

Introduction

Machine Learning 1 — Lecture 1

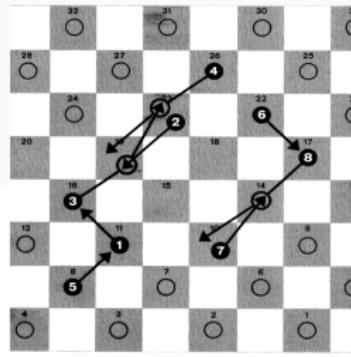
5th March 2024

Robert Peharz

Institute of Theoretical Computer Science
Graz University of Technology

Arthur L. Samuel, 1959

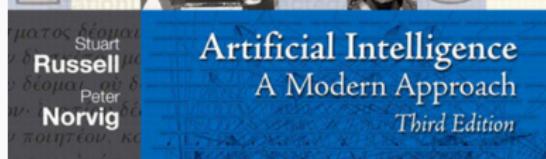
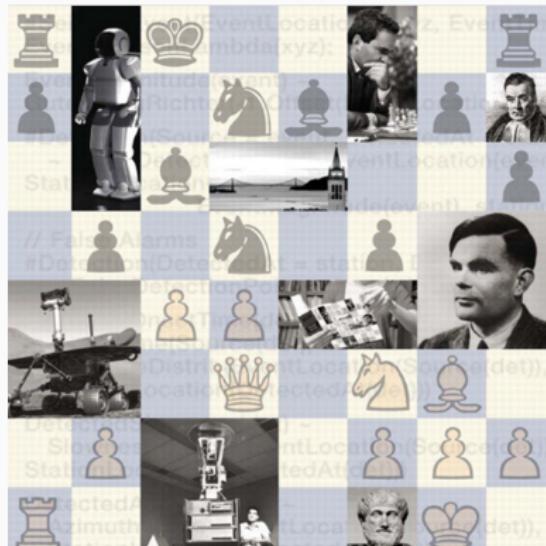
Some Studies in Machine Learning Using the Game of Checkers



Eight moves starting utilizing perceptron learning. See Appendix B. Score 0-43.

Abstract: Two machine-learning procedures have been investigated in some detail using the game of checkers. Enough work has been done to verify the fact that a computer can be programmed so that it will learn to play a better game of checkers than can be played by the person who wrote the program. Furthermore, it can learn to do this in a remarkably short period of time (8 or 10 hours of machine-playing time) when given only the rules of the game, a sense of direction, and a redundant and incomplete list of parameters which are thought to have something to do with the game, but whose correct signs and relative weights are unknown and unspecified. The principles of machine learning verified by these experiments are, of course, applicable to many other situations.

Machine Learning: Subfield of Artificial Intelligence



Four Historical Approaches to AI

Thinking Humanly <p>“The exciting new effort to make computers think . . . <i>machines with minds</i>, in the full and literal sense.” (Haugeland, 1985)</p> <p>“[The automation of] activities that we associate with human thinking, activities such as decision-making, problem solving, learning . . .” (Bellman, 1978)</p>	Thinking Rationally <p>“The study of mental faculties through the use of computational models.” (Charniak and McDermott, 1985)</p> <p>“The study of the computations that make it possible to perceive, reason, and act.” (Winston, 1992)</p>
Acting Humanly <p>“The art of creating machines that perform functions that require intelligence when performed by people.” (Kurzweil, 1990)</p> <p>“The study of how to make computers do things at which, at the moment, people are better.” (Rich and Knight, 1991)</p>	Acting Rationally <p>“Computational Intelligence is the study of the design of intelligent agents.” (Poole <i>et al.</i>, 1998)</p> <p>“AI . . . is concerned with intelligent behavior in artifacts.” (Nilsson, 1998)</p>
Figure 1.1 Some definitions of artificial intelligence, organized into four categories.	

Learning is any process by which a system improves performance from experience.

Herbert Simon

Learning is any process by which a system improves performance from experience.

Herbert Simon

- **system**: (internal state of) animal/human/learning machine
- **process**: adaptation, change, optimization, algorithm
- **experience**: input, data, observations
- **performance**: loss, fitness, score, utility, objective

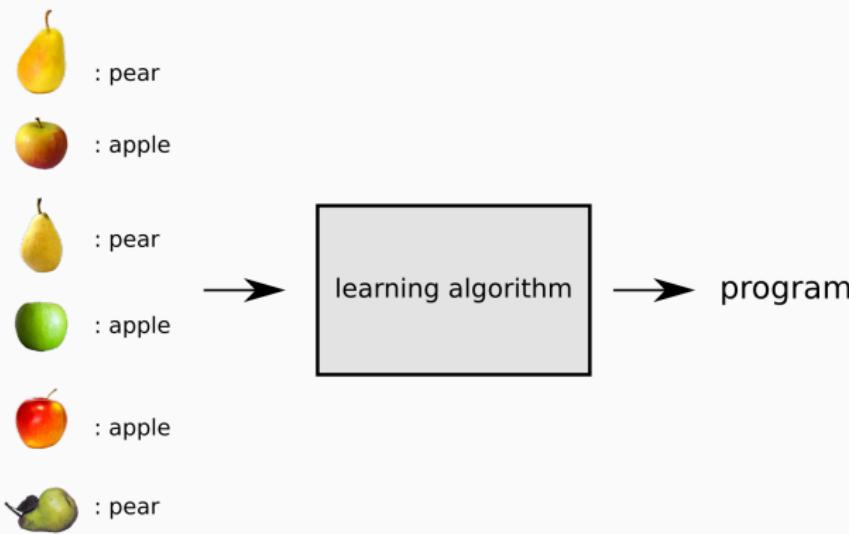
“Write a program which distinguishes apples from pears.”

