Machine Learning SoSe23

How much? 2 lectures and 1 tutorial per week, 9 credits

When? Lectures on Mon. 14.15 h & Thu. 12.15h (in person + Zoom + Recordings)
Tutorials to pick between: Wednesday, 12 pm (in person)
Thursday, 16 pm (in person)
Fridays at 10 am (in person)

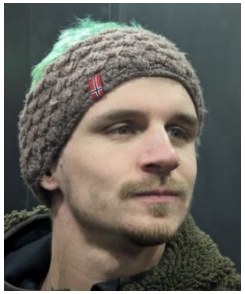
What? Theoretical & practical foundations of Machine Learning

More information and registration (by May 31th): https://cms.sic.saarland/ml24/

ML Team

Isabel Valera
(Lecturer)

Ayan Majumdar (TA)



Jonas Klesen (TA)

Deborah Kanubala (TA)

Gabriele Merlin (TA)

Michael Schott
(Tutor)

Shilpa Sharma
(Tutor)

Sofia Molotkova
(Tutor)

Aiman Al-Azazi
(Tutor)

Ghada Nait Said
(Tutor)

Fatemeh Moghaddam
(Tutor)

ML Timeline

	Week	Date	Lecture Nr.	Title
	1	18-Apr	1	Introduction
	2	22-Apr	2	Bayesian decision Theory
Block I - Foundations	2	25-Apr	3	Empirical Risk Minimization
	3	29-Apr	4	Linear regression I
	3	2-May	5	Linear regression II
	4	6-May	6	Linear classification
	4	9-May	Holiday	No Class
	5	13-May	7	Performance measures I
	5	16-May	8	Performance measures II
	6	20-May	Holiday	No Class
	6	23-May	9	Q&A session
	7	27-May	10	**Project Description (recording)**
Block II - Linear Models	7	30-May	Holiday	No Class
	8	3-Jun	**No** Holiday	No Class
	8	6-Jun	11	Convex Optimization I
	9	10-Jun	12	Convex Optimization II
	9	13-Jun	13	Linear SVM
	10	17-Jun	14	Intro to Kernels
Block III - Optimization & Kernel Meth	10	20-Jun	15	Learning with Kernels
	11	24-Jun	16	Societal impact I
Block IV - Societal Impact	11	27-Jun	17	Societal impact II
	12	1-Jul	18	Clustering I - Kmeans
	12	4-Jul	19	Clustering II - GMMs
Block V - Unsupervised Learning	13	8-Jul	20	PCA + Kernel PCA
	13	11-Jul	21	Deep learning I - MLPs + BP
	14	15-Jul	22	Deep learning II - CNNs
	14	18-Jul	23	Deep learning III - Transformers
	15	22-Jul	24	Beyond supervised DL
Block VI - Intro to Deep Learning	15	25-Jul	25	Q&A

Pre-requisites:

- linear algebra
- good basis on statistics

Tutorials:

- Help tutorials
- Solution tutorials
- Programing tutorials

Evaluation:

- Final written exam
- Project Report
- (Bonus) Project Challenge

ML Tutorials

If you would like to attend the ML tutorials, please **register to (ONLY!) one of the tutorial slots by Tuesday 23rd via CMS**!!

Id	Name	# Students	Start	End
1	**Tutorial group I -- Wednesdays at 14 hours**	0/150	18.04.24	23.04.24
2	**Tutorial group II -- Thursdays 16 hours**	0/99	18.04.24	23.04.24
3	**Tutorial group III -- Fridays at 10 hours**	0/99	18.04.24	23.04.24

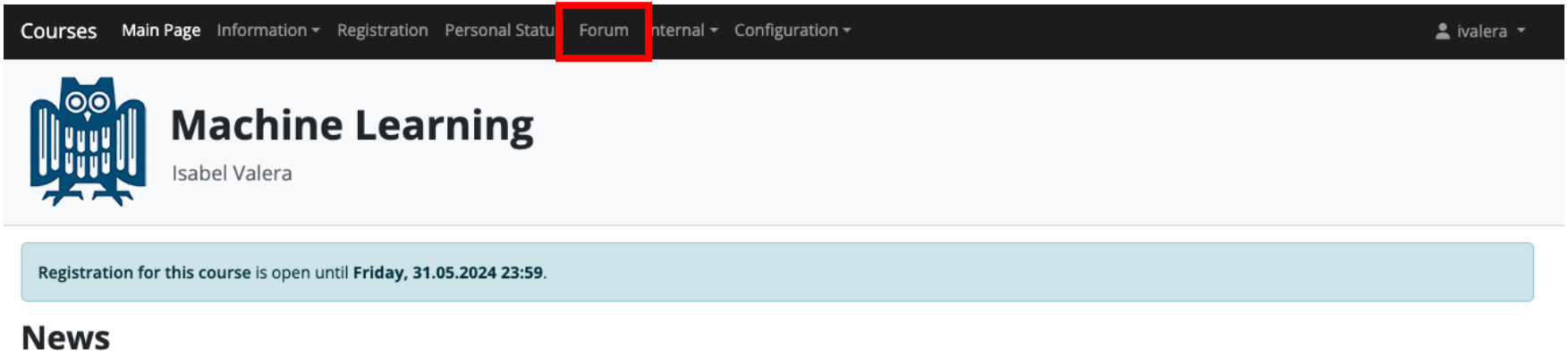
ML Evaluation

- Grade = Max(0.75* Final written exam + 0.25* Project report, Final written exam)
- Grade Bonus for top 5% teams in Project Challenge
- For exam registration, no admission pre-requisites (beyond University registration)
- Project voluntary but strongly recommended!
  (more details on May 27th)

