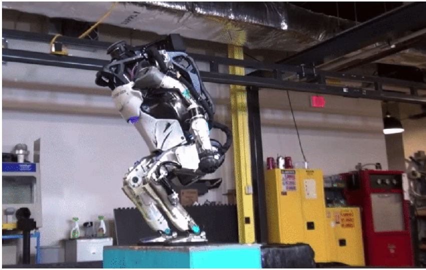
Chapter 01

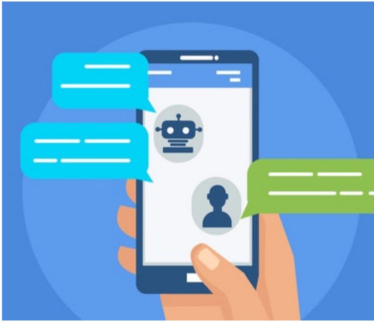
Machine Learning Basics

Prof. Dr. Adrian Ulges

RheinMain University of Applied Sciences

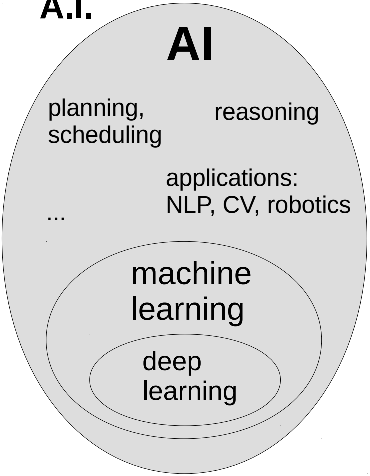
Machine Learning: Recent Successes images: [13] [14] [4] [12] [15]



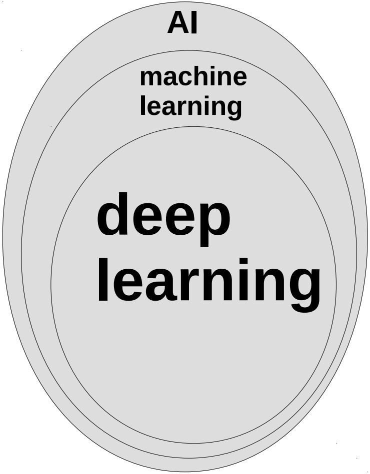


AI vs. Machine Learning vs. Deep Learning

Machine Learning is a sub-area of A.I.



public perception these days



Example Applications

Most common ML Applications

- object recognition, OCR+handwriting recognition
- search engines, recommender systems
- natural language processing, document categorization.

But there's a lot more! Check out kaggle.com...

- Flight Quest
  optimize flight routes based on wheather and traffic
- TFI Restaurant Revenue Prediction
  predict annual sales of restaurants to open
- Job Recommendation Challenge
  predict which jobs users will apply to
- Whale Detection Challenge
  detect whale calls from audio, prevent collision with ship traffic
- Discovering trolling in user comments
- ...

An ML Sample Application images: [3] [7]

- A computer system is to make a non-trivial decision.
- example: spam filtering
- Why not hard-code the decision logic?

Problems

- high effort to grasp problem's complexity.
- easy to code *something*, difficult to reach the optimal logic.
- feasibility checking: What accuracy can be reached by a decision?
- code is extremely difficult to maintain.
- keeping track of data changes (e.g., when spammers change strategies) is almost impossible.
- there is no way to take user feedback into account.

Outline

1. Basic Terminology

2. Benchmarking

Machine Learning: Definition

> *"Machine learning is a scientific discipline that explores the construction and study of* **algorithms that can learn from data**. *Such algorithms operate by building a* **model** *from* **example inputs** *and using that to make* **predictions or decisions**, *rather than following strictly static program instructions."*
>
> (en.wikipedia.org)

> *"The field of study that gives computers the ability to learn* **without being explicitly programmed.**"
>
> (Arthur Samuel (1959))

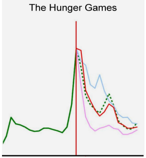
> *"A computer program is said to* **learn** *from experience E with respect to some task T and some performance measure P, if its performance on T, as measured by P, improves with Experience E."*
>
> (Tom Mitchell (1998))

Remark

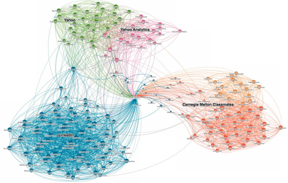
These definitions are entirely **non-operational**!