“Write a program which distinguishes apples from pears.”

How to even start thinking about this task?

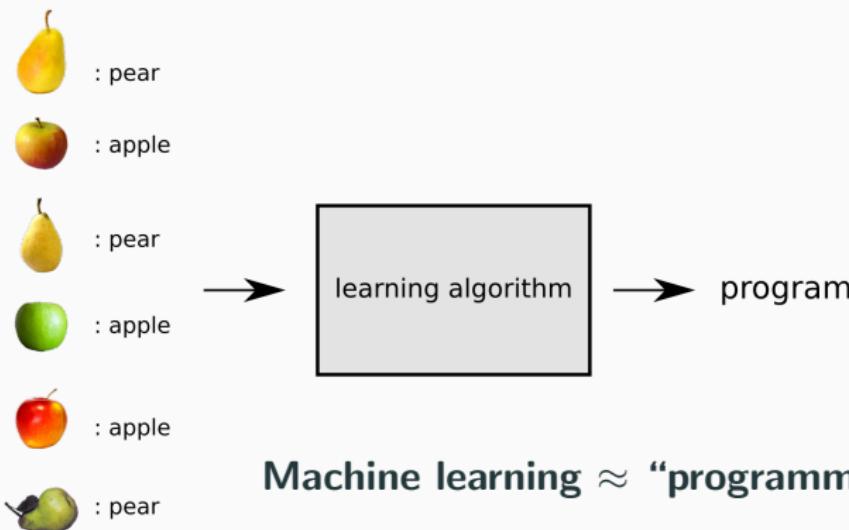
It might be easier to:

- write a *meta-program/learning machine*, which is
- provided with *training data (examples of input-output pairs)*
- and *outputs a program* solving the task



It might be easier to:

- write a *meta-program/learning machine*, which is
- provided with *training data (examples of input-output pairs)*
- and *outputs a program* solving the task



Machine learning \approx “programming with data”

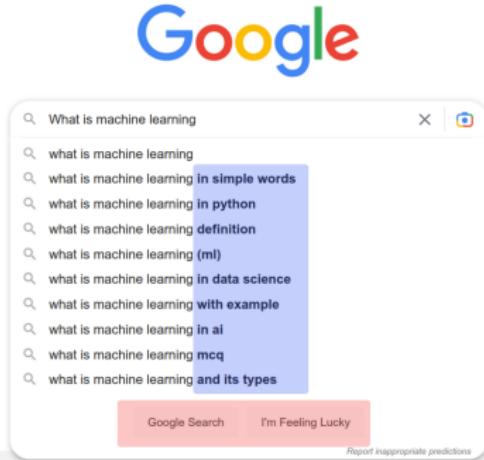
Spam Filtering

Spam in email started to become a problem when the Internet was opened up to the general public in the mid-1990s. It grew exponentially over the following years, and today composes some 80 to 85% of all the email in the world, by a “conservative estimate”

<https://en.wikipedia.org/wiki/Spamming>

data	prediction
<p>Natural _LoseWeight SuperFood Endorsed by Oprah Winfrey, Free Trial 1 bottle, pay only \$5.95 for shipping mfw rik <small>Spam X</small></p> <p>★ Jaquelyn Halley to nherrlein, bcc: thehorney, bcc: ang <small>show details 9:52 PM (1 hour ago)</small> </p> <p>==== Natural WeightLOSS Solution ==== Vital Acai is a natural WeightLOSS product that Enables people to lose weight and cleansing their bodies faster than most other products on the market. Here are some of the benefits of Vital Acai that You might not be aware of. These benefits have helped people who have been using Vital Acai daily to Achieve goals and reach new heights in there dieting that they never thought they could.</p> <ul style="list-style-type: none">* Rapid WeightLOSS* Increased metabolism - BurnFat & calories easily!* Better Mood and Attitude* More Self Confidence* Cleanse and Detoxify Your Body* Much More Energy* BetterSexLife* A Natural Colon Cleanse <p>http://sfsl.kaeconomic.cn http://slvk.kaeconomic.cn</p>	 <p>Spam vs. Not Spam</p>

