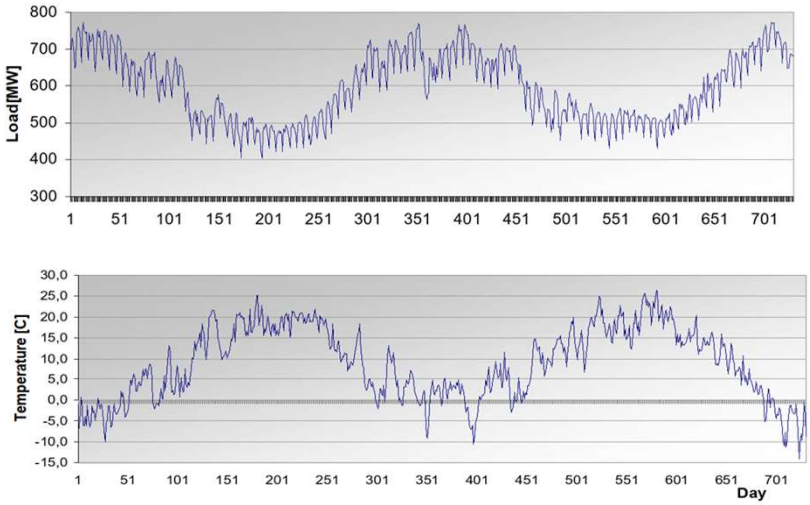


Aarshay Jain

https://www.analyticsvidhya.com/blog/2016/03/introduction-deep-learning-fundamentals-neural-networks/

37

Loads for Two Years



38

Input to the Learning Model

```
dataset.head()
```

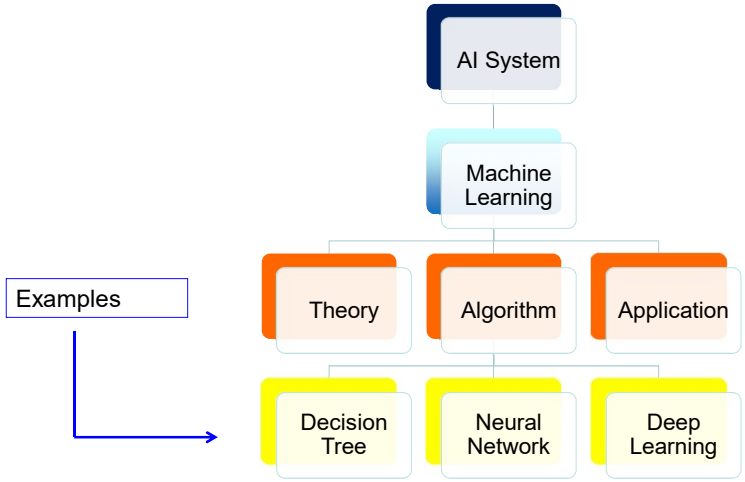
datetime	Load	Temperture	weekday	Holday
1997-01-01 00:30:00	797.0	-7.55	2	1
1997-01-01 01:00:00	794.0	-7.55	2	1
1997-01-01 01:30:00	784.0	-7.55	2	1
1997-01-01 02:00:00	787.0	-7.55	2	1
1997-01-01 02:30:00	763.0	-7.55	2	1

39

Prediction for Maximum Usages

40

Summary



41

References

1. Real-World Machine Leanring, H. Brink, J. W. Richards and M. Fetherolf, Manning Publications 2017.
2. Introduction to Neural Networks with Java, 2nd edition, Jeff Heaton, Heaton Research, Inc. 2008.
3. Neural Networks, A Comprehensive Foundation, Simon Haykin, 2nd edition, Prentice Hall.
4. Artificial Intelligence: A Modern Approach, S. Russell and P. Norvig, 2nd edition, Prentice Hall.
5. www.myreaders.info
6. Introduction to Artificial Intelligence: Padhraic Smyth, http://www.ics.uci.edu/~smyth/courses/cs271/
7. Open source documents, mainly internet.

42