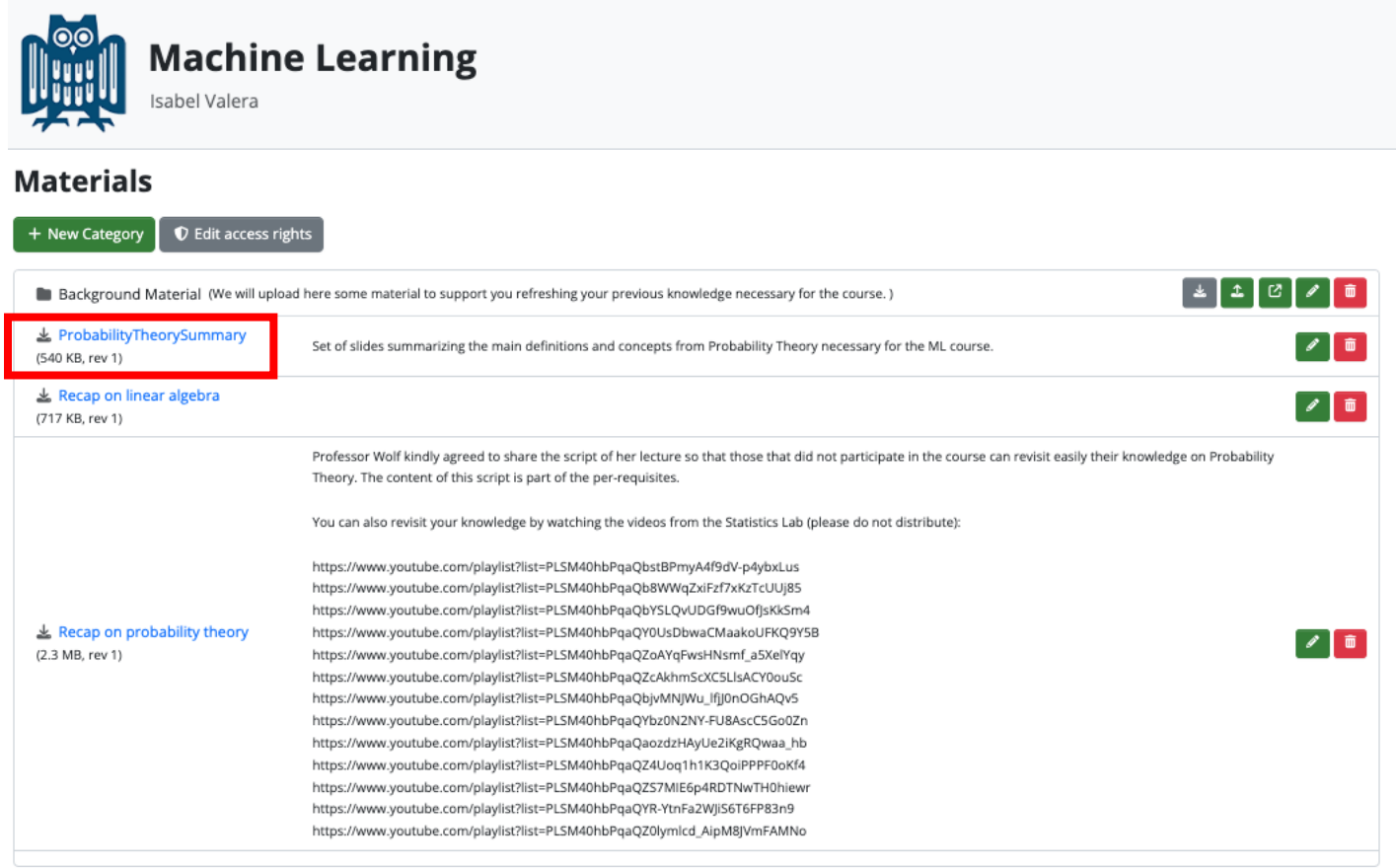
ML Q&A

Please ask your questions using one of the following options:

- Ask during or after (if private request) class
- Ask during turorials to our tutors
- Use the cms forum
- Do not send me emails, please!



ML Materials



Machine Learning

Isabel Valera

Materials

+ New Category | Edit access rights

Background Material (We will upload here some material to support you refreshing your previous knowledge necessary for the course.)

ProbabilityTheorySummary
(540 KB, rev 1)

Set of slides summarizing the main definitions and concepts from Probability Theory necessary for the ML course.

Recap on linear algebra
(717 KB, rev 1)

Recap on probability theory
(2.3 MB, rev 1)

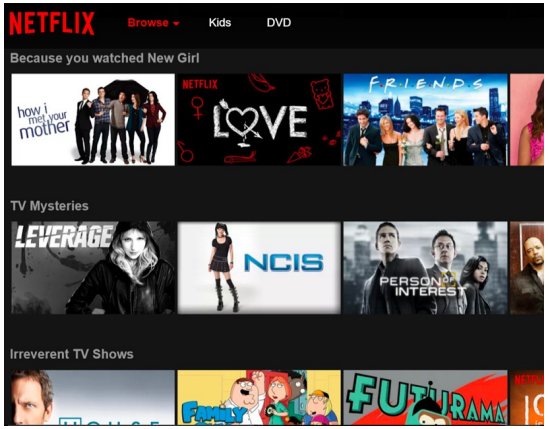
Professor Wolf kindly agreed to share the script of her lecture so that those that did not participate in the course can revisit easily their knowledge on Probability Theory. The content of this script is part of the per-requisites.