Web Applications



Web Search

Natural Language Processing

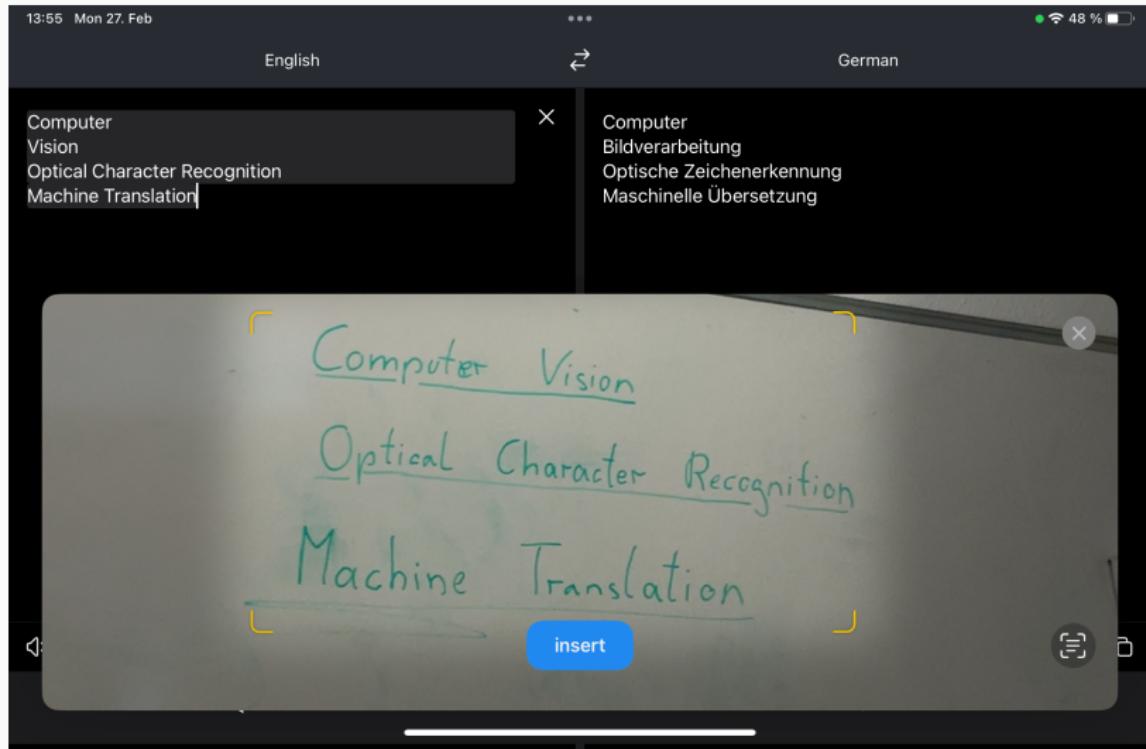
Advertising & Recommender Systems

A screenshot of a web page displaying a shopping cart for machine learning books. The page has a light green background with a white header bar. The header bar contains the text "Ads · Shop machine learning book". Below the header, there is a grid of book covers and their details. The books listed are:

Book Title	Author(s)	Publisher	Price	Shipping	Condition
Machine Learning	Tom M. Mitchell	MIT Press	\$85.55	Free shipping	New
An Introduction to Statistical Learning	Gareth James, Daniela Witten, Trevor Hastie, Robert Tibshirani	Springer	\$68.99	Free shipping	New
Hands-on Machine Learning with Scikit-Learn, Keras & TensorFlow	Aurélien Géron	O'Reilly	\$67.50	Free shipping	New
Introduction to Machine Learning with Python	Ciaran O'Donnell	MIT Press	\$47.99	Free shipping	New
An Introduction to Machine Learning	Kevin P. Murphy	MIT Press	\$120.99	Free shipping	New
Artificial Intelligence: A Modern Approach	John McCarthy, Peter M. Newell, Stuart Russell, Peter Norvig	Prentice Hall	\$57.99	Free shipping	New
Deep Learning (Adaptive Computation and Machine Learning series)	Ian Goodfellow, Yoshua Bengio, Aaron Courville	MIT Press	\$79.49	Free shipping	New
MACHINE LEARNING: AN ALGORITHMIC APPROACH, SECOND EDITION	Tom M. Mitchell	MIT Press	\$42.99	+€11.00 shipping	Used

The page also includes a "Next" button on the right side of the grid.

Computer Vision, Machine Translation



Chat GPT



How many atoms are there in the Harry Potter universe?



It is impossible to determine the exact number of atoms in the Harry Potter universe, as it is a fictional world with its own set of physical laws and properties. However, we can estimate the number of atoms in some of the physical objects that exist within the universe.



For example, if we consider Hogwarts Castle, we could estimate the number of atoms in the building materials used to construct it, such as stone, wood, and metal. However, this would be a very rough estimate, as we do not know the exact size, shape, and composition of the castle or its materials.

