Introduction to Machine Learning

Borchert, Dr. Schapranow

Data Management for Digital Health

Winter 2023

Hasso Plattner Institut
Digital Engineering · Universität Potsdam

Agenda
Pillars of the Lecture

Medical Use Cases
- Biology Recap
- Oncology
- Nephrology
- Infectious Diseases

Technology Foundation
- Data Sources
- Data Formats
- Processing and Analysis
- Software Architectures

Machine Learning
- Data
- ML (Refine, Evaluate)
- Prediction + Probability

Agenda
Pillars of the Lecture

Medical Use Cases

- Biology Recap
- Oncology
- Nephrology
- Infectious Diseases

Technology Foundation

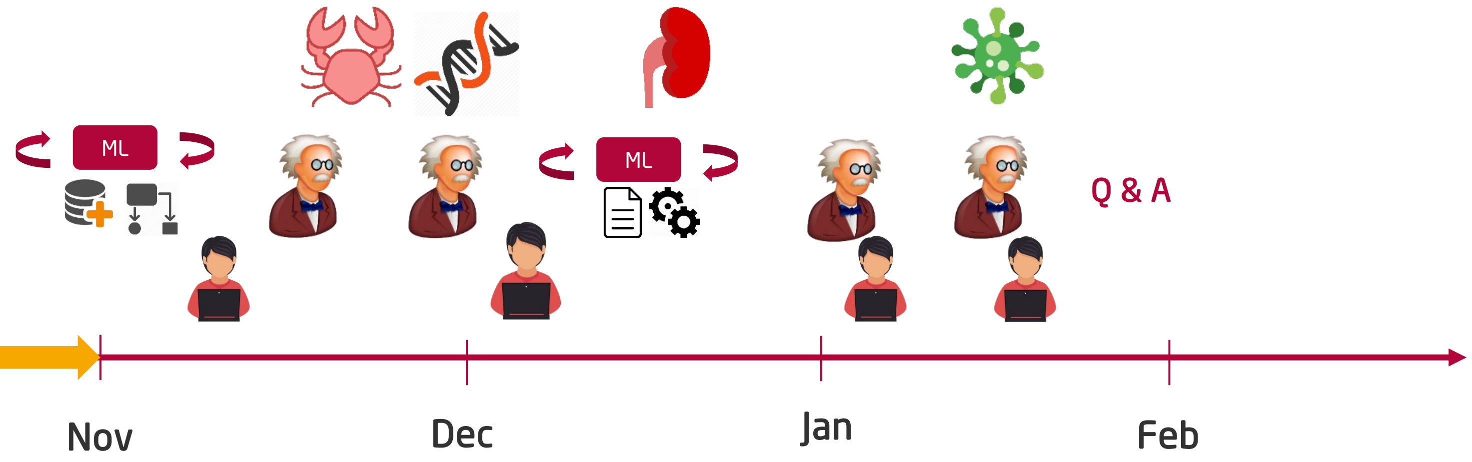
- Data Sources
- Data Formats
- Processing and Analysis
- Software Architectures

Machine Learning

- Data
- ML (Refine, Evaluate)
- Prediction + Probability

Hasso Plattner Institut

Lecture Schedule



- Lecture Kickoff
- Actors in Healthcare
- Digital Health Data

- Machine Learning (ML) Foundations
- Use Case Oncology
- Biology Recap

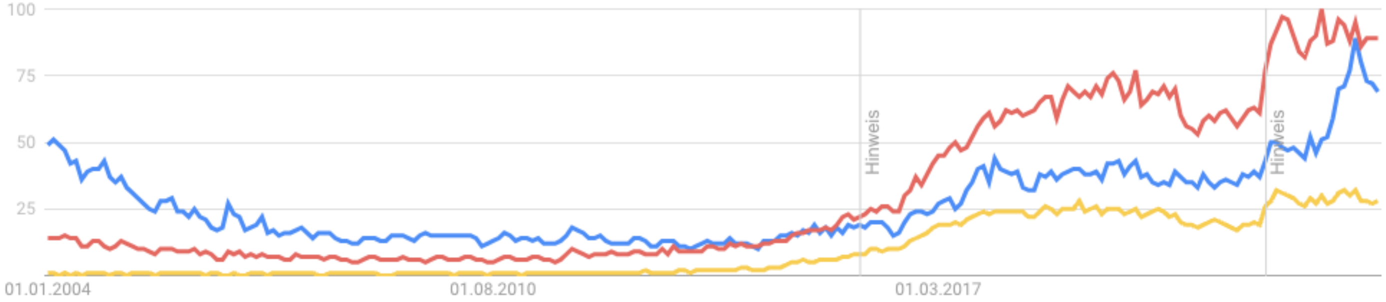
- Natural Language Processing
- Use Case Nephrology & Intensive Care
- Supervised ML & Deep Learning

- Use Case Infectious Diseases
- Unsupervised ML

Agenda

- Basic ML Terminology
- Problem Settings in ML
- First ML Algorithm

AI, ML, DL: Who's Afraid of AI, ML and DL?



https://trends.google.com/ on Sep 18, 2023

Artificial Intelligence Machine Learning Deep Learning

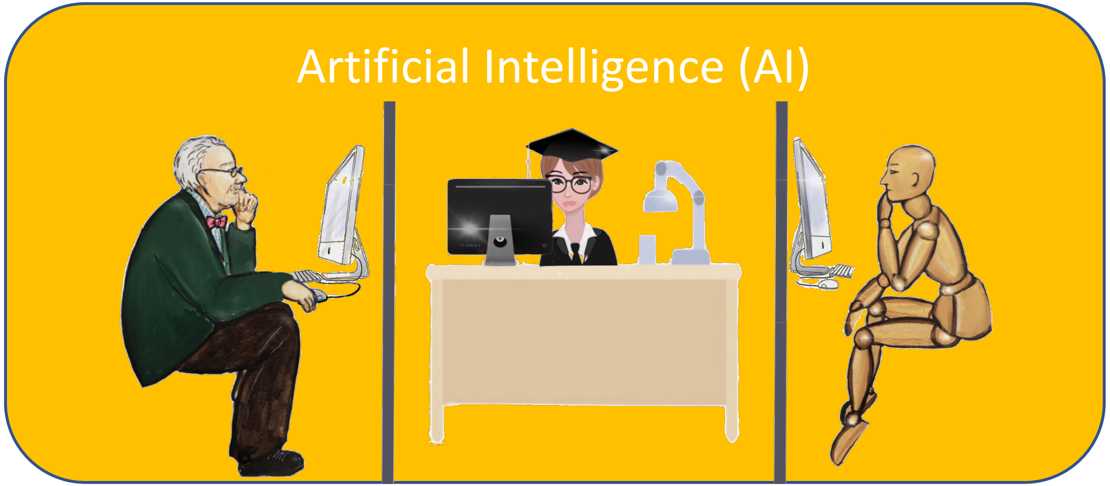
Artificial Intelligence: Solution for All Data Handling Problems?

Hasso Plattner Institut

Artificial Intelligence (AI)

Artificial Intelligence: Roots

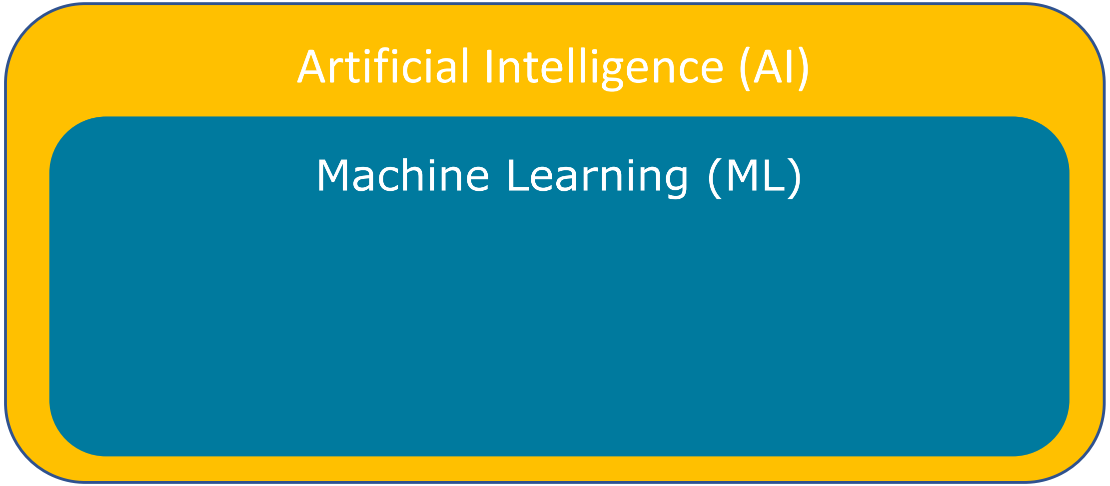
- Idea is as old as computer science
- Alan Turing already proposed an AI test in 1950
  - Ask a physically remote client questions (human and computer)
  - If tester cannot distinguish b/w human and computer Turing test passed



https://wsimag.com/science-and-technology/36961-no-turing-test-for-consciousness

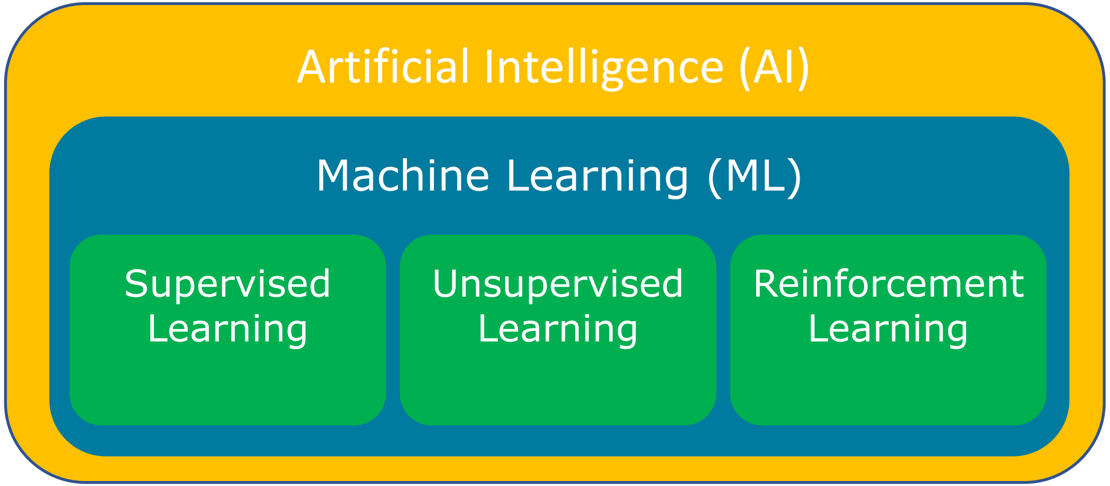
Artificial Intelligence: Key Facts

- Today, often ML algorithms are referred to when someone talks about AI
- Machine learning := pattern recognition through data analysis