Machine Learning: Tasks images: [8] [5] [2] [1]

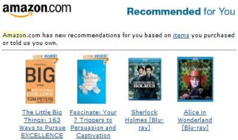
Regression



Clustering / Segmentation



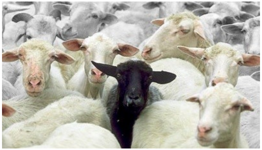
Recommendation



Classification

Data Reduction

Anomaly Detection



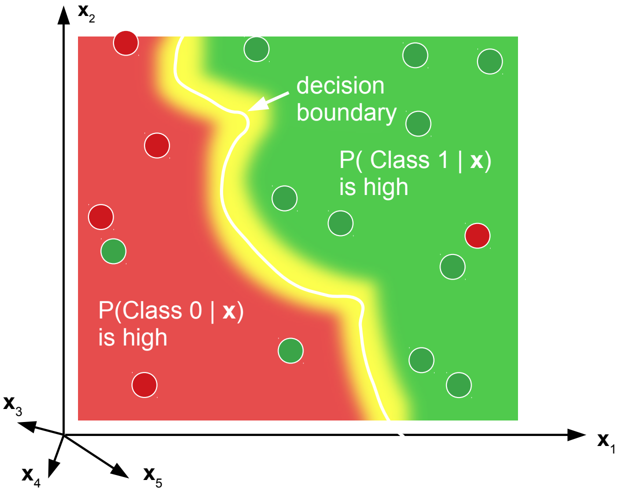
Machine Learning: Features+Labels image: [9]

- Machine learning makes predictions about real-world objects.
- We describe an object by a **feature vector** $\mathbf{x}$ *(bold=vectors!)*.
- Our goal is to predict a **label** y for the object.

	A	B	C	D	E	F	G	H	I	J	K	L
1	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked
2	1	0	3	Braund, Mr. Owen Harris	male	22	1	0	A/5 21171	7.25		S
3	2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Thayer)	female	38	1	0	PC 17599	71.2833	C85	C
4	3	1	3	Heikkinen, Miss. Laina	female	26	0	0	STON/O2. 3101282	7925		S
5	4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35	1	0	113803	53.1	C123	S
6	5	0	3	Allen, Mr. William Henry	male	35	0	0	373450	8.05		S
7	6	0	3	Moran, Mr. James	male		0	0	330877	8.4583		Q
8	7	0	1	McCarthy, Mr. Timothy J	male	54	0	0	17463	51.8625	E46	S
9	8	0	3	Palsson, Master. Gosta Leonard	male	2	3	1	349909	21075		S
10	9	1	3	Johnson, Mrs. Oscar W (Elisabeth Vilhelmina Berg)	female	27	0	2	347742	11.1333		S
11	10	1	2	Nasser, Mrs. Nicholas (Adele Achem)								
12	11	1	3	Sandstrom, Miss. Marguerite Rut								
13	12	1	1	Bonnell, Miss. Elizabeth								
14	13	0	3	Saundercock, Mr. William Henry								
15	14	0	3	Andersson, Mr. Anders Johan								
16	15	0	3	Vestrom, Miss. Hulda Amanda Adolfina								
17	16	1	2	Hewlett, Mrs. (Mary D Kingcome)								
18	17	0	3	Rice, Master. Eugene								
19	18	1	2	Williams, Mr. Charles Eugene								
20	19	0	3	Vander Planke, Mrs. Julius (Emelia Maria Van								
21	20	1	3	Masselmani, Mrs. Fatima								
22	21	0	2	Fynney, Mr. Joseph J								
23	22	1	2	Beesley, Mr. Lawrence								
24	23	1	3	McGowan, Miss. Anna "Annie"								
25	24	1	1	Sloper, Mr. William Thompson								
26	25	0	3	Palsson, Miss. Torborg Danira								
27	26	1	3	Asplund, Mrs. Carl Oscar (Selma Augusta Emi								
28	27	0	3	Emir, Mr. Farred Chehab								
29	28	0	1	Fortune, Mr. Charles Alexander								
30	29	1	3	O'Dwyer, Miss. Ellen "Nellie"								
31	30	0	3	Todoroff, Mr. Lalio								
32	31	0	1	Uruchurtu, Don. Manuel E								
33	32	1	1	Spencer, Mrs. William Augustus (Marie Eugen								
34	33	1	3	Glynn, Miss. Mary Agatha								
35	34	0	2	Wheadon, Mr. Edward H								
36	35	0	1	Meyer, Mr. Edgar Joseph								
37	36	0	1	Holverson, Mr. Alexander Oskar								
38	37	1	3	Mamee, Mr. Hanna								
39	38	0	3	Cann, Mr. Ernest Charles								
40	39	0	3	Vander Planke, Miss. Augusta Maria								
41	40	1	3	Nicola-Yarred, Miss. Jamila								
42	41	0	3	Ahlin, Mrs. Johan (Johanna Persdotter Larsso								
43	42	0	2	Turpin, Mrs. William John Robert (Dorothy Ann								
44	43	0	3	Kraeff, Mr. Theodor								