Self-driving Cars

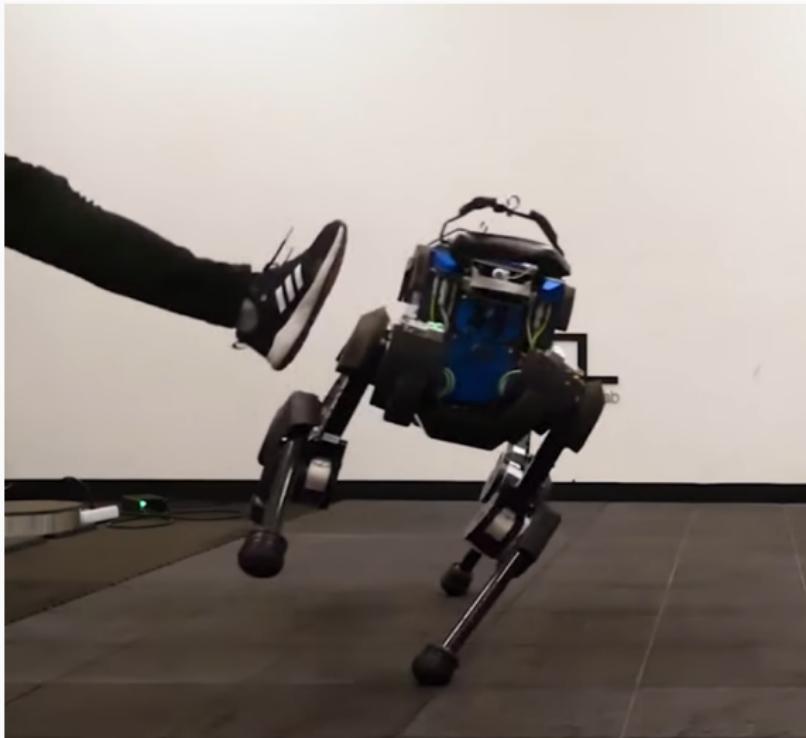
<https://www.youtube.com/watch?v=TsaES--0TzM>



- Many states have partially legalized self-driving cars (e.g. for testing)
- Full autonomous driving (SAE level 5, “steering wheel optional”) still a long way down the road

Robotics

<https://www.youtube.com/watch?v=3jDoPobFgwA>



Game Playing (Chess, Go, Video Games)



<https://www.youtube.com/watch?v=7L2sUGc0gh0>

Bioinformatics, Protein Folding

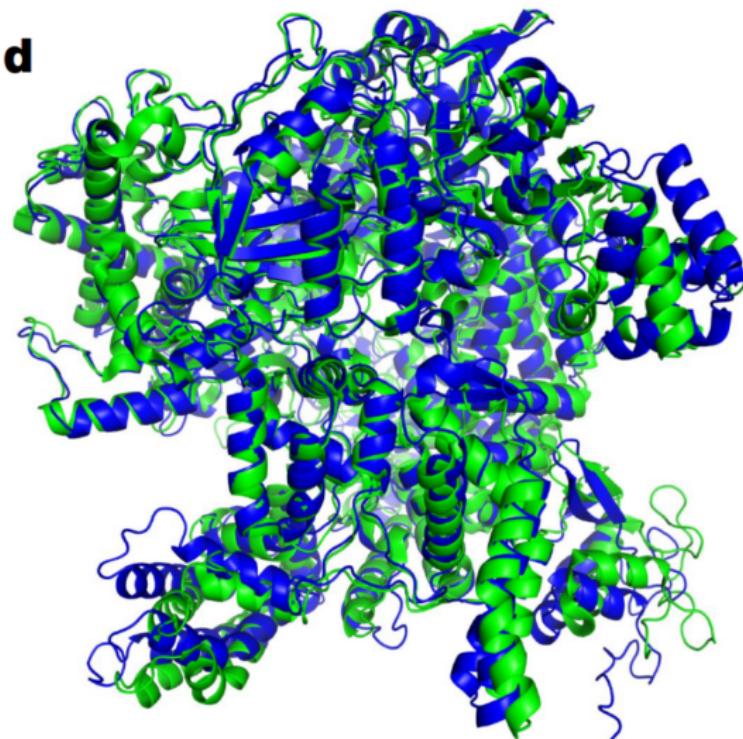


Image: Jumper et al., Highly accurate protein structure prediction with AlphaFold, Nature 2021

Image Generation

“Monkey with glasses and IPad, desparately trying to teach Machine Learning in front of black board with fancy math on it”



<https://stablediffusionweb.com>

Applications

- astronomy
- autonomous driving/navigation
- bio-informatics
- brain-machine interfaces
- chemistry
- computational finance
- data visualization
- drug discovery
- event detection
- fraud detection
- game playing
- image generation
- information retrieval
- medical diagnosis (support)
- natural language processing
- network analysis
- online advertising
- personal assistants
- recommender systems
- search engines
- speech recognition
- speech synthesis
- speech and language translation
- visual object detection
- ...

When is ML especially useful?

- When human expert knowledge is missing or when humans can only “do it intuitively”
 - driving a car
 - recognize visual objects
 - process language
- When task changes frequently
 - stock market prediction
 - weather forecasting
 - spam filters
- When task needs to be customized
 - recommender systems

Overview

Organizational Matters

Machine Learning Paradigms

Organizational Matters

What will you get from this course?