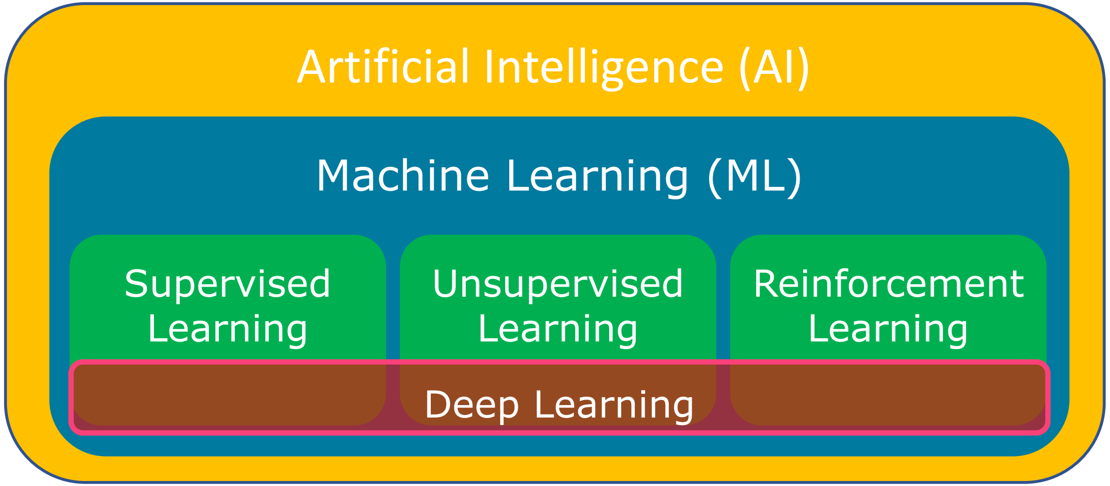
Artificial Intelligence: Key Facts

- Supervised learning := **operator defines** what to detect and provides **input data**
- Unsupervised learning := **algorithms detects** what to detect based on **input data**
- Reinforcement learning := actor interacts with **environment** to maximize reward

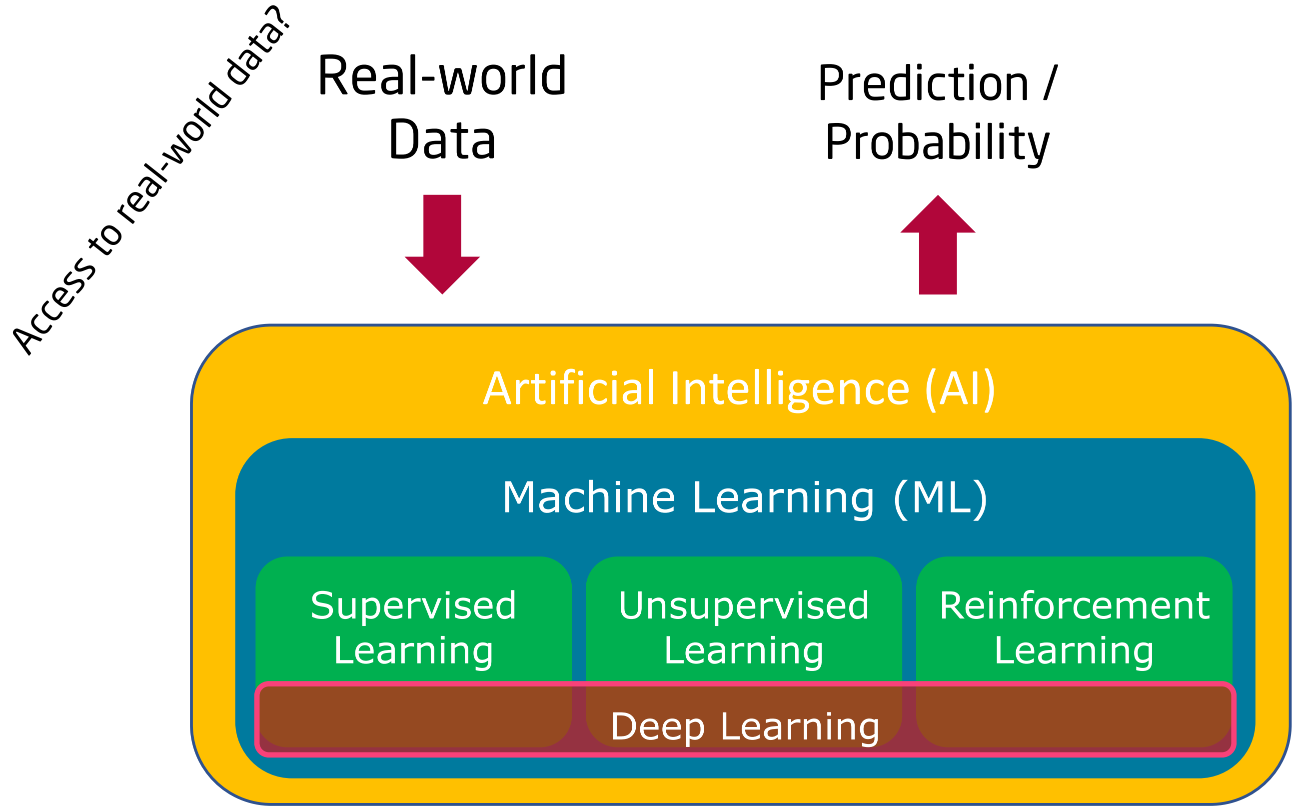


Artificial Intelligence: Key Facts

- Supervised learning := **operator defines** what to detect and provides **input data**
- Unsupervised learning := **algorithms detects** what to detect based on **input data**
- Reinforcement learning := actor interacts with **environment** to maximize reward
- Deep learning := Machine Learning with deep neural networks



Artificial Intelligence: Key Facts



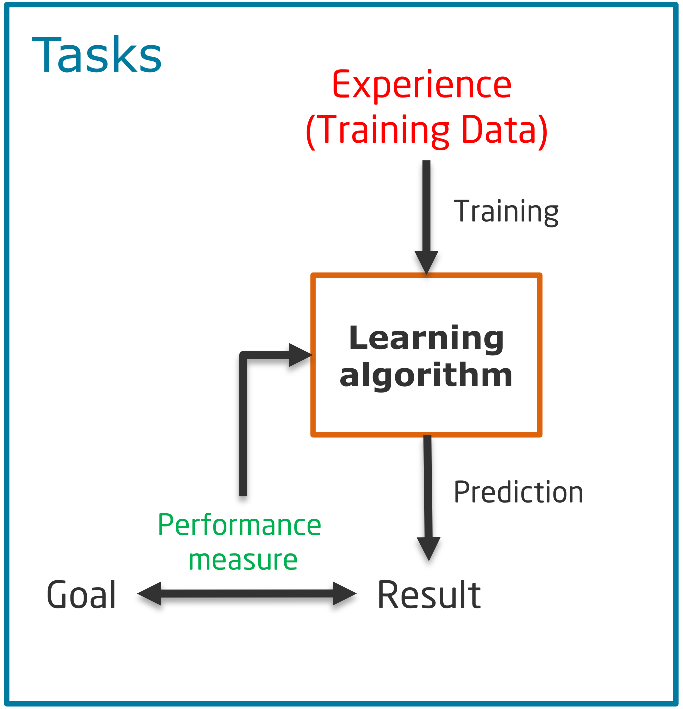
Using Data to Train Machines

Hasso Plattner Institut

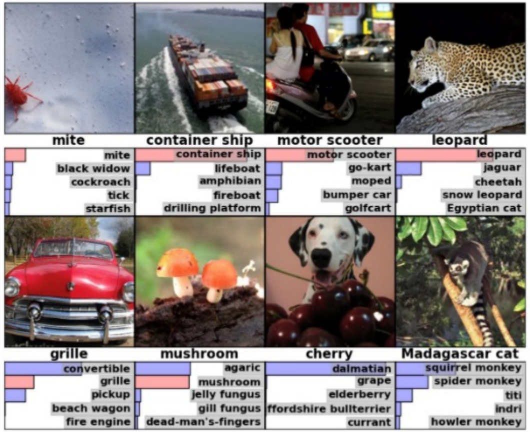
Machine Learning

“Field of study that gives computers the ability to learn without being explicitly programmed”
(Arthur Samuel, 1959)

“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.”
(Tom Mitchell, 1997)



What Can You Do with ML?