Machine Learning: Geometric View

- Feature vectors x are points in feature space.
- The ML model estimates decision boundaries between classes.



ML: Feature Engineering / Feature Extraction

Often, we preprocess features before applying machine learning.

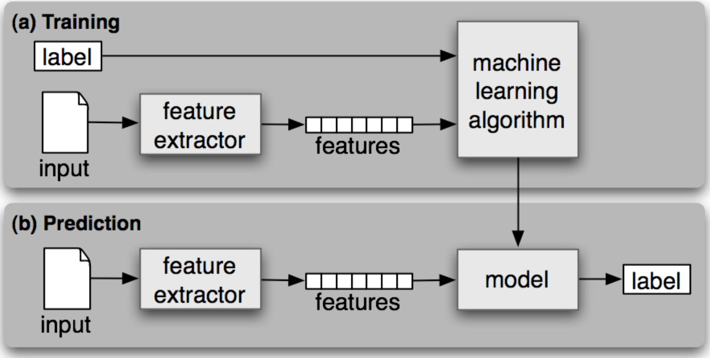
1. Features may be missing, i.e. x is *incomplete*.
   Approach: estimate missing values (*imputation*)
2. Categorical features may have to be transformed into numerical ones, typically by introducing dummy variables *(one-hot encoding)*.

one-hot encoding

	PS	color	PS	is_green	is_silver	is_red
Prof. Ulges' car	70	white	73	0	0	0
Prof. Ulges' wives' car	690	red	690	0	0	1

3. We may want to discard outliers.
4. We may want to discard uninformative features.
5. Often, we normalize features, for example by standardizing them to mean 0 and standard deviation 1 ($x := (x - \bar{x})/s$).

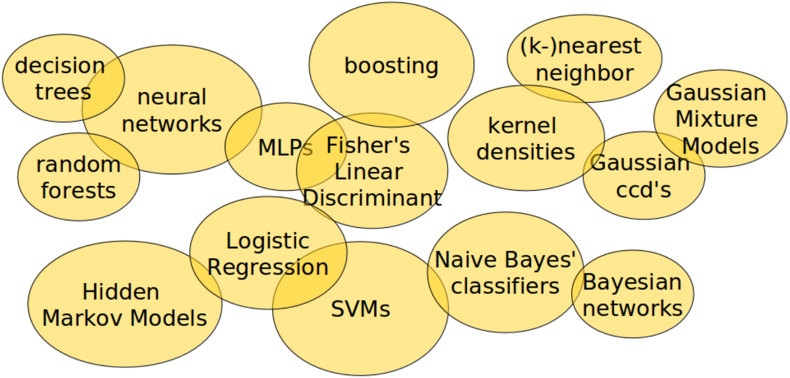
ML: System Pipeline image: [11]



Here: Batch Learning / Offline Learning

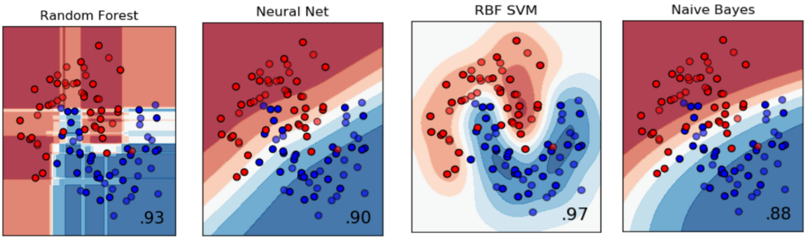
1. We train the system on training data $x_1, x_2, \ldots, x_n$ with labels $y_1, y_2, \ldots, y_n$, obtaining a model ϕ.
2. Given a new object x', the model predicts the label $\phi(x')$.
3. Training happens offline, the application of the model online.

ML: Approaches image: [6]



- There are many different methods in machine learning to solve the classification problem *(and also other problems)*.
- There is no universally best classifier *("no-free-lunch theorem")*.

