



base.camp Talks
24.04.2020



Eugen Ruppert
ruppert at informatik.uni-hamburg.de

BASE.CAMP TALKS
MACHINE LEARNING I – INTRODUCTION

Contents



Zoom

Introduction

Machine Learning and AI – What is it?

Hands on

Perceptron

Hands on

Support Vector Machine

Hands on

Evaluation

Decision for algorithm

Outlook

Questions?

Zoom

Zoom

Nettiquette

- Turn on your video
- If available, use a headset
- Please mute your microphone when not in the conversation
mute on/off: Alt + a – Push to Talk: Space
- We want to record the session and make it available for people who could not join today
We take privacy seriously. Thus, we will blur/black out your faces, so don't worry about asking questions and giving feedback!

Introduction

- Focus areas: **Big Data**, **Artificial Intelligence**, **Security (BASE)**
- Offering students possibilities
 - student projects
(focus on practice and iterative development)
 - hackathons
 - workshops
- Organizing events
- Networking for interdisciplinary events and courses

<https://basecamp.informatik.uni-hamburg.de/>

- lecture series on current and general topics for Computer Science
- practically oriented
- easy to follow
- interactive examples, not just theoretical knowledge

`https://www.inf.uni-hamburg.de/inst/basecamp/events/
basecamp-talks.html`

- practical introduction to ML
- understand the key terms for ML
- understand how to perform a ML project
- code examples
- **not** a full lecture about ML and learning theory, partial derivatives, etc.
There is a lecture ‘Machine Learning’ by Dr. Victor Uc Cetina

Machine Learning and AI – What is it?

Machine Learning and AI – What is it?



- finding regularities in data
- classification based on training:
 - SPAM classification
 - image recognition (tagging people in photos)
 - text categorization (invoice, technical issue report, contract cancelation)
- emergent 'intelligent' systems (Artificial Intelligence)
- distinction between weak and strong AI
- our topic: useful AI – reduce human effort



Machine Learning and AI – What is it?

Strong AI: Chess



IBM Deep Blue:

[https://en.wikipedia.org/wiki/Deep_Blue_\(chess_computer\)](https://en.wikipedia.org/wiki/Deep_Blue_(chess_computer))

Machine Learning and AI – What is it?

Strong AI: Question Answering



IBM Watson Jeopardy

[https://en.wikipedia.org/wiki/Watson_\(computer\)](https://en.wikipedia.org/wiki/Watson_(computer))

Machine Learning and AI – What is it?

Strong AI: Go



Google Research: Alpha Go

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

Easy vs Hard Problems

Language identification

?

Computerspiele machen Spaß und sind ein Riesemarkt. Und sie sind “Treiber für innovative Technologien”, sagt der zuständige Minister Dobrindt.

?

It's easy to travel the world and feel oddly at home, particularly if you're in one of the planet's 24,000 Starbucks.

Easy vs Hard Problems

Language identification



?

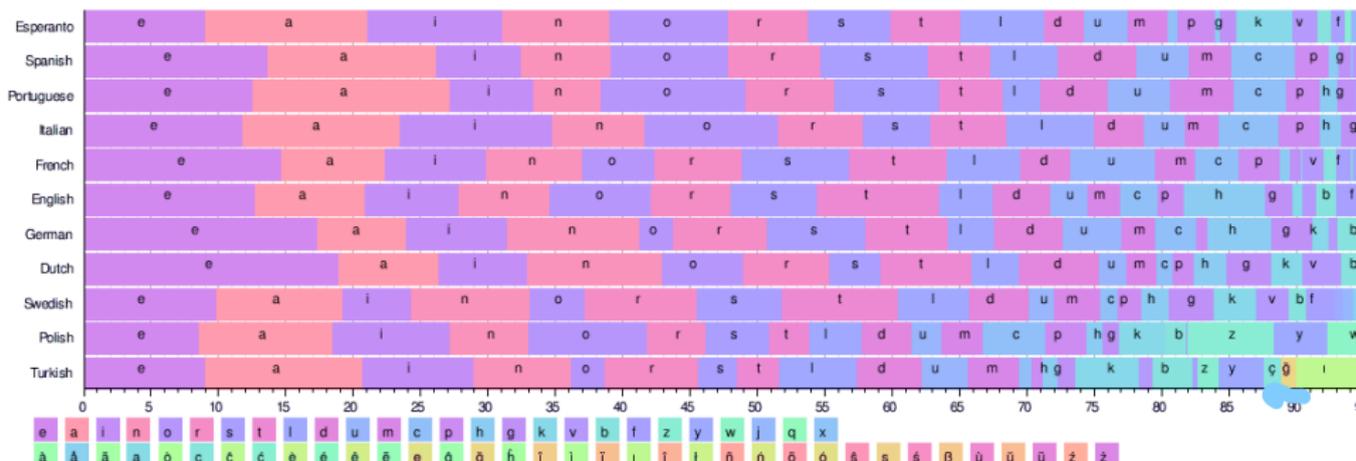
Dörtnala gelip Uzak Asya'dan Akdeniz'e bir kısrak başı gibi uzanan bu memleket, bizim.