What Can You Do with ML?

Google

KI

News | Bilder | Videos | Bücher | Chatbot | Roboter | Singularität | Bedeutung | Intelligenz

Alle Filter | Suchfilter

Ungefähr 11.180.000.000 Ergebnisse (0,43 Sekunden)

Gesponsert

ai-aktientipps.de
https://www.ai-aktientipps.de

Mit KI Aktienchancen erkennen - Clever Investieren mit KI

Gewinnoptimierung an der Börse durch den Einsatz von Artificial Intelligence. So geht´s. Künstliche Intelligenz als Schlüssel zum Erfolg. Börsengewinne durch Depot-Optimierung. AI-basierte Aktientipps. Künstliche Intelligenz. Hier investieren Computer.

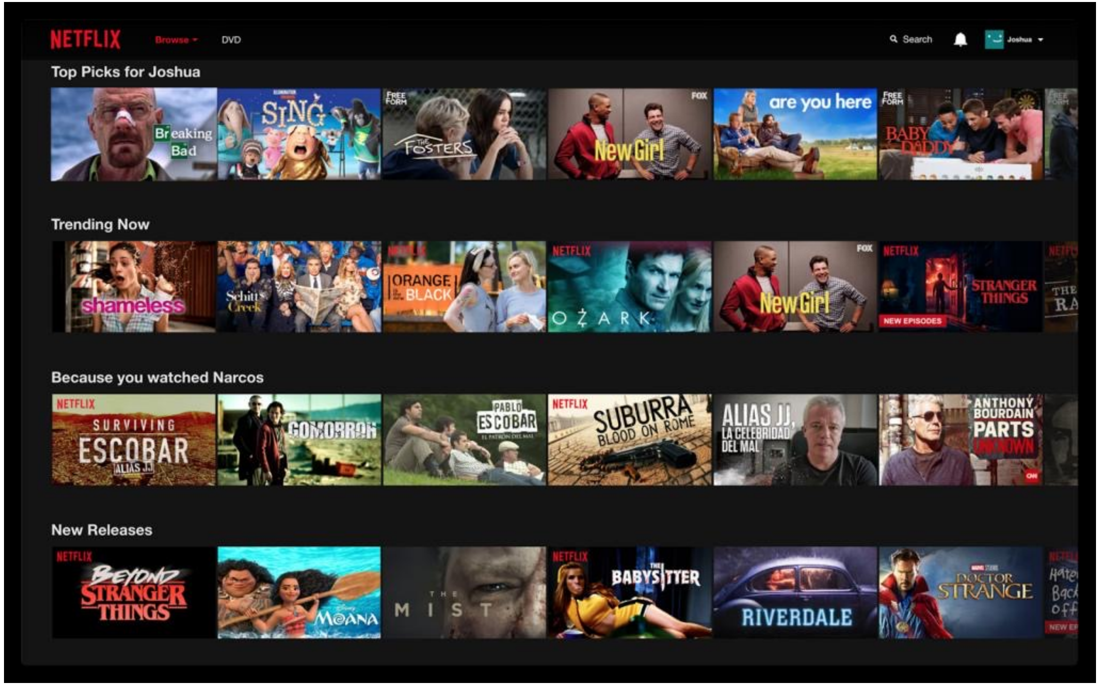
Künstliche Intelligenz (KI) ist ein Teilgebiet der Informatik. Sie imitiert menschliche kognitive Fähigkeiten, indem sie Informationen aus Eingabedaten erkennt und sortiert. Diese Intelligenz kann auf programmierten Abläufen basieren oder durch maschinelles Lernen erzeugt werden.

Fraunhofer IKS
https://www.iks.fraunhofer.de › kuenstliche-intelligenz

Künstliche Intelligenz (KI) und maschinelles Lernen

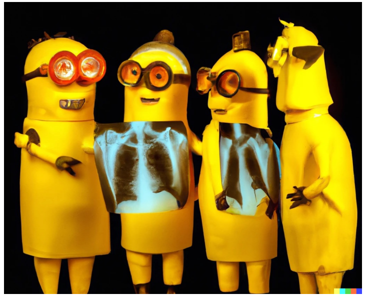
What Can You Do with ML?

Hasso Plattner Institut

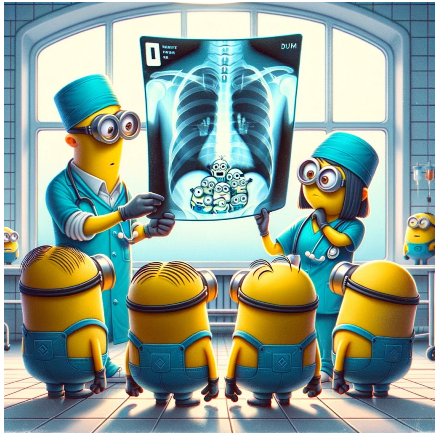


What Can You Do with ML?

DALL·E 2



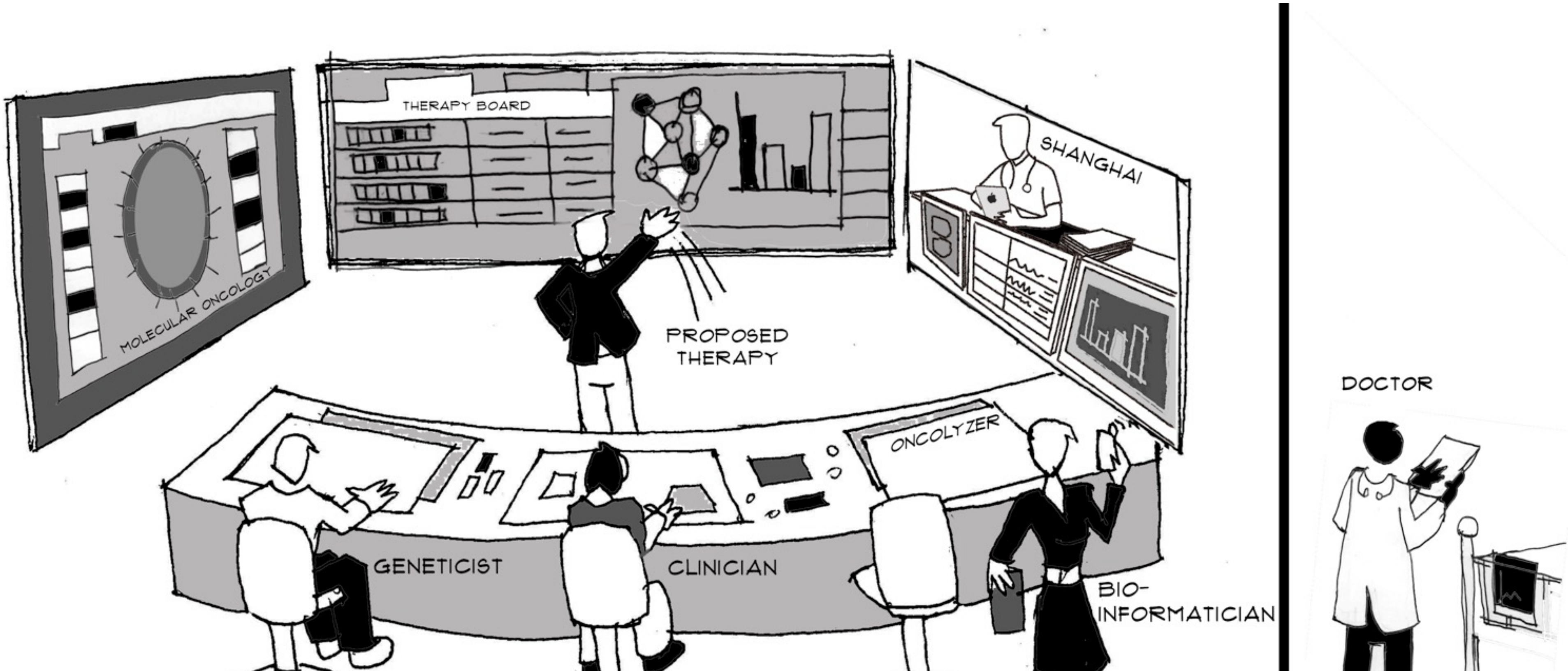
DALL·E 3



"Photo of a Group of Radiologists made from minions discussing X-Ray image made from minions, in the style of minions, trending on artstation, in the style of Pixar”

Our Vision
Real-time Access to Latest International Medical Knowledge

Hasso Plattner Institut



What Can You Do with ML?

"Group of minions discussing in a molecular tumor board. The image should include clinical data (e.g., lab values), images of organs and genetic information. One minion is joining remotely from Shanghai."

What Can You Do with ML?