ML: Sample Approaches



Different Approaches

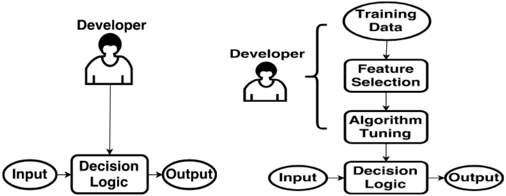
- **Random Forests**: Recursive Splitting of feature space.
- **Neural Networks**: Stacking of linear decision units.
- **SVMs**: Similarity comparison of objects.
- **Probabilistic Models (e.g., Naive Bayes)**: Use probabilities.

Types of Machine Learning

- **classification vs. regression**
  in classification, the labels $y_1, ..., y_n$ are categorical. In regression, they are real-valued.
- **supervised learning**
  learning from samples $x_1, ..., x_n$ with labels $y_1, ..., y_n$.
- **unsupervised learning**
  learning only from samples $x_1, ..., x_n$, no labels.
- **semi-supervised learning**
  learning from samples $x_1, ..., x_n$, some with labels.
- **active learning**
  ... where the system can pick which samples to label.
- **ensemble learning**
  ... is about combining learners for a more robust decision.
- **reinforcement learning**
  ... is about learning from feedback instead of labels.
- **transfer learning**
  ... is about applying models trained on Task X to Task Y.

Outline

ML in Practice

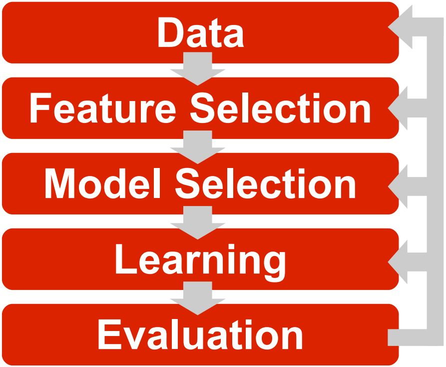


Verification in ML Practice ...

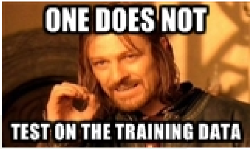
- is done by measuring accuracy on sample data: Benchmarking.

The Machine Learning Development Cycle

- We conduct an iterative search for good features, models, and input data.



Benchmarking: Ground Rule image: [10]

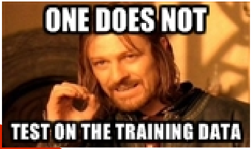


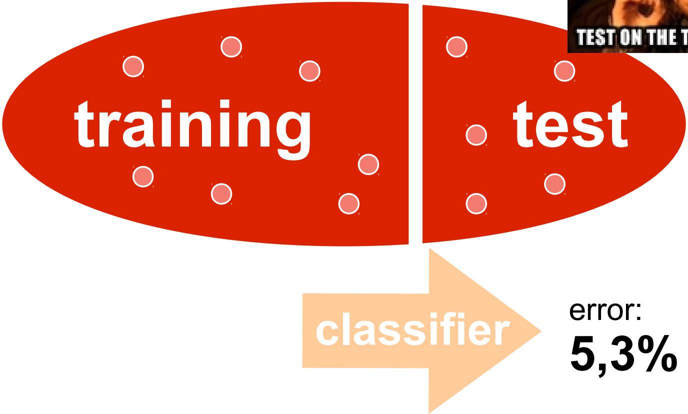
> *"The most common mistake among machine learning beginners is to test on the training data and have the illusion of success."*
>
> (P. Domingos, *A few Useful Things to Know about Machine Learning*)

Why?

- All classifiers are prone to overfitting.
- Achieving perfect accuracy on the training samples is simple *(e.g., nearest neighbor classifiers simply memorize them)*.
- It is the generalization to new data that matters.
- When building classifiers, always set some test data aside!
- Use the test data as rarely as possible!

Machine Learning: Benchmarking

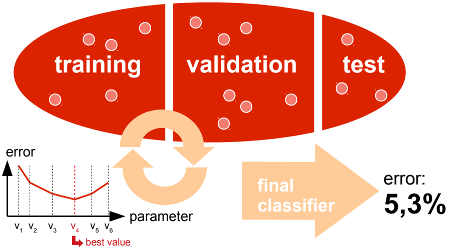




We separate our dataset into **training and testing data**:

- **Tip 1**: choose 'just enough' test data, use the rest for training.
- **Tip 2**: use a roughly balanced class distribution for training.