Lectures (ML1 VO)

Robert Peharz

- A general and broad introduction to machine learning
- Foundational and mathematical principles of machine learning
- Theoretical understanding of most important machine learning algorithms

Practicals (ML1 KU)

Thomas Wedenig

- Hands-on experience in implementing and applying algorithms
- Formally separate from VO, but highly recommended to be combined
- Wednesday, 6th March, 10:30 i13

Lectures

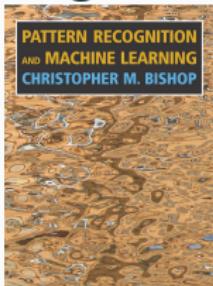
- Lectures are on Tuesdays, 11:30–13:15, HS i13
- 2 × 45 minutes, ~ 10 minutes break
- Easter break 26th March and 2nd April
- Pentecost 21st May
- In-person lectures, but no mandatory attendance
- Lectures will be recorded
- Last year's recordings are still available; however, content will slightly change this year

Materials

- Lecture slides on TeachCenter
- Bishop, **Pattern Recognition and Machine Learning**

Freely available under:

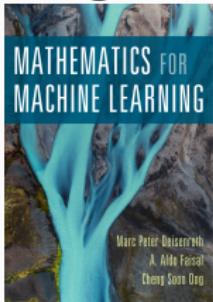
<https://www.microsoft.com/en-us/research/people/cmbishop/prml-book/>



- Deisenroth et al., **Mathematics for Machine Learning**

Freely available under:

<https://mml-book.github.io/>



Exam

- **Early exam:** (since some exchange students need to depart early)
Friday, 7th June, 14:15, HS i13
- **Main exam:**
Tuesday, 2nd July, 11:00, HS i13
- See TUGonline for further exam dates in the future
- Duration: 120 minutes
- 60% multiple choice questions, 40% open questions
- Grading scheme:

87.5	-	100	Sehr Gut (1)
75	-	87.4	Gut (2)
62.5	-	74.9	Befriedigend (3)
50	-	62.4	Genügend (4)
0	-	49.9	Nicht Genügend (5)

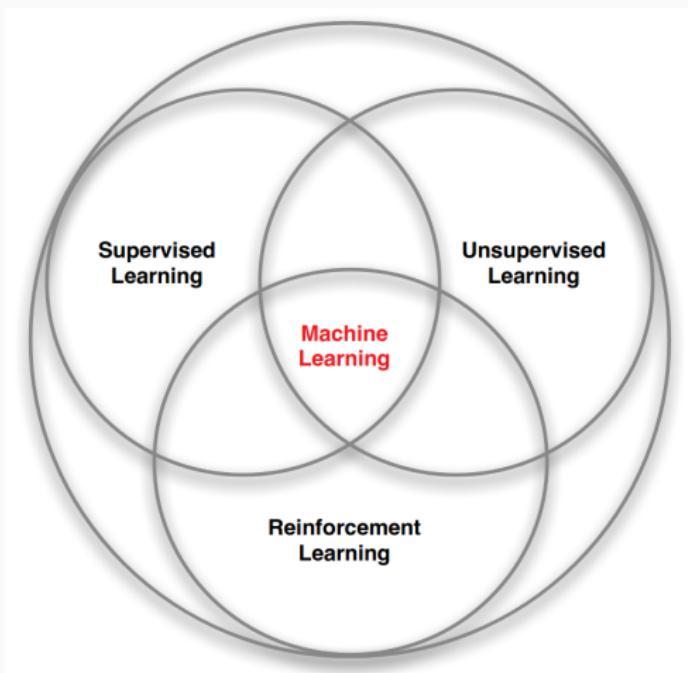
Contact

- The live lecture on Tuesday
- For general questions about the course, please use the forum on TeachCenter
- In particular, please ask content related questions in the forum (others might be interested in them as well)
- For all other questions, you can contact me under
`robert.peharz@tugraz.at`

Machine Learning Paradigms

Machine Learning Paradigms

Machine Learning is understood to consists of three paradigms:



Supervised Learning

- Teacher-student situation
 - Teacher gives you many examples of a concept (e.g. addition)
 - For the final exam, you shall solve new (but similar) examples on your own
-
- **Input-output characteristics**
 - **Predict target from input**

Supervised Learning (Classification)

Example

training targets
(labels)

	0	1	2	3	4	5	6	7	8	9
0	1	2	3	4	5	6	7	8	9	0
1	2	3	4	5	6	7	8	9	0	1
2	3	4	5	6	7	8	9	0	1	2
3	4	5	6	7	8	9	0	1	2	3
4	5	6	7	8	9	0	1	2	3	4
5	6	7	8	9	0	1	2	3	4	5
6	7	8	9	0	1	2	3	4	5	6
7	8	9	0	1	2	3	4	5	6	7
8	9	0	1	2	3	4	5	6	7	8
9	0	1	2	3	4	5	6	7	8	9

training inputs

Supervised Learning (Classification)