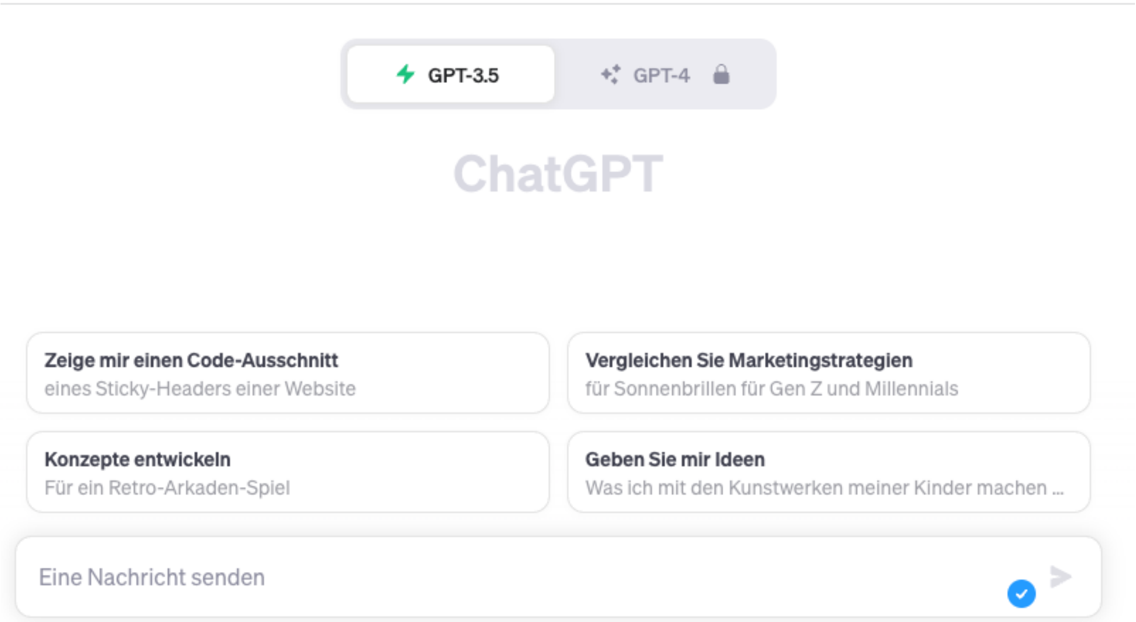
THE SHIFT

An A.I.-Generated Picture Won an Art Prize. Artists Aren't Happy.

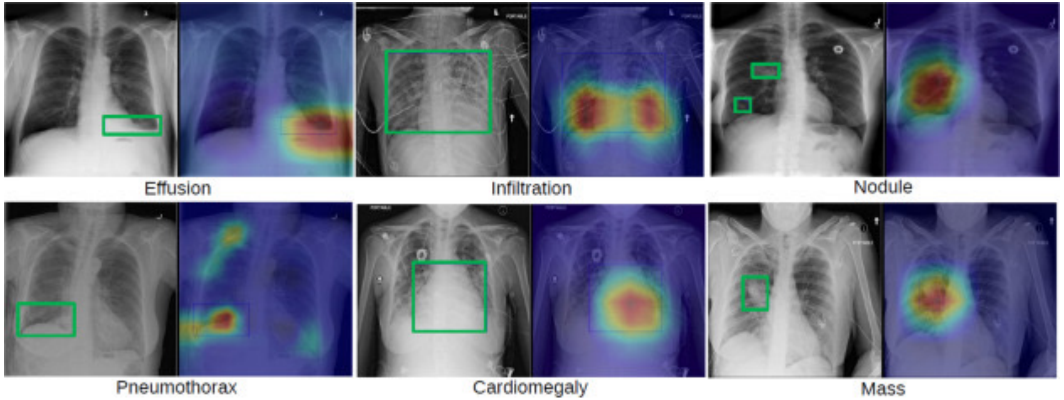
"I won, and I didn't break any rules," the artwork's creator says.

https://www.nytimes.com/2022/09/02/technology/ai-artificial-intelligence-artists.html

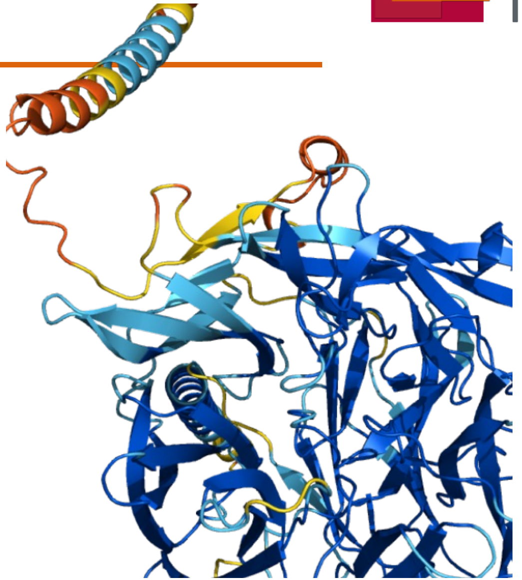
What Can You Do with ML?



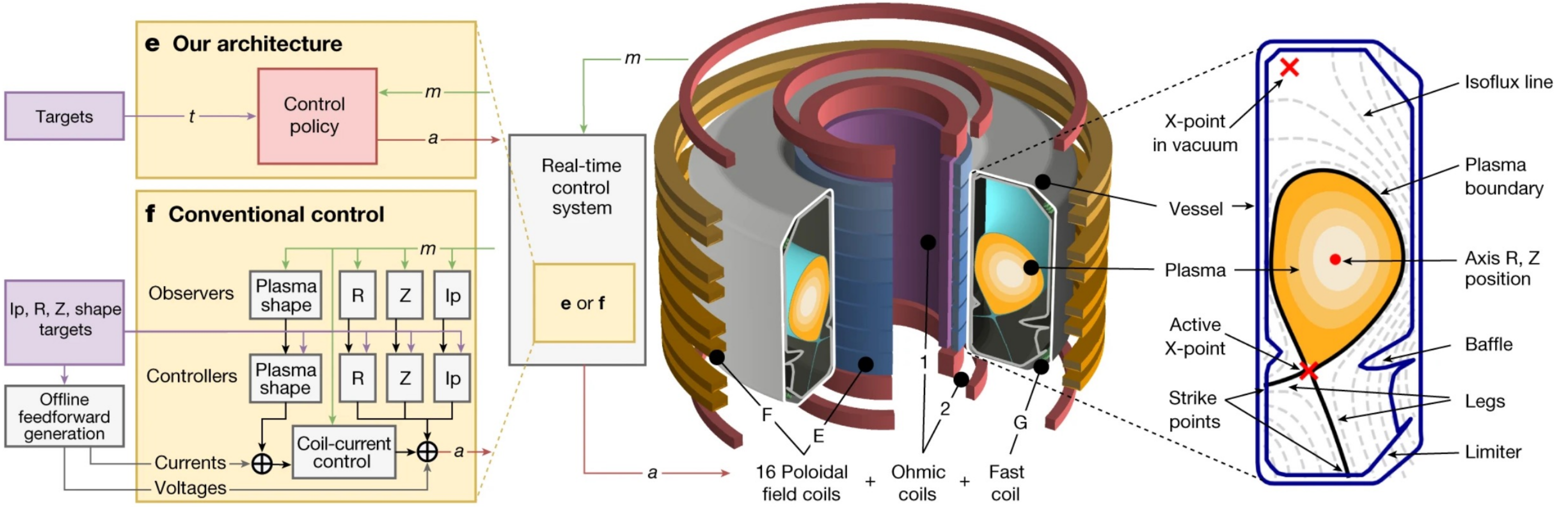
What Can You Do with ML?



https://www.sciencedirect.com/science/article/pii/B9780128212592000144



What Can You Do with ML?



https://www.nature.com/articles/s41586-021-04301-9

What is ML?
Basic Terminology

- Training data:
  - Inputs: data instances $\boldsymbol{x}^{(1)}, \boldsymbol{x}^{(2)}, \dots, \boldsymbol{x}^{(n)} \in \boldsymbol{X}$
    - e.g., X is the set of fruits, $\boldsymbol{x}^{(1)} =$  , $\boldsymbol{x}^{(2)} =$ 
  - (optionally) Labels: Annotations we want to predict $y^{(1)}, y^{(2)} \dots, y^{(n)} \in Y$
    - e.g., $\boldsymbol{Y} = \{\text{"apple"}, \text{"orange"}\}$ and $y^{(1)} = \text{"apple"}, y^{(2)} = \text{"orange"}$
- **Goal in ML**: Find function (model of the data) $\boldsymbol{f}: \boldsymbol{X} \rightarrow \boldsymbol{Y}$, that performs well on unseen instances (test data) 
- Training: choose optimal $\boldsymbol{f}^*$ out of a class of functions in an optimization procedure, such that objective function $\boldsymbol{J}$ is maximized / error is minimized
- Data instances are represented by features, e.g.:
  - attributes of fruits (color, shape) ($\boldsymbol{X} = \mathbb{R}^d$)
  - raw pixels of images of fruits ($\boldsymbol{X} = \mathbb{R}^{w \times h}$)

https://de.wikipedia.org/wiki/Datei:Honeycrisp-Apple.jpg
https://blog.paperspace.com/intro-to-optimization-in-deep-learning-gradient-descent/
http://images6.fanpop.com/image/photos/35300000/Food-Oranges-food-35302708-500-500.png

Most common problem settings in Machine Learning

Supervised Learning (Labels available for training)

Classification

Categorical output

e.g. $x \in$ Fruits, $y \in$ {"apple", "orange"}

f()= "apple"

f()= "orange"

Regression

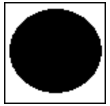
Continuous output

e.g.: $x \in$ Fruits, $y \in \mathbb{R}_+ \triangleq$ t until ripe)

f () = 12 days

Structured Prediction

e.g. $x \in \mathbb{R}^{w \times h \times d}$, $y \in \mathbb{R}^{w \times h} \triangleq$ pixels

f() = 

Most common problem settings in Machine Learning

Supervised Learning (Labels available for training)

Classification

Categorical output

e.g. $x \in$ Fruits, $y \in$ {“apple”, “orange”}

f()= “apple”

f()= “orange”

Regression

Continuous output

e.g.: $x \in$ Fruits, $y \in \mathbb{R}_{+} \triangleq$ t until ripe)

f () = 12 days

Structured Prediction

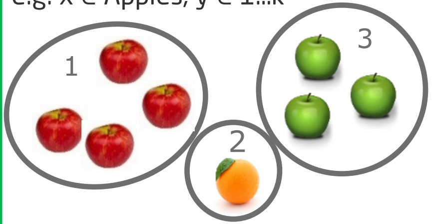
e.g. $x \in \mathbb{R}^{w \times h \times d}$, $y \in \mathbb{R}^{w \times h} \triangleq$ pixels

f() = 

Unsupervised Learning (No labels during training)