You can also revisit your knowledge by watching the videos from the Statistics Lab (please do not distribute):

https://www.youtube.com/playlist?list=PLSM40hbPqaQbstBPmyA4f9dV-p4ybxLus
https://www.youtube.com/playlist?list=PLSM40hbPqaQb8WWqZxiFzf7xKzTcUUj85
https://www.youtube.com/playlist?list=PLSM40hbPqaQbYSLQvUDGf9wuOfJsKkSm4
https://www.youtube.com/playlist?list=PLSM40hbPqaQY0UsDbwaCMaakoUFKQ9Y5B
https://www.youtube.com/playlist?list=PLSM40hbPqaQZoAYqFwsHNsmf_a5XelYqy
https://www.youtube.com/playlist?list=PLSM40hbPqaQZcAkhmScXC5LlsACY0ouSc
https://www.youtube.com/playlist?list=PLSM40hbPqaQbjvMNJWu_lfJJ0nOGhAQv5
https://www.youtube.com/playlist?list=PLSM40hbPqaQYbz0N2NY-FU8AscC5Go0Zn
https://www.youtube.com/playlist?list=PLSM40hbPqaQaozdzHAyUe2iKgRQwaa_hb
https://www.youtube.com/playlist?list=PLSM40hbPqaQZ4Uoq1h1K3QoiPPPF0oKf4
https://www.youtube.com/playlist?list=PLSM40hbPqaQZS7MIE6p4RDTNwTH0hiewr
https://www.youtube.com/playlist?list=PLSM40hbPqaQYR-YtnFa2WJiS6T6FP83n9
https://www.youtube.com/playlist?list=PLSM40hbPqaQZ0lymIcd_AipM8JVmFAMNo

ML is everywhere

[Amazon]



[Netflix]

[AppleNews]

The New York Times

Banking Start-Ups Adopt New Tools for Lending

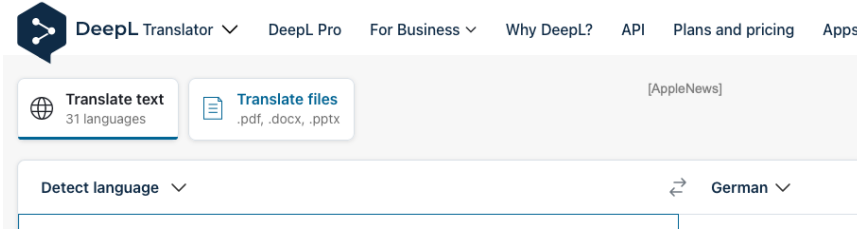
``Big-data lending, though, relies on software algorithms...“

FINANCE MAGNATES

AI and Personalization in Insurance: Customizing Policies and Customer Experiences

Monday, 10/04/2023 | 15:19 GMT by Finance Magnates Staff

The AI revolution is here



[Deepl.com]

FORBES > LEADERSHIP

HR Recruiting Software In The Era Of AI

Maxime Legardez Coquin Forbes Councils Member
Forbes Human Resources Council
COUNCIL POST | Membership (Fee-based)

Apr 11, 2023, 07:30am EDT

CEO of Maki People, a skills assessment tool to hire better and faster.

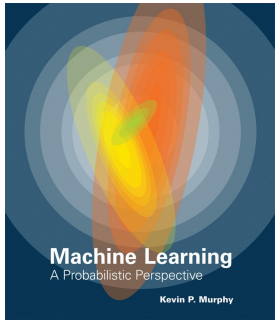
ML Definitions

(Picture from Wikipedia)

Arthur Samuel (1959): *"Machine Learning is the field of study that gives the computer the ability to learn without being explicitly programmed."*

(Image from Tom Mitchell's homepage)

Tom Mitchell (1998): *"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."*



Kevin Murphy's book (2012): *"we define machine learning as a set of methods that can automatically detect patterns in data, and then use the uncovered patterns to predict future data, or to perform other kinds of decision making under uncertainty (such as planning how to collect more data!)"*

ML Definition by ChatGPT

what is machine learning?

Machine learning is a subfield of artificial intelligence that involves developing algorithms and models that can automatically learn from data, without being explicitly programmed. It involves building mathematical models and algorithms that can identify patterns and make predictions or decisions based on the data.

The process of machine learning typically involves several steps, including collecting and preparing data, selecting a model or algorithm, training the model on the data, testing the model to evaluate its performance, and then deploying the model to make predictions or decisions on new data.

There are several different types of machine learning, including supervised learning, unsupervised learning, and reinforcement learning. In supervised learning, the model is trained on labeled data, where the correct output is already known, and the goal is to learn to generalize to new, unseen data. In unsupervised learning, the model is trained on unlabeled data, where there is no predetermined correct output, and the goal is to discover underlying patterns or structure in the data. Reinforcement learning involves training a model to interact with an environment and learn through trial and error to maximize a reward signal.

why is it important to learn the theoretical foundations of machine learning?