Example

training targets
(labels)

	0	1	2	3	4	5	6	7	8	9
0	1	2	3	4	5	6	7	8	9	0
1	2	3	4	5	6	7	8	9	0	1
2	3	4	5	6	7	8	9	0	1	2
3	4	5	6	7	8	9	0	1	2	3
4	5	6	7	8	9	0	1	2	3	4
5	6	7	8	9	0	1	2	3	4	5
6	7	8	9	0	1	2	3	4	5	6
7	8	9	0	1	2	3	4	5	6	7
8	9	0	1	2	3	4	5	6	7	8
9	0	1	2	3	4	5	6	7	8	9

training inputs

test input (not in training data)

2

Supervised Learning (Classification)

Example

training targets
(labels)

	0	1	2	3	4	5	6	7	8	9
0	1	2	3	4	5	6	7	8	9	0
1	2	3	4	5	6	7	8	9	0	1
2	3	4	5	6	7	8	9	0	1	2
3	4	5	6	7	8	9	0	1	2	3
4	5	6	7	8	9	0	1	2	3	4
5	6	7	8	9	0	1	2	3	4	5
6	7	8	9	0	1	2	3	4	5	6
7	8	9	0	1	2	3	4	5	6	7
8	9	0	1	2	3	4	5	6	7	8
9	0	1	2	3	4	5	6	7	8	9

training inputs

test input (not in training data)

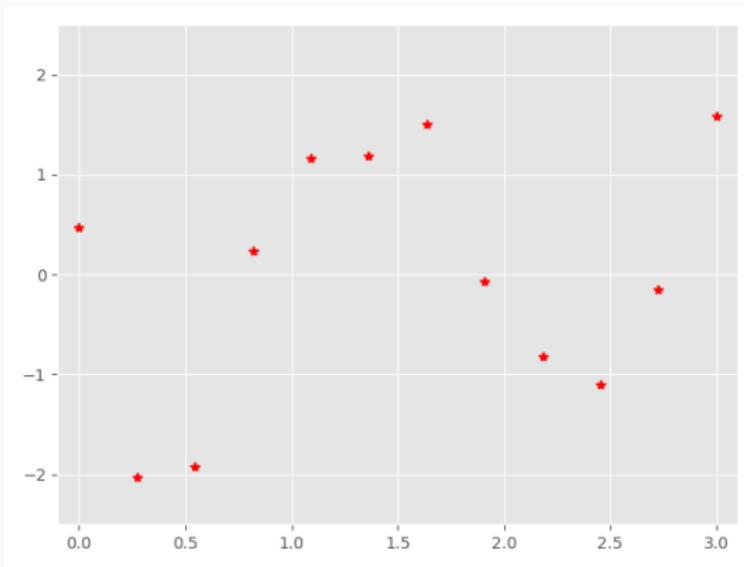


⇒ classifier will (hopefully) output '2'

Supervised Learning (Regression)

Example

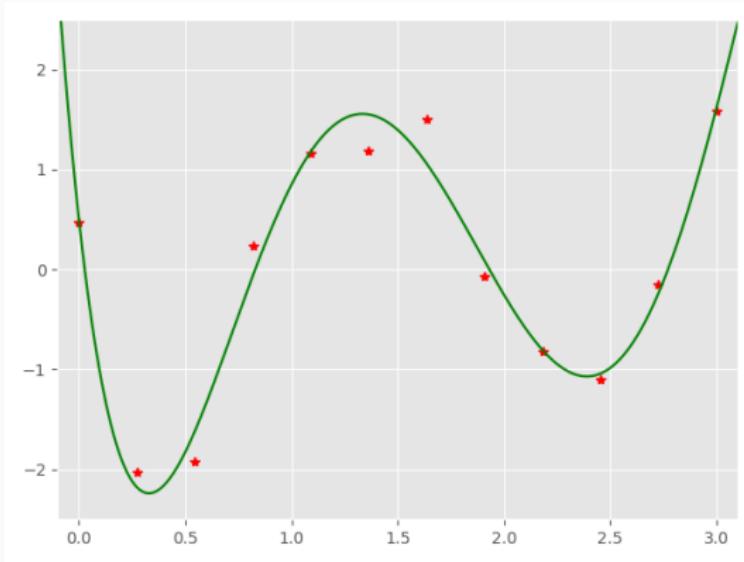
training data (independent variable x , target y)



Supervised Learning (Regression)