Clustering

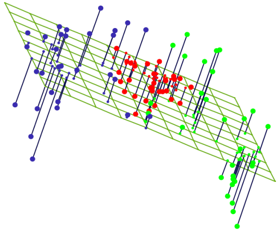
e.g. $x \in$ Apples, $y \in 1 \dots k$



Dimensionality reduction

$x \in \mathbb{R}^d$, $x' \in \mathbb{R}^p$, $p < d$

e.g., projecting all features of a fruit to 2 dimensions for visualization



Most common problem settings in Machine Learning

Supervised Learning (Labels available for training)

Classification

Categorical output

e.g. x ∈ Fruits, y ∈ {"apple", "orange"}

f()= "apple"

f()= "orange"

Regression

Continuous output

e.g.: x ∈ Fruits, y ∈ $\mathbb{R}_+$ ≙ t until ripe)

f () = 12 days

Structured Prediction

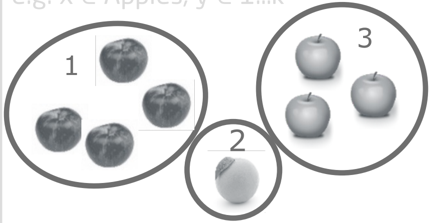
e.g. x ∈ $\mathbb{R}^{w \times h \times d}$, y ∈ $\mathbb{R}^{w \times h}$ ≙ pixels

f() = 

Unsupervised Learning (No labels during training)

Clustering

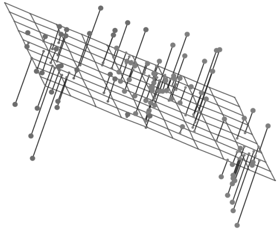
e.g. x ∈ Apples, y ∈ 1…k



Dimensionality reduction

x ∈ $\mathbb{R}^d$, x' ∈ $\mathbb{R}^p$, p < d

e.g., projecting all features of a fruit to 2 dimensions for visualization

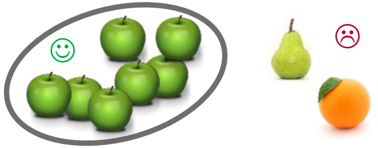


Semi-Supervised Learning (Some labels for training)

Anomaly / novelty detection

trained only on "normal" samples

e.g. x ∈ Apples, y ∈ { ☺ , ☹ }



Most common problem settings in Machine Learning

Supervised Learning (Labels available for training)

Classification

Categorical output

e.g. $x \in$ Fruits, $y \in$ {"apple", "orange"}

f( )= "apple"

f( )= "orange"

Regression

Continuous output

e.g.: $x \in$ Fruits, $y \in \mathbb{R}_+ \triangleq$ t until ripe)

f ( ) = 12 days

Structured Prediction

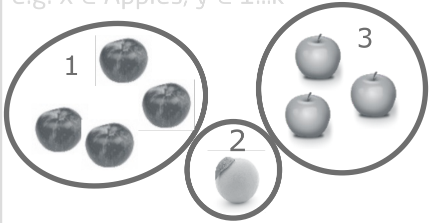
e.g. $x \in \mathbb{R}^{w \times h \times d}$, $y \in \mathbb{R}^{w \times h} \triangleq$ pixels

f() =

Unsupervised Learning (No labels during training)

Clustering

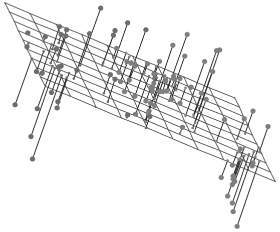
e.g. $x \in$ Apples, $y \in 1 \ldots k$



Dimensionality reduction

$x \in \mathbb{R}^d$, $x' \in \mathbb{R}^p$, $p < d$

e.g., projecting all features of a fruit to 2 dimensions for visualization

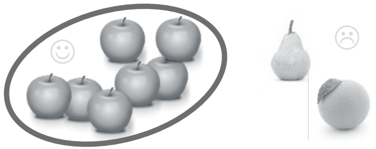


Semi-Supervised Learning (Some labels for training)

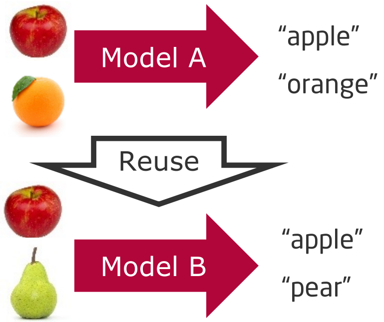
Anomaly / novelty detection

trained only on "normal" samples

e.g. $x \in$ Apples, $y \in$ { ☺ , ☹ }



Transfer Learning



Introduction to ML

Most common problem settings in Machine Learning

Hasso Plattner Institut

Supervised Learning (Labels available for training)

Classification

Categorical output

e.g. x ∈ Fruits, y ∈ {"apple", "orange"}

f()= "apple"

f()= "orange"

Regression

Continuous output

e.g.: x ∈ Fruits, y ∈ $\mathbb{R}_+$ ≙ t until ripe)

f () = 12 days

Structured Prediction

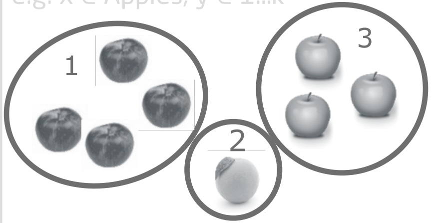
e.g. $x \in \mathbb{R}^{w \times h \times d}$, $y \in \mathbb{R}^{w \times h}$ ≙ pixels

f() = 

Unsupervised Learning (No labels during training)

Clustering

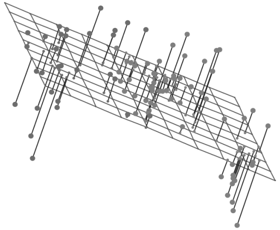
e.g. x ∈ Apples, y ∈ 1…k



Dimensionality reduction

$x \in \mathbb{R}^d$, $x' \in \mathbb{R}^p$, p < d

e.g., projecting all features of a fruit to 2 dimensions for visualization

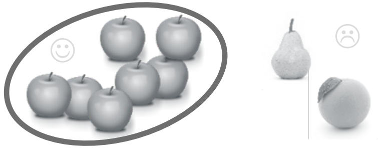


Semi-Supervised Learning (Some labels for training)

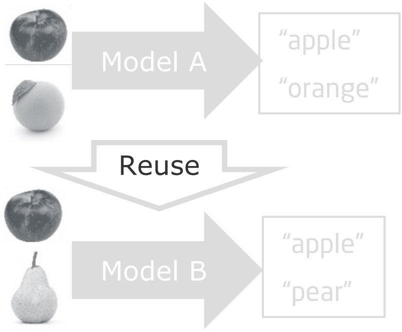
Anomaly / novelty detection

trained only on "normal" samples

e.g. x ∈ Apples, y ∈ { ☺ , ☹ }



Transfer Learning



Reinforcement Learning

https://en.wikipedia.org/wiki/Apple
https://cdn4.vectorstock.com/i/1000x1000/16/58/robot-arm-line-icon-sign-on-vector-17841658.jpg

Most common problem settings in Machine Learning

Hasso Plattner Institut

Supervised Learning (Labels available for training)

Classification

Categorical output

e.g. x ∈ Fruits, y ∈ {“apple”, “orange”}

f()= “apple”

f()= “orange”

Regression

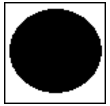
Continuous output

e.g.: x ∈ Fruits, y ∈ $\mathbb{R}_+$ ≙ t until ripe)

f () = 12 days

Structured Prediction

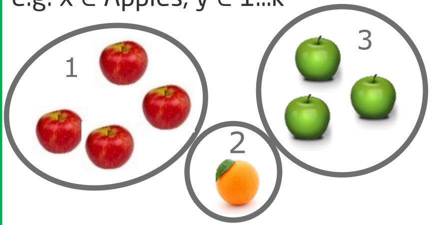
e.g. x ∈ $\mathbb{R}^{w \times h \times d}$, y ∈ $\mathbb{R}^{w \times h}$ ≙ pixels

f() = 

Unsupervised Learning (No labels during training)

Clustering

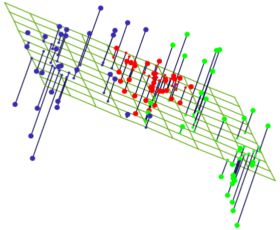
e.g. x ∈ Apples, y ∈ 1…k



Dimensionality reduction

x ∈ $\mathbb{R}^d$, x’ ∈ $\mathbb{R}^p$, p < d

e.g., projecting all features of a fruit to 2 dimensions for visualization

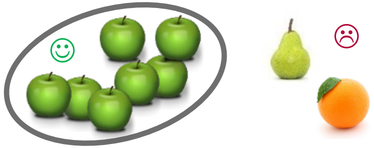


Semi-Supervised Learning (Some labels for training)