Bilekler kan içinde, dişler kenetli, ayaklar çıplak Ve ipek bir halıya benzeyen toprak bu cehennem, bu cennet bizim. Kapansın el kapıları, bir daha açılmasın, Yok edin insanın insana kulluğunu, bu dâvet bizim....

Yaşamak bir ağaç gibi tek ve hür ve bir orman gibi kardeşçesine, bu hasret bizim...

Easy vs Hard Problems

Language identification



Easy vs Hard Problems

Language understanding



Früher stellten die Frauen der Inseln am Wochenende Kopftücher mit Blumentmotiven her, die ihre Männer an den folgenden Montagen auf dem Markt im Zentrum der Hauptinsel verkauften.

<https://slideplayer.org/slide/645960/>

Easy vs Hard Problems

Language understanding

Früher stellten die Frauen der Inseln am Wochenende Kopftücher mit Blumentmotiven her, die ihre Männer an den folgenden Montagen auf dem Markt im Zentrum der Hauptinsel verkauften.

- more than 250,000 interpretations possible!

<https://slideplayer.org/slide/645960/>

Machine Learning and AI – What is it?

Impact



- AI has a large impact on the world
- ethical considerations are important
- Big Data and privacy issues
- misuse opportunities

Interesting talks from the EIT (Prof. Judith Simon and Prof. Ingrid Schneider)

Taming the Machines:

<https://www.inf.uni-hamburg.de/en/inst/ab/eit/taming-the-machines/lecture2go.html>

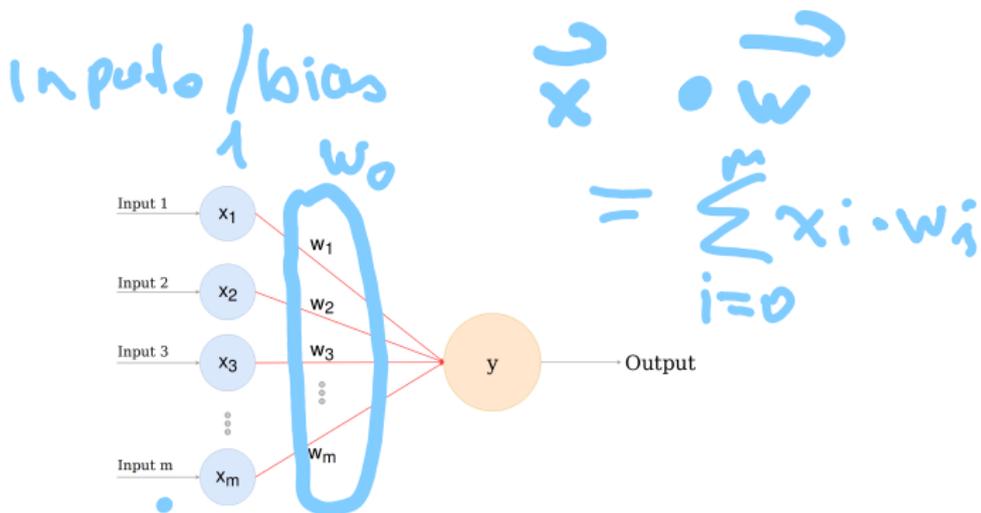
Hands on

- `https://ltdemos.informatik.uni-hamburg.de/rserver/`
- `code: https://git.informatik.uni-hamburg.de/base.camp/ml-intro-talk/-/tree/master`