Example

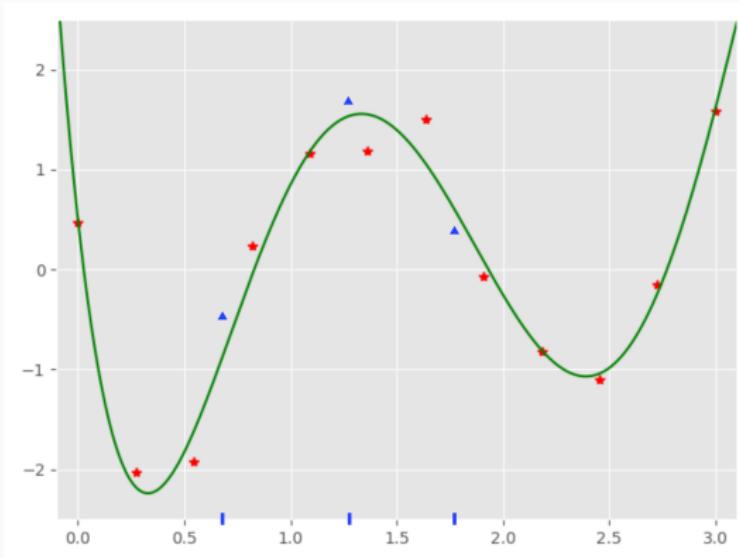
training data (independent variable x , target y)
model (polynomial degree 5)



training data (independent variable x , target y)

model (polynomial degree 5)

test data prediction error on test data will (hopefully) be small



What is the key difference between Classification and Regression?

What is the key difference between Classification and Regression?

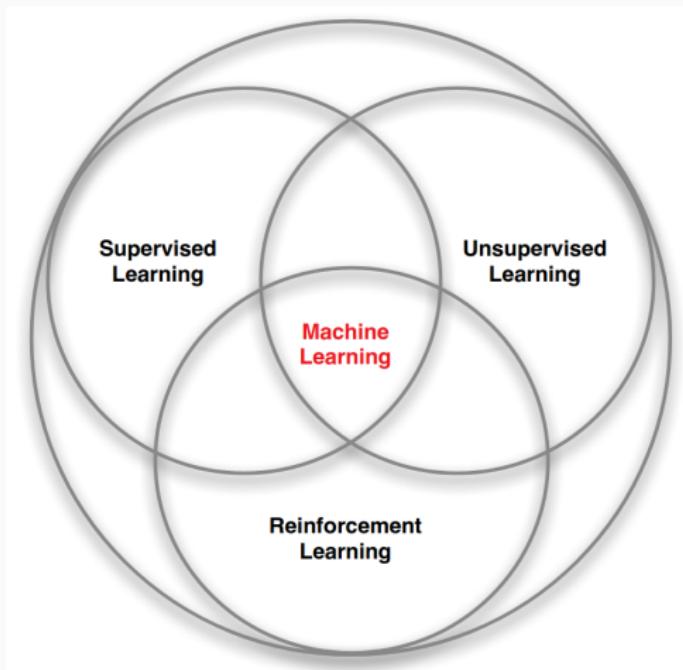
- Classification: discrete, unordered targets
- Regression: continuous, ordered targets

What is the key difference between Classification and Regression?

- Classification: discrete, unordered targets
- Regression: continuous, ordered targets

Note: there are also prediction tasks which are in between classification and regression, like Ordinal Regression, which uses discrete, but ordered targets (not covered in this course).

Machine Learning Paradigms



Unsupervised Learning

- You go to work by bike every day
- One day, you note a strange noise
- **No teacher**
- **Discover the world on your own**
- **Only inputs, or, outputs unknown**
- **Find interesting patterns in data**

0	1	2	3	4	5	6	7	8	9
0	1	2	3	4	5	6	7	8	9
0	1	2	3	4	5	6	7	8	9
0	1	2	3	4	5	6	7	8	9
0	1	2	3	4	5	6	7	8	9
0	1	2	3	4	5	6	7	8	9
0	1	2	3	4	5	6	7	8	9

0	1	2	3	4	5	6	7	8	9
0	1	2	3	4	5	6	7	8	9
0	1	2	3	4	5	6	7	8	9
0	1	2	3	4	5	6	7	8	9
0	1	2	3	4	5	6	7	8	9
0	1	2	3	4	5	6	7	8	9
0	1	2	3	4	5	6	7	8	9

- What are the clusters are in these data? (Clustering)

0	1	2	3	4	5	6	7	8	9
0	1	2	3	4	5	6	7	8	9
0	1	2	3	4	5	6	7	8	9
0	1	2	3	4	5	6	7	8	9
0	1	2	3	4	5	6	7	8	9
0	1	2	3	4	5	6	7	8	9
0	1	2	3	4	5	6	7	8	9

- What are the clusters are in these data? (Clustering)
- Which distribution generated these data? (Density Estimation, Generative Modeling)

0	1	2	3	4	5	6	7	8	9
0	1	2	3	4	5	6	7	8	9
0	1	2	3	4	5	6	7	8	9
0	1	2	3	4	5	6	7	8	9
0	1	2	3	4	5	6	7	8	9
0	1	2	3	4	5	6	7	8	9
0	1	2	3	4	5	6	7	8	9