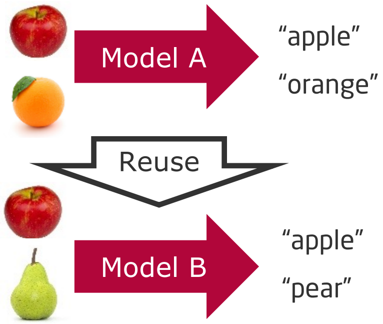
Anomaly / novelty detection

trained only on “normal” samples

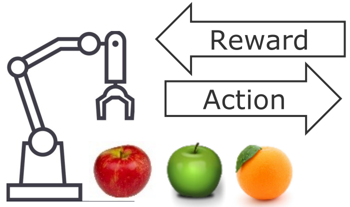
e.g. x ∈ Apples, y ∈ { ☺ , ☹ }



Transfer Learning



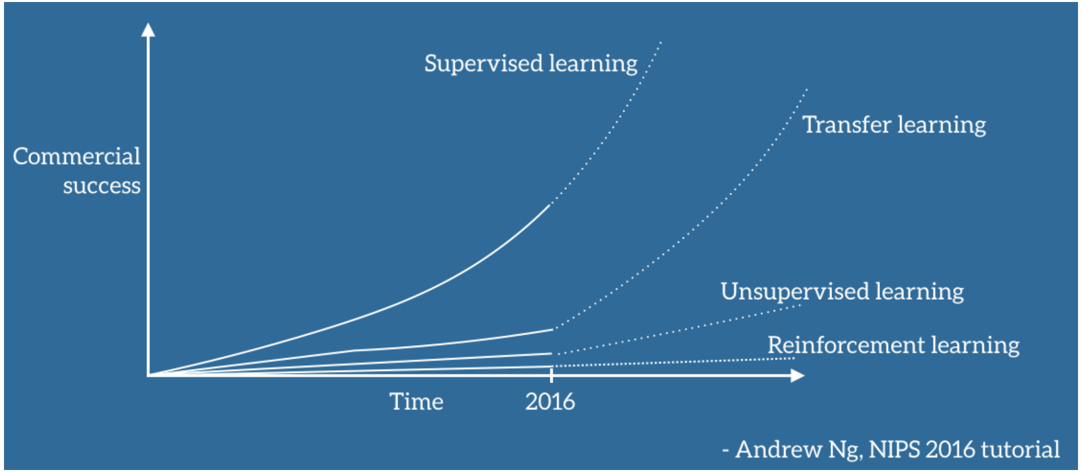
Reinforcement Learning



Introduction to ML

Drivers of ML Success in Industry

- Most of the recent successes of ML are based on:
  - Large, labelled datasets for Supervised Learning
  - Computational power (GPU acceleration) to carry out training
  - Pre-trained models to reduce necessary data in new domains via transfer learning

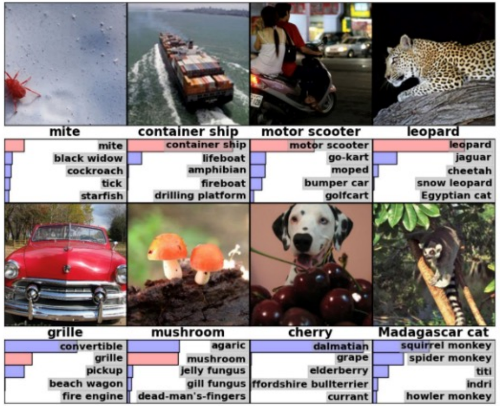


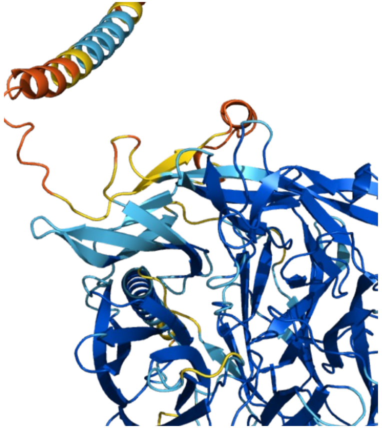
http://ruder.io/transfer-learning/

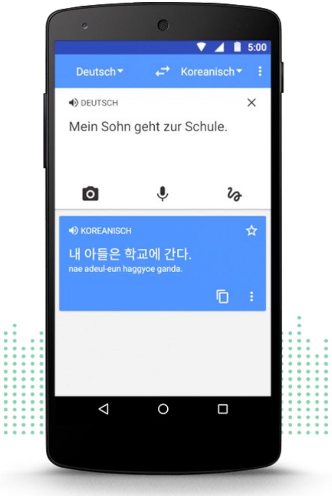
Supervised Learning

IMAGENET

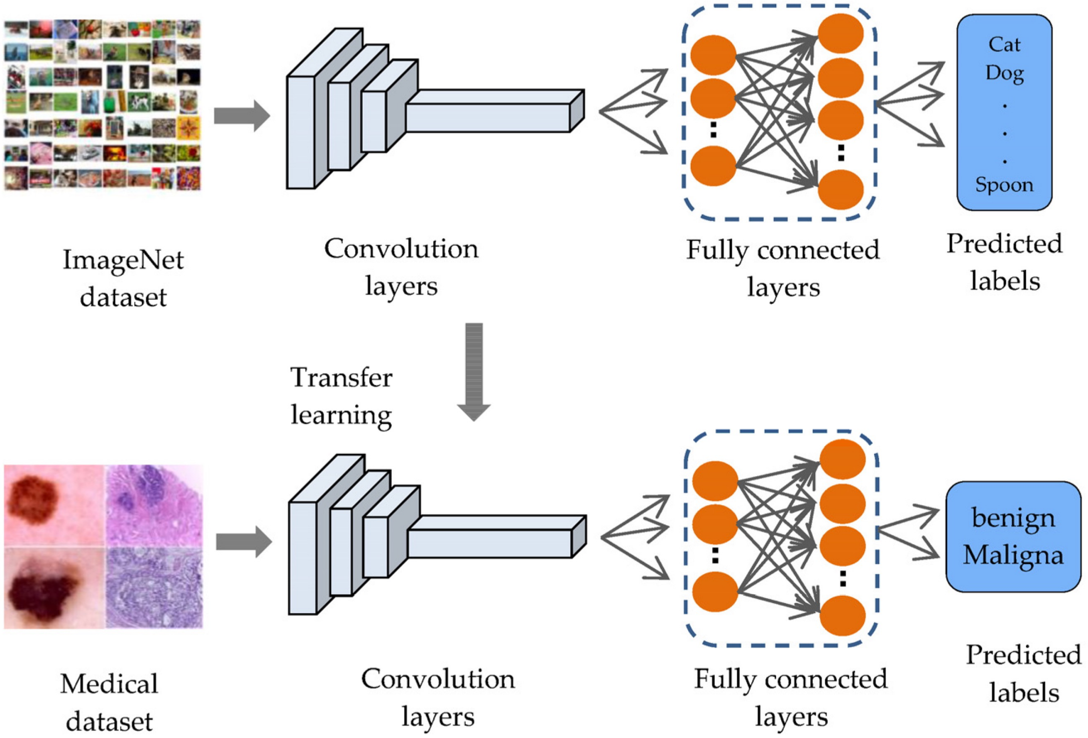
- 1,000 object classes (categories).
- Images:
    - 1.2 M train
    - 100k test.





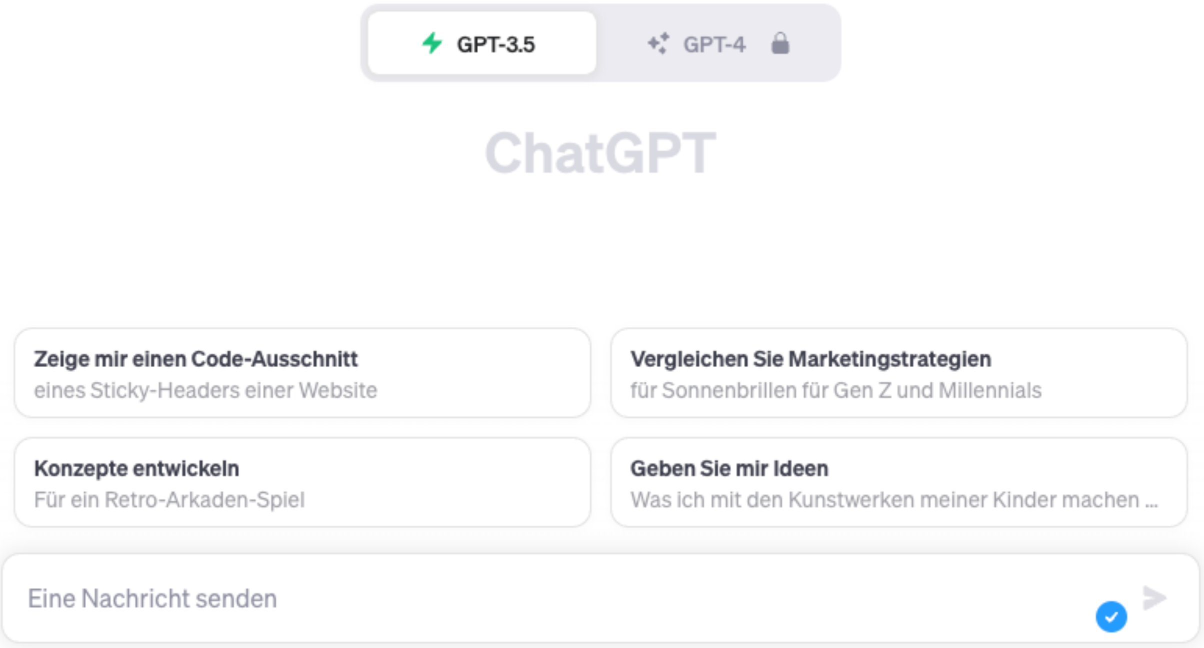


Transfer Learning



https://www.mdpi.com/1424-8220/23/2/570

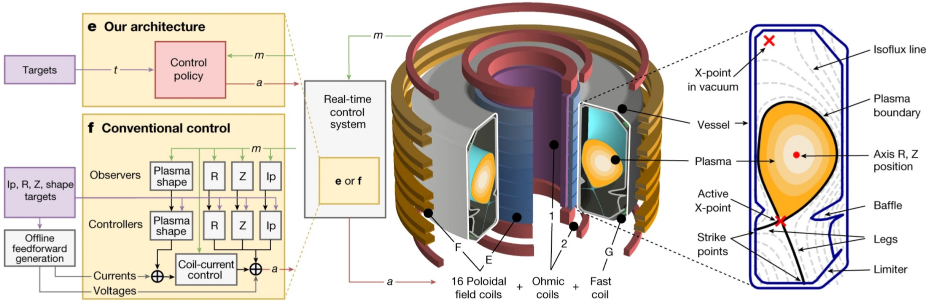
Unsupervised Learning
(+ a few clever tricks)