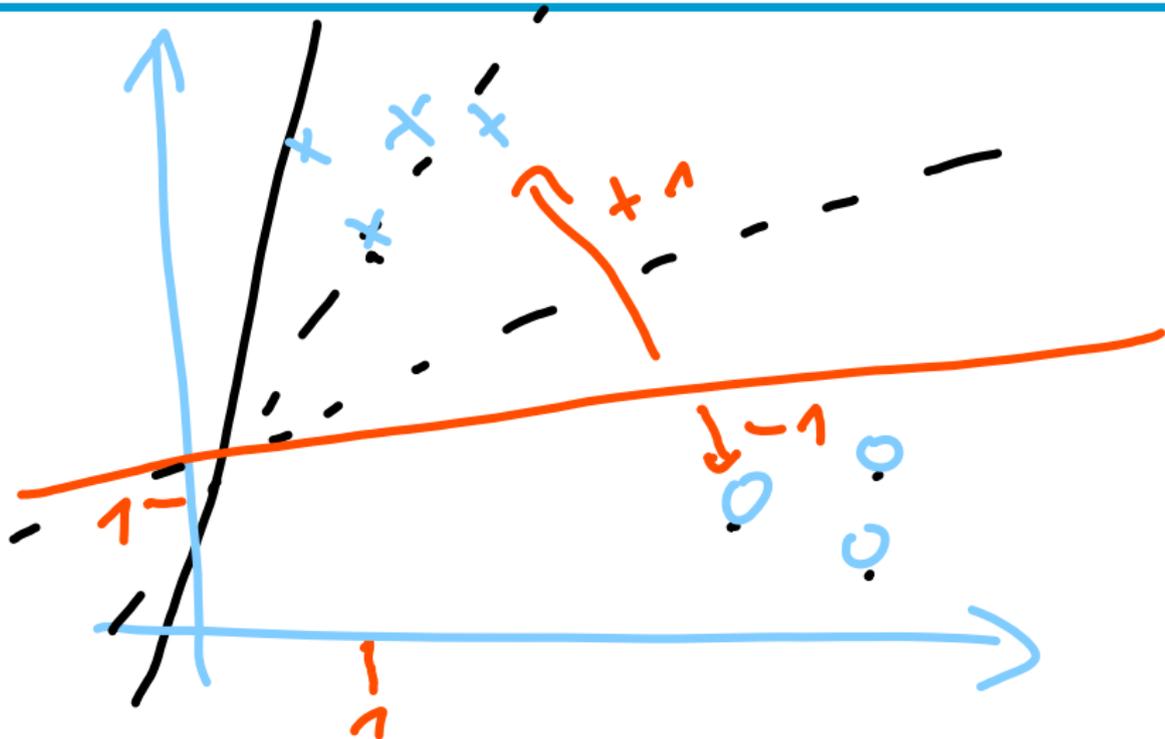
Perceptron

Perceptron

Overview



$$f(x) = mx + b$$
$$= 0.1 \cdot x + 1.5$$



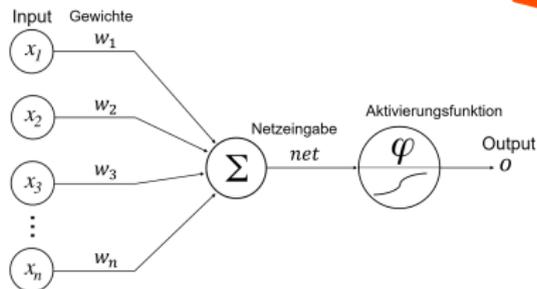
Perceptron

Evaluation