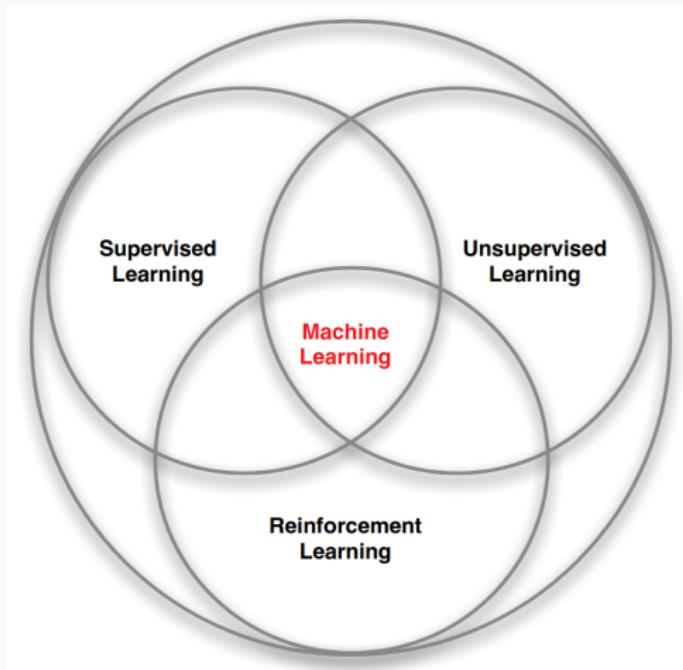
- What are the clusters are in these data? (Clustering)
- Which distribution generated these data? (Density Estimation, Generative Modeling)
- Is this a typical example?

2

0	1	2	3	4	5	6	7	8	9
0	1	2	3	4	5	6	7	8	9
0	1	2	3	4	5	6	7	8	9
0	1	2	3	4	5	6	7	8	9
0	1	2	3	4	5	6	7	8	9
0	1	2	3	4	5	6	7	8	9
0	1	2	3	4	5	6	7	8	9

- What are the clusters are in these data? (Clustering)
- Which distribution generated these data? (Density Estimation, Generative Modeling)
- Is this a typical example? 
- And this?  (Outlier Detection, Novelty Detection)

Machine Learning Paradigms



Reinforcement Learning [not covered in this course]

- You are a dog learning to sit on command
 - First, you don't know what is going on...
 - ...but at some point, you try to sit after command "sit!"
 - And you get a tasty reward!
-
- **Interact with your environment**
 - **Learn to make (sequential) decisions**
 - **Only feedback: reward signal**

Reinforcement Learning [not covered in this course]

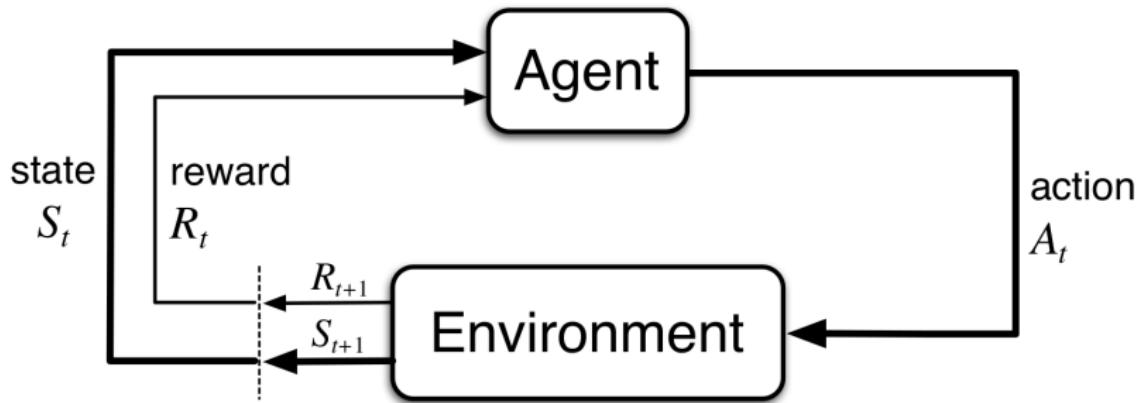


Image: Sutton & Barto

See course **Reinforcement Learning (VO + KU)** at master level

Paradigms, Paradigms, Paradigms

The traditional split of ML into Supervised, Unsupervised and Reinforcement Learning is perhaps somewhat outdated. Many more paradigms have been invented:

- Semi-supervised learning (labeled and unlabeled examples)
- Active learning (learn by picking examples to be labeled)
- Bayesian optimization (optimize an unknown function, while learning the function)
- Imitation learning (copy behavior)
- Coactive learning (user returns improved solution)
- Self-supervised learning (learn about the data by generating your own tasks)
- Transfer learning (adapt learned task to a similar task)
- Meta learning (learning to learn)

Course Content (might be subject to change)

1. Introduction
2. Math recap: calculus, linear algebra, probability, optimization
3. Math recap: calculus, linear algebra, probability, optimization
4. Linear regression
5. Logistic regression
6. Linear transformations
7. Neural Networks 1
8. Model selection, Evaluation
9. Neural Networks 2
10. K-Nearest Neighbours, Decision Trees, Random Forests
11. Support Vector Machines, Kernel Trick
12. K-Means clustering
13. Gaussian Mixture Models
14. Graphical Models