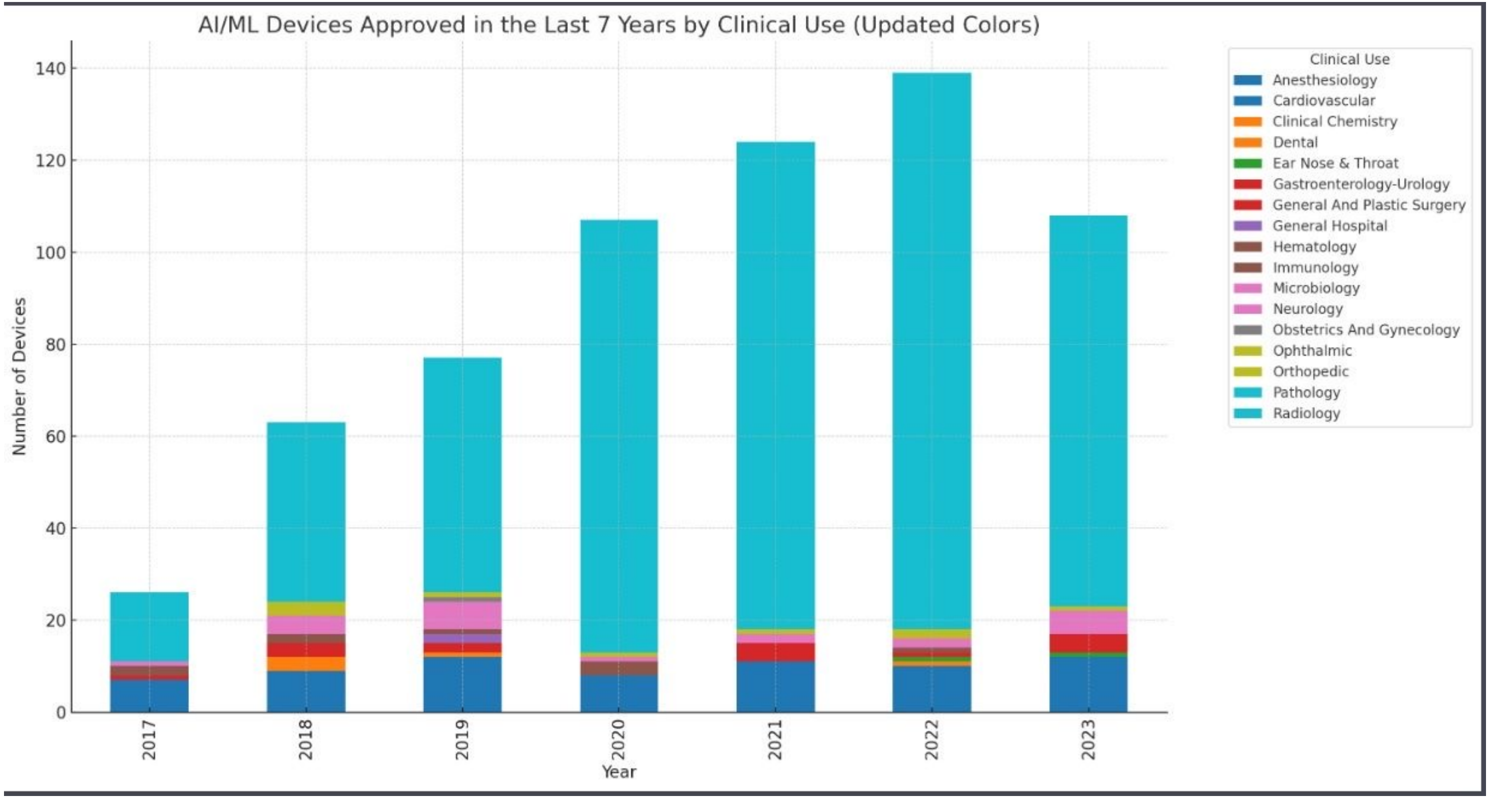
Reinforcement Learning

AlphaGO
1202 CPUs, 176 GPUs,
100+ Scientists.

Lee Se-dol
1 Human Brain,
1 Coffee.



AI/ML-Enabled Medical Devices (FDA Approved)



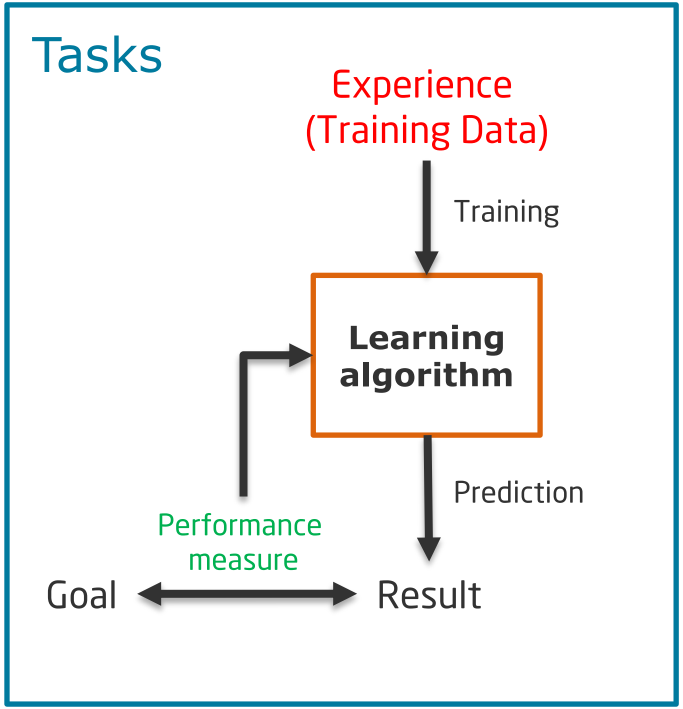
https://www.fda.gov/medical-devices/software-medical-device-samd/artificial-intelligence-and-machine-learning-aiml-enabled-medical-devices

A Very Simple First Example

Machine Learning

“Field of study that gives computers the ability to learn without being explicitly programmed”
(Arthur Samuel, 1959)

“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.”
(Tom Mitchell, 1997)



Running Example: Binary Classification

	x_1	x_2	y
	Age	Systolic Blood Pressure	Coronary Heart Disease
$x^{(1)}$	17	118	0
$x^{(2)}$	46	117	0
$x^{(3)}$	53	146	1
$x^{(4)}$	62	158	1
$x^{(5)}$	20	106	1
.	...	...	...
.	20	124	0
.	48	144	1
.	42	154	0
	51	124	1
$x^{(n)}$	58	214	0

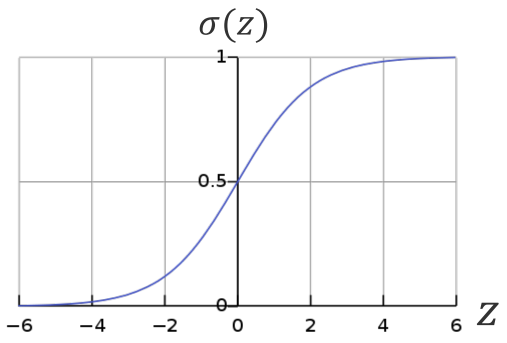
Predict risk of coronary heart disease!

Hasso Plattner Institut

Simple Binary Classification Algorithm: Logistic Regression

- Linear Model:
  - Features $\boldsymbol{x}^{(i)} = (x_1^{(i)}, x_2^{(i)}, \dots, x_d^{(i)})$
  - Weights $\boldsymbol{w} = (w_1, w_2, \dots, w_n)$, bias b
  - Model $P\left(y^{(i)} = 1 \,\middle|\, \boldsymbol{x}^{(i)}\right) = \sigma(\boldsymbol{w}^T \boldsymbol{x}^{(i)} + b)$
  - Logistic function $\sigma(z) = \frac{1}{1+e^{-z}}$
  - Classification function: $f\left(\boldsymbol{x}^{(i)}\right) = \begin{cases} 1, & P(=1\,|\boldsymbol{x}^{(i)}) > t \\ 0, & otherwise \end{cases}$

    where $0 \leq t \leq 1$ is a threshold
- Goal: find values for $\boldsymbol{w}^*$and b^*s.t. prediction $f(\boldsymbol{x})$ is likely to be close to true label $y^{(i)}$ for many data points $\boldsymbol{x}^{(i)}$
- Training: find parameters that minimize loss on training set $(\boldsymbol{x}^{(1)}, \boldsymbol{x}^{(2)}, \dots, \boldsymbol{x}^{(n)})$ with labels $(y^{(1)}, y^{(2)}, \dots, y^{(n)})$



Simple Binary Classification: Logistic Regression

	x_1	x_2	y
	Age	Systolic Blood Pressure	Coronary Heart Disease
$x^{(1)}$	17	118	0
$x^{(2)}$	46	117	0
$x^{(3)}$	53	146	1
$x^{(4)}$	62	158	1
$x^{(5)}$	20	106	1
	...	...	...
.	20	124	0
.	48	144	1
.	42	154	0
.	51	124	1
$x^{(n)}$	58	214	0