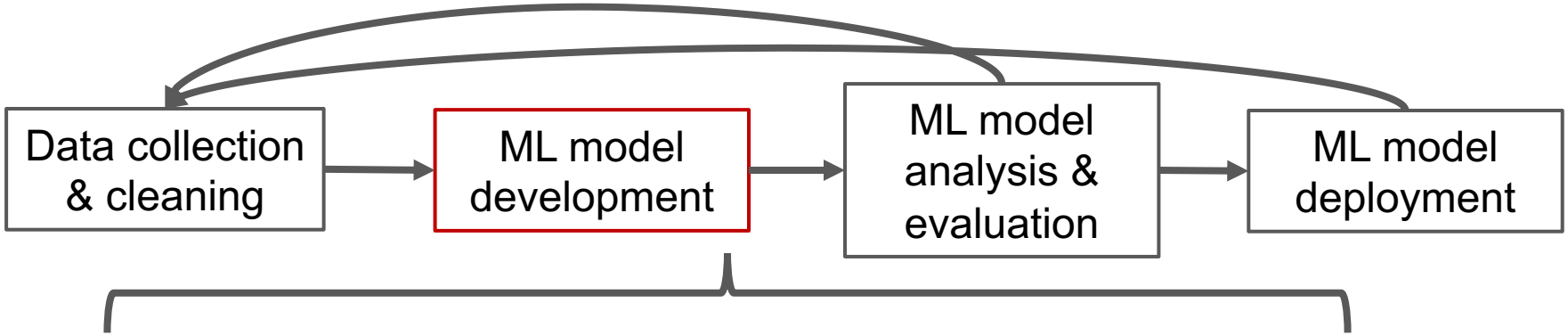
It is important to learn the theoretical foundations of machine learning because it provides a deep understanding of the underlying concepts and principles that drive the development and application of machine learning models. Understanding the theory behind machine learning can help practitioners develop better models, make more informed decisions, and avoid common pitfalls.

Some of the key benefits of learning the theoretical foundations of machine learning include:

1. Better model selection: Understanding the theoretical foundations of machine learning can help practitioners select the right model for a given problem. This can help improve the accuracy and efficiency of the model.
2. Robustness: Understanding the theory behind machine learning can help practitioners develop models that are more robust to changes in the data or the environment. This can help ensure that the model continues to perform well even when faced with new or unexpected situations.
3. Interpretability: Understanding the theory behind machine learning can help practitioners develop models that are more interpretable, meaning that they can provide insights into how the model is making its predictions or decisions.
4. Generalization: Understanding the theoretical foundations of machine learning can help practitioners develop models that generalize well to new, unseen data. This is critical for ensuring that the model can be applied to real-world problems.
5. Innovation: Understanding the theoretical foundations of machine learning can help practitioners develop new algorithms and techniques that can push the field forward and lead to new breakthroughs.

In short, learning the theoretical foundations of machine learning is critical for anyone who wants to develop and apply machine learning models effectively and efficiently.

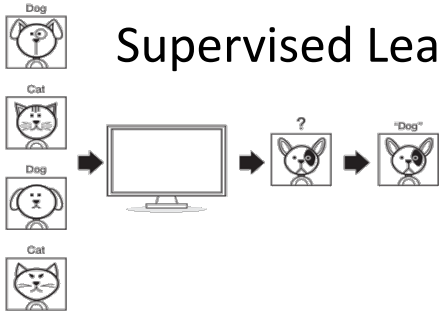
ML Pipeline



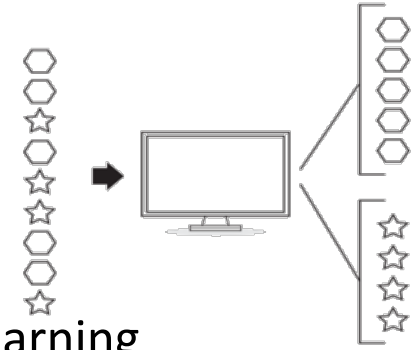
- Data pre-processing (e.g., data normalization, feature selection, etc.)
- **Model family choice (based on task, data properties, etc.)**
- **Model training & validation**