Machine Learning: Benchmarking

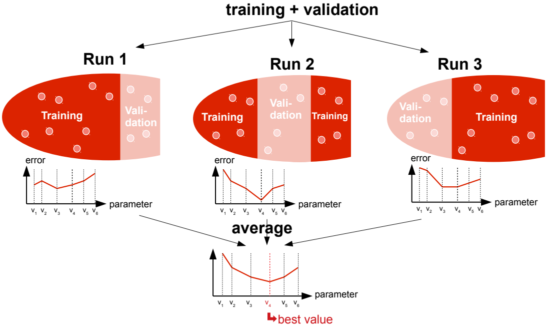


Some of a classifiers' parameters are **learned**, others (also called *hyperparameters*) must be **chosen**.

Common Approach: Grid Search

- For each parameter Θ_i, test a few values $\mathcal{R}_i := \{v_i^1, ..., v_i^{n_i}\}$.
- Train the classifier for each parameter combination in $\mathcal{R}_1 \times \mathcal{R}_2 \times ... \times \mathcal{R}_m$.
- Evaluate the resulting classifier on a separate **validation set**.

Machine Learning: Benchmarking



If there is **little data**, we apply **cross-validation**:

- Split the data into subsets ("*folds*"), train/validate **multiple times**, and average the results.
- Above: **3-fold cross-validation**.

References I

[1] Animali i animaletti: owca nie tylko czarna.
http://studiaparlaama.pl/animali-i-animaletti-owca-nie-tylko-czarna/ (retrieved: Oct 2016).

[2] Axay Pithava: Perceptual Mapping Techniques.
http://www.slideshare.net/AxayPithava/perceptual-mapping-techniques (retrieved: Oct 2016).

[3] Brizzle born and Bread.
https://www.flickr.com/photos/brizzlebornandbred/5292576151/ (retrieved: Oct 2016).

[4] Google DeepDream robot: 10 weirdest images produced by AI 'inceptionism' and users online (Photo: Reuters).
http://www.straitstimes.com/asia/east-asia/alphago-wins-4th-victory-over-lee-se-dol-in-final-go-match (retrieved: Nov 2016).

[5] Google opens Cloud Vision API beta, world + dog asked to try it.
http://www.theregister.co.uk/2016/02/19/google_opens_cloud_vision_api_beta/ (retrieved: Oct 2016).

[6] Scikit-Learn Landing Page.
http://scikit-learn.org (retrieved: Oct 2016).

[7] Spam (Monty Python).
https://en.wikipedia.org/wiki/Spam_(Monty_Python) (retrieved: Oct 2016).

[8] The Value of a Professional Network?
https://www.linkedin.com/pulse/value-professional-network-daniel-tunkelang (retrieved: Oct 2016).

[9] 'Untergang der Titanic' Illustration von Willy Stöwer für die Zeitschrift Die Gartenlaube.
https://de.wikipedia.org/wiki/RMS_Titanic (retrieved: Oct 2016).

[10] The fellowship of the ring, 2001.

References II

[11] S. Bird, E. Klein, and E. Loper.
Natural Language Processing with Python.
O'Reilly Media, Inc., 2009.

[12] Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg S Corrado, and Jeff Dean.
Distributed Representations of Words and Phrases and their Compositionality.
In Advances in Neural Information Processing Systems 26, pages 3111–3119. Curran Associates, Inc., 2013.

[13] Volodymyr Mnih, Koray Kavukcuoglu, David Silver, Andrei A. Rusu, Joel Veness, Marc G. Bellemare, Alex Graves, Martin Riedmiller, Andreas K. Fidjeland, Georg Ostrovski, Stig Petersen, Charles Beattie, Amir Sadik, Ioannis Antonoglou, Helen King, Dharshan Kumaran, Daan Wierstra, Shane Legg, and Demis Hassabis.
Human-level control through deep reinforcement learning.
Nature, 518(7540):529–533, 02 2015.

[14] Mary-Ann Russon.
Google DeepDream robot: 10 weirdest images produced by AI 'inceptionism' and users online.
http://www.ibtimes.co.uk/google-deepdream-robot-10-weirdest-images-produced-by-ai-inceptionism-users-online-1509518
(retrieved: Nov 2016).

[15] C. Szegedy.
Building a deeper understanding of images (Google Research Blog).
https://research.googleblog.com/2014/09/building-deeper-understanding-of-images.html
(retrieved: Nov 2016).