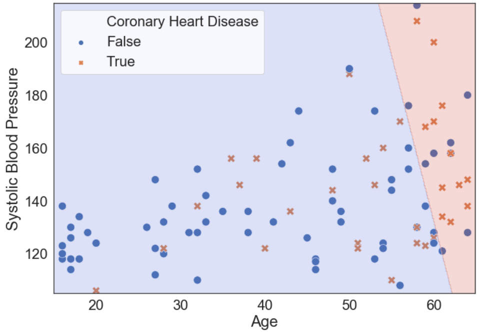
Model $P\left(y^{(i)} = 1 \,\middle|\, \boldsymbol{x}^{(i)}\right) = \sigma(\boldsymbol{w}^T \boldsymbol{x}^{(i)} + b)$

$$= \sigma(w_1 \; x_1^{(i)} + w_2 \; x_2^{(i)} + b)$$

Features $\boldsymbol{x}_i = (x_i^{(1)}, x_i^{(2)})$

Weights $\boldsymbol{w} = (w_1 \;, w_2 \;)$, bias b

Logistic function $\sigma(z) = \frac{1}{1+e^{-z}}$



→ Result of running logistic regression:

$w_1 \;\; = 0.057 \quad w_2 \;\; = 0.0044 \quad b = -3.857$

$P(y = 1 \mid \boldsymbol{x}^{(1)}) \; = \sigma(0.057 \cdot 17 + 0.0044 \cdot 118 \; - 3.857) = \mathbf{0.08}$

$P(y = 1 \mid \boldsymbol{x}^{(4)}) \; = \sigma(0.057 \cdot 62 + 0.0044 \cdot 158 \; - 3.857) = \mathbf{0.56}$

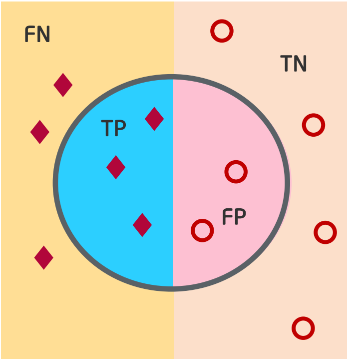
Performance Measures for Binary Classification

In binary classification, each instance is classified either correctly or incorrectly

Predicted

Actual

	Positive	Negative
Positive	True Positive	False Negative
Negative	False Positive	True Negative



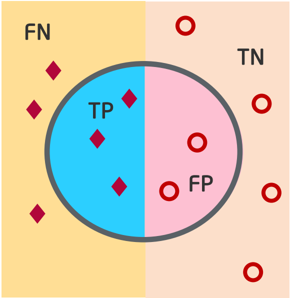
… Also works for multi-class classification

Metrics

- Accuracy := $\dfrac{TP + TN}{FP + FN + TP + TN}$
- Precision := $\dfrac{TP}{TP + FP}$
- Recall := $\dfrac{TP}{TP + FN}$
- F-measure := $2 \cdot \dfrac{Precision \cdot Recall}{Precision + Recall}$
- Sensitivity := True Positive Rate := Recall
- Specificity := True Negative Rate

Predicted

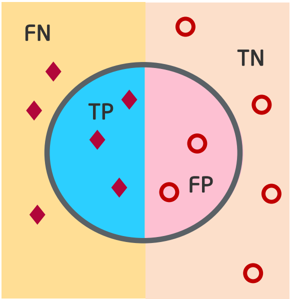
Actual	Positive	Negative
Positive	True Positive	False Negative
Negative	False Positive	True Negative



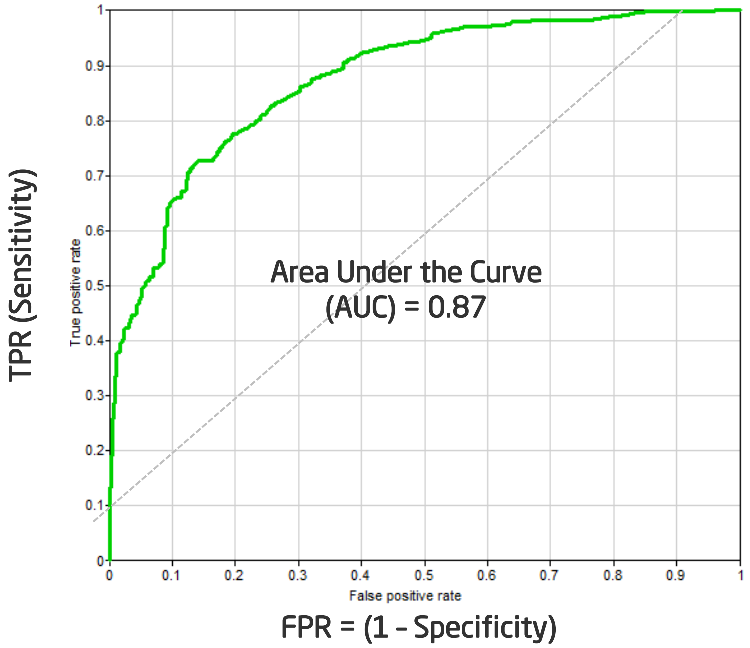
Measures of Performance: Confusion Matrix

- Precision := $\frac{3}{3+2} = \frac{3}{5} = 0.6$
- Recall := $\frac{3}{3+3} = \frac{3}{6} = 0.5$
- F-measure := $2 \cdot \frac{0.5 \cdot 0.6}{0.5+0.6} = 0.55$

Actual \ Predicted	Positive	Negative
Positive	3	3
Negative	2	4



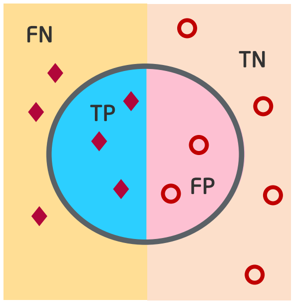
Evaluation
Receiver Operating Characteristic (ROC) Curve



TPR (sensitivity) = TP / (TP + FN)

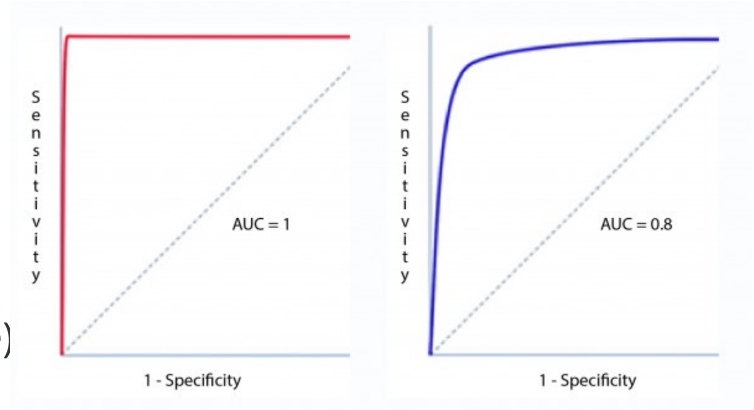
TNR (specificity) = TN / (TN + FP)

FPR (1 - specificity) = FP / (TN + FP)



Evaluation
Receiver Operating Characteristic (ROC) Curve

- Performance of a binary classifier
- Plot showing TPR and FPR
- Varying *classification thresholds*
- Allows comparison between classifiers
- The higher the AUC (Area Under the Curve) the better

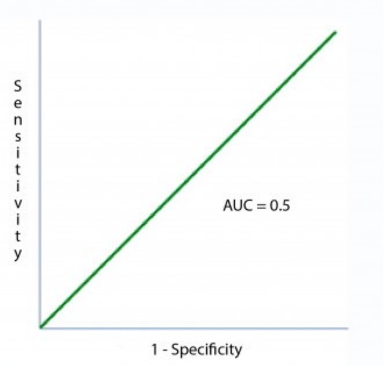


$$TPR = \underbrace{\frac{TP}{TP + FN}}_{\text{all positive instances}}$$

all positive instances

$$FPR = \underbrace{\frac{FP}{FP + TN}}_{\text{all negative instances}}$$

all negative instances



http://www.cienciasinseso.com/en/tag/roc-curve/

Try it at Home

- https://github.com/hpi-dhc/dm4dh-2023

README.md

Data Management for Digital Health 2023/24

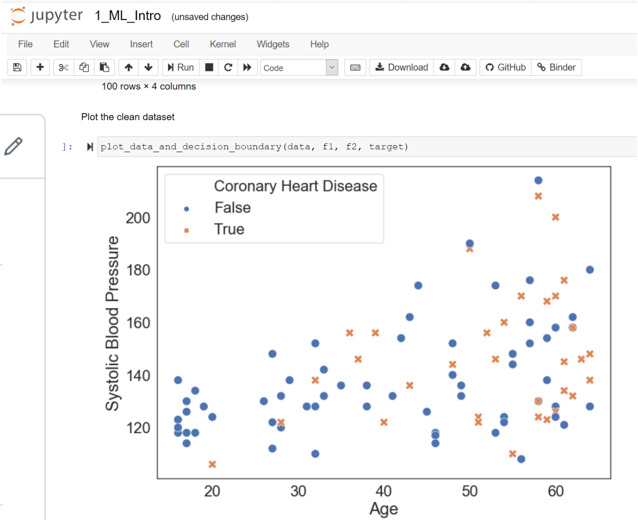
This repository contains the code to reproduce figures, metrics, and models for the 2023/24 version of the course.

To run all notebooks interactively with MyBinder, click here (and wait for a few seconds):

launch binder

Contents:

- Week 2: Introduction to Machine Learning



- View the Python code to reproduce many of the graphs / tables in this lecture
- Browse the data and experiment yourself

What to take home?

- ML Problem Settings
- ML Applications (in Digital Health)
- Binary Classification Algorithm + Metrics