ML Taxonomy

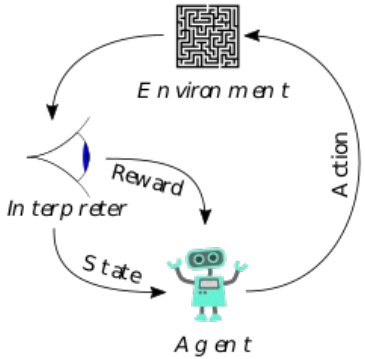
Supervised Learning



Unsupervised Learning

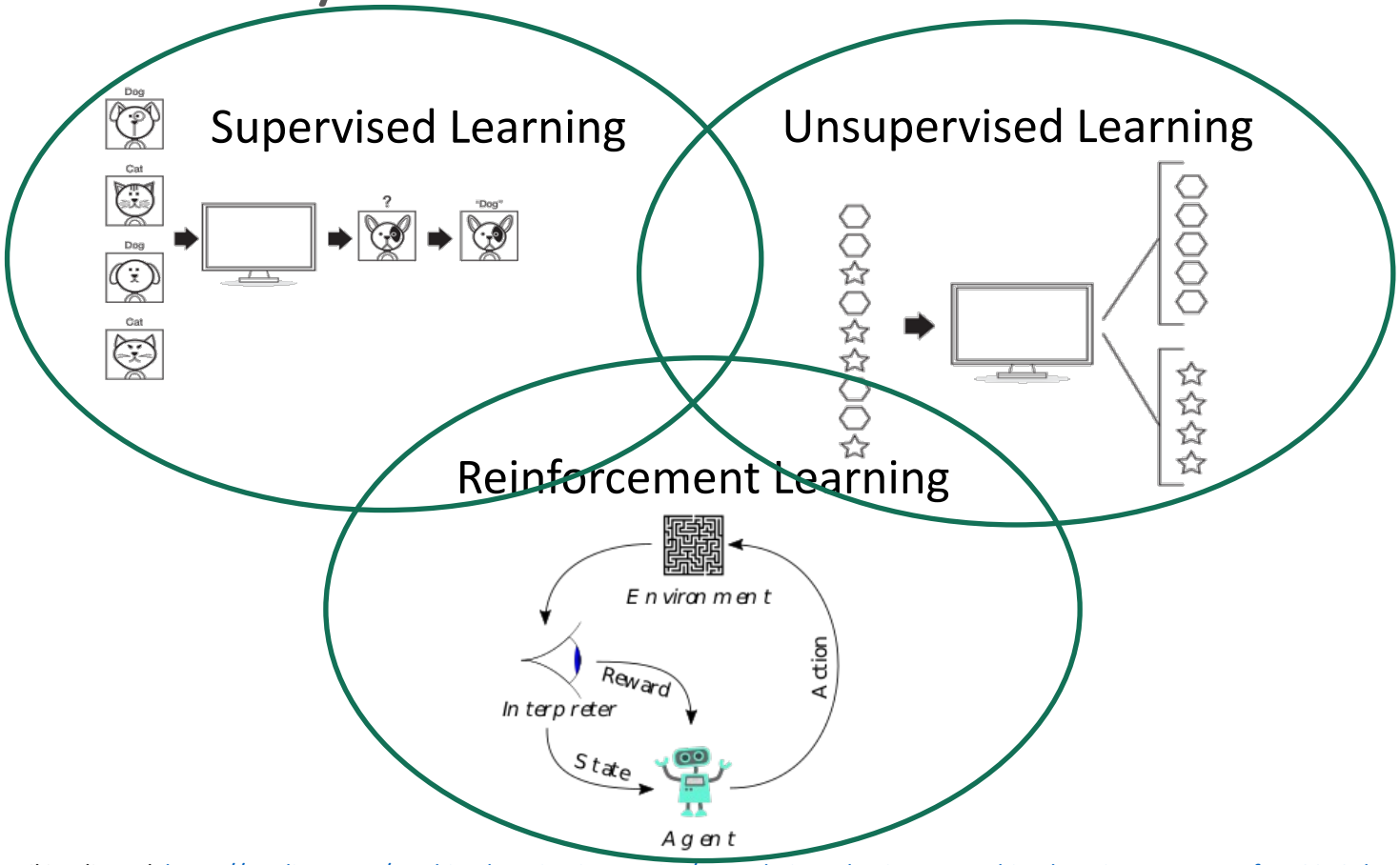


Reinforcement Learning



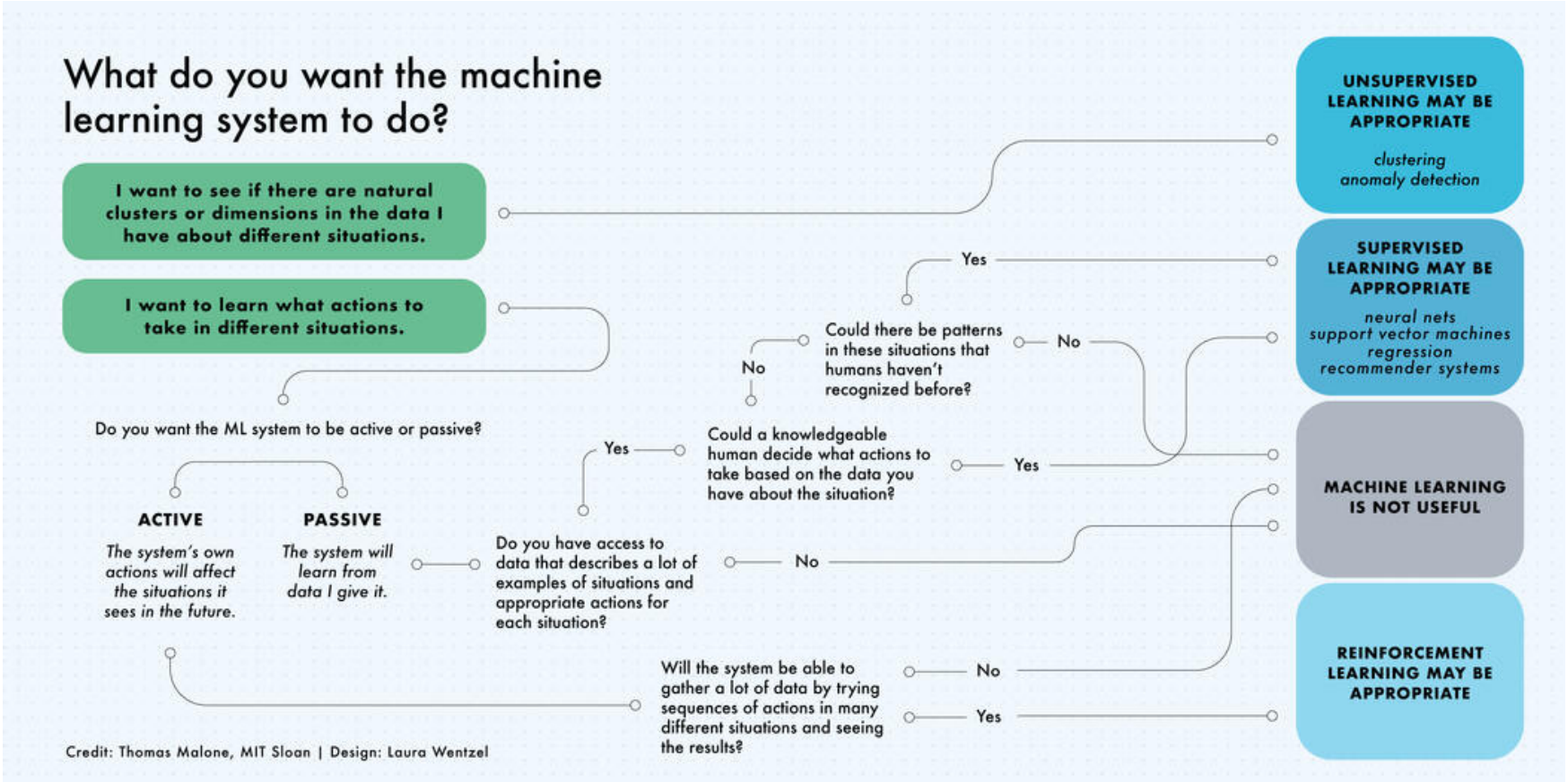
Images from Wikipedia and: https://medium.com/machine-learning-in-practice/a-gentle-introduction-to-machine-learning-concepts-cfe710910eb

ML Taxonomy



Images from Wikipedia and: https://medium.com/machine-learning-in-practice/a-gentle-introduction-to-machine-learning-concepts-cfe710910eb

ML Taxonomy



Images from https://mitsloan.mit.edu/ideas-made-to-matter/machine-learning-explained

ML, the basics (Block I)

- Probability theory

Bishop's book: *"A key concept in the field of pattern recognition is that of **uncertainty**. It **arises both through noise on measurements, as well as through the finite size of data sets**. Probability theory provides a consistent framework for the quantification and manipulation of uncertainty and forms one of the central foundations for pattern recognition."*

- Bayesian decision Theory

Bishop's book: *"decision theory that, when combined with probability theory, allows us to **make optimal decisions in situations involving uncertainty** such as those encountered in pattern recognition."*

- Empirical Risk Minimization

Vapnik 1991: *"**Learning is posed as a problem of function estimation**, for which two principles of solution are considered: empirical risk minimization and structural risk minimization."*

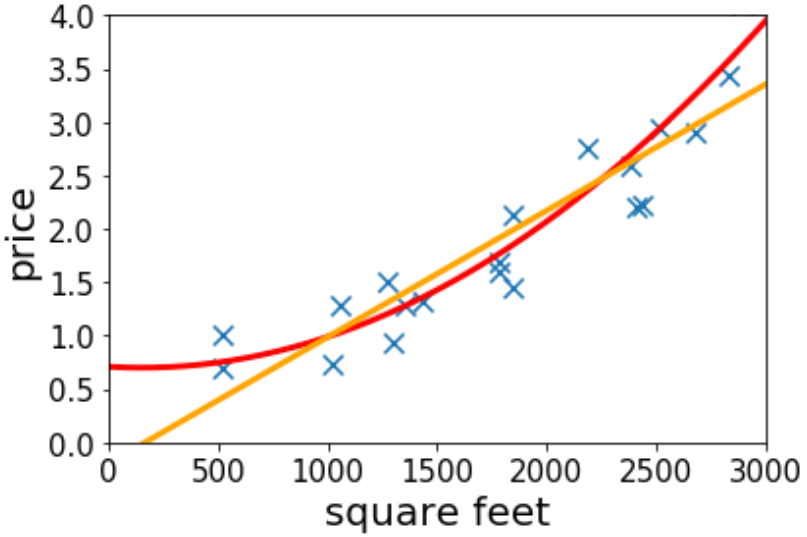
Supervised learning (Blocks II & III)

- Given a dataset that contains samples (or features/outcomes pairs):
$(x_n, y_n)_{n=1...N}$

- Task: Predict outcome (y) from features (x).

- Supervision comes from y!

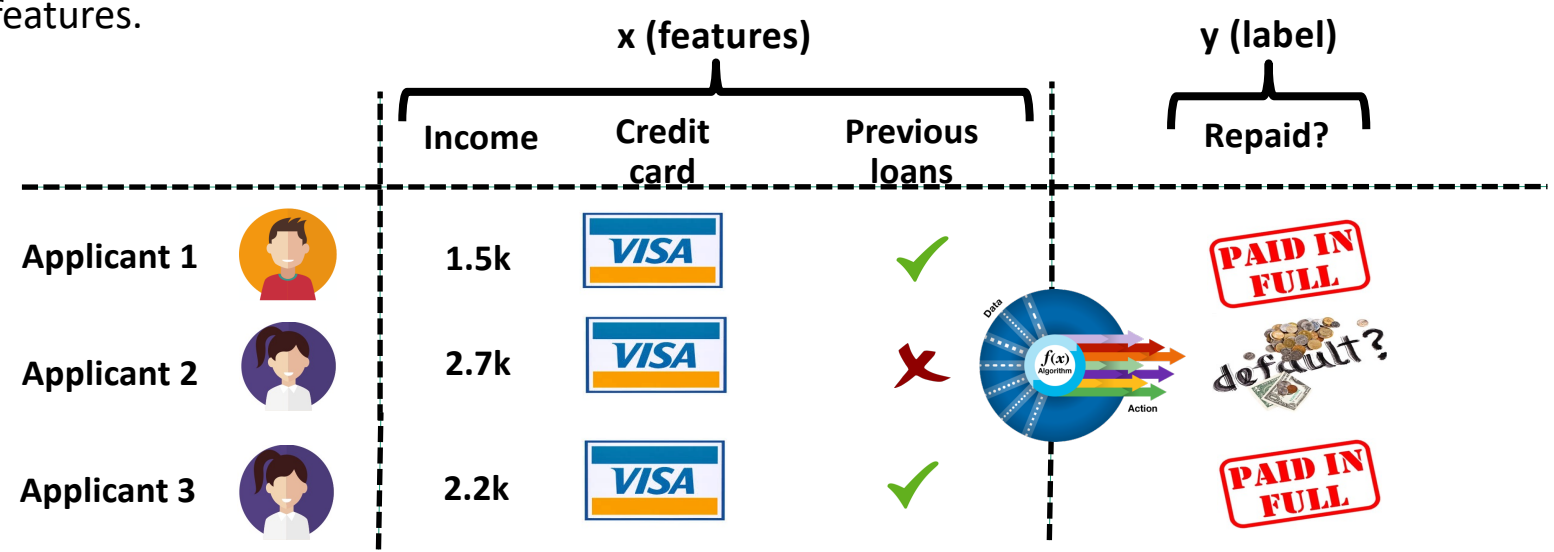
Example of regression problem: We observe the square feet of a house and its price. We aim to predict the price of unseen houses based on their size.



Example from: http://cs229.stanford.edu/syllabus.html

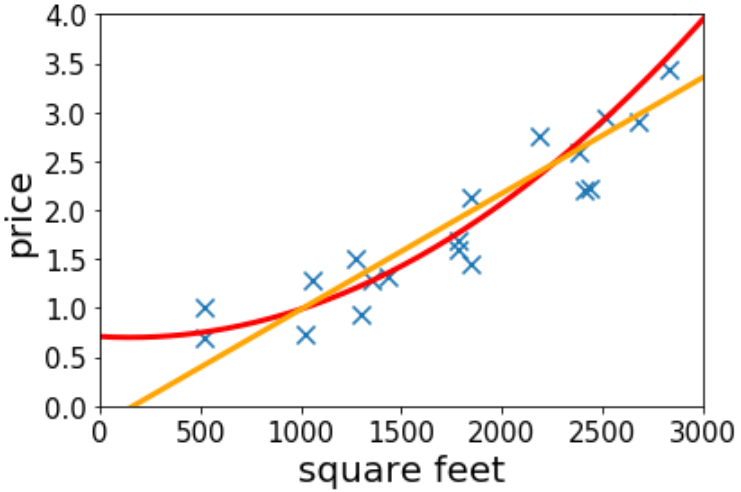
Supervised learning (Blocks II & III)

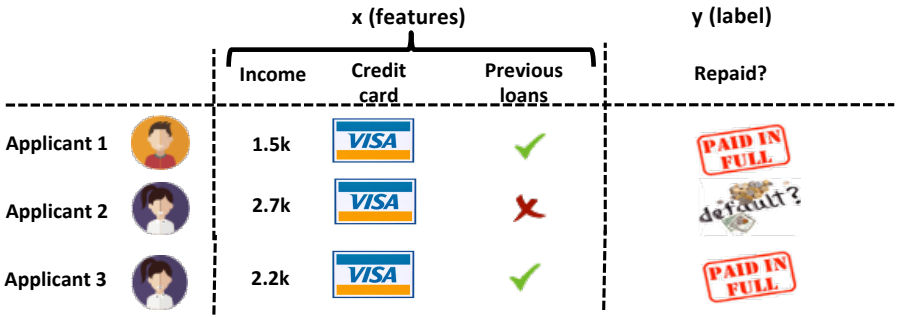
Example of classification problem: We observe a set of features characterizing every loan/credit applicant together with information on whether they repaid or defaulted to repay the credit in the past. The algorithm aims to predict for a new applicant is he/she will repay based on the set of features.



Supervised learning – Classification & Regression

- Regression: if is a continuous variable
  - e.g., price prediction
- Classification: the label is a discrete (nominal) variable
  - e.g., the task of predicting loan repayment





Supervised learning – Societal Impact (Block IV)

TIME

Google Has a Striking History of Bias Against Black Girls

MIT News
ON CAMPUS AND AROUND THE WORLD

The privacy risks of compiling mobility data

Merging different types of location-stamped data can make it easier to discern users' identities, even when the data is anonymized.

Bloomberg Businessweek

Artificial Intelligence Has Some Explaining to Do

Healthcare IT News

AI in healthcare - not so fast? Study outlines challenges, dangers for machine learning

INDEPENDENT

London terror attack: Uber slammed for being slow to turn off 'surge pricing' after rampage

BBC

AAAS: Machine learning 'causing science crisis'

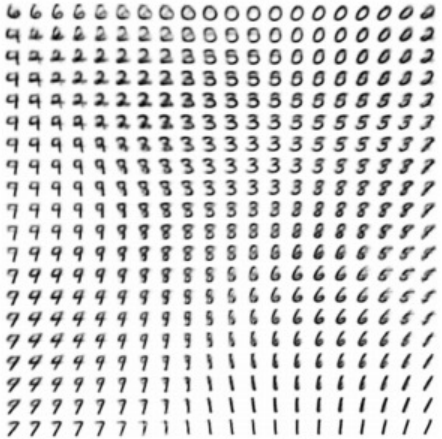
ImageNet Large Scale Visual Recognition Challenge. Russakovsky et al.'2015

Unsupervised learning (Block V)

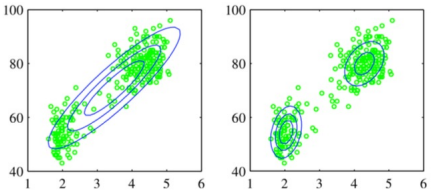
- Dataset contains no labels.
- Goal (vaguely-posed): to find interesting structures/patterns in the data

- Examples:

Generate new samples



Density estimation



Data exploration

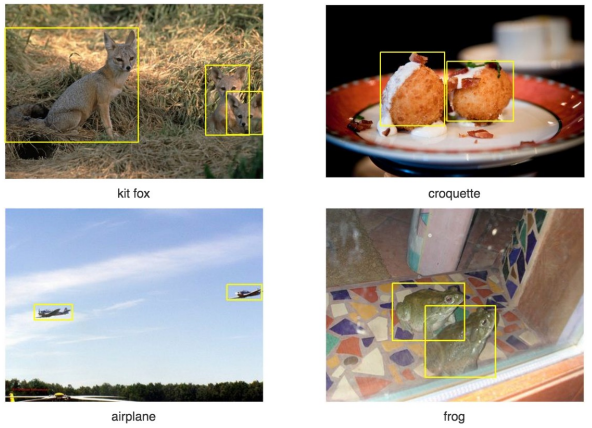
In the ML course:

- Clustering: Kmeans & GMMs
- (Kernel) PCA

Supervised learning – Deep Learning (Block VI)

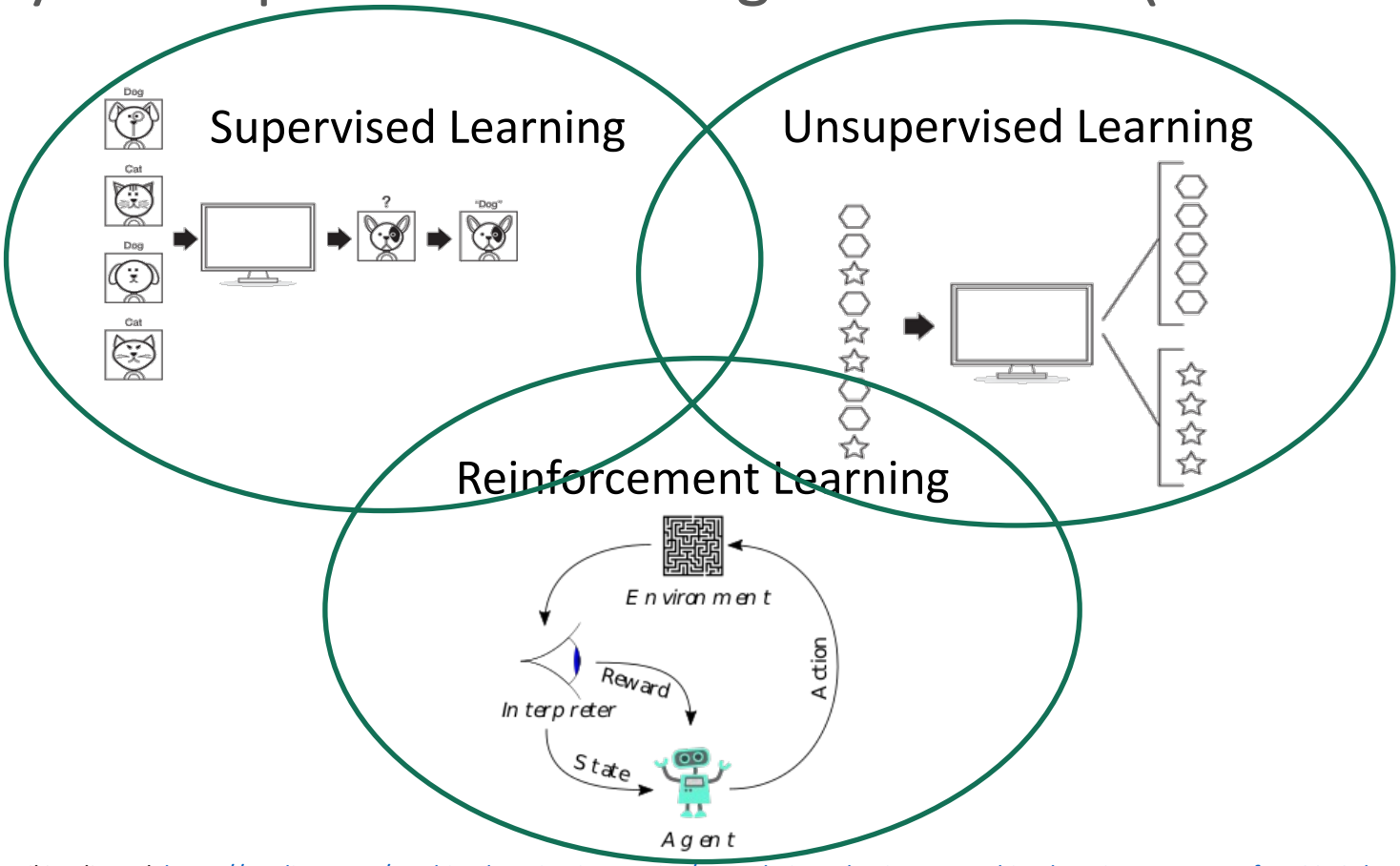
- Image Classification
  - raw pixels of the image,

- Object location and detection
  - raw pixels of the image
  - bounding box



ImageNet Large Scale Visual Recognition Challenge. Russakovsky et al.'2015

Beyond supervised learning – Overview (Last lecture)



Images from Wikipedia and: https://medium.com/machine-learning-in-practice/a-gentle-introduction-to-machine-learning-concepts-cfe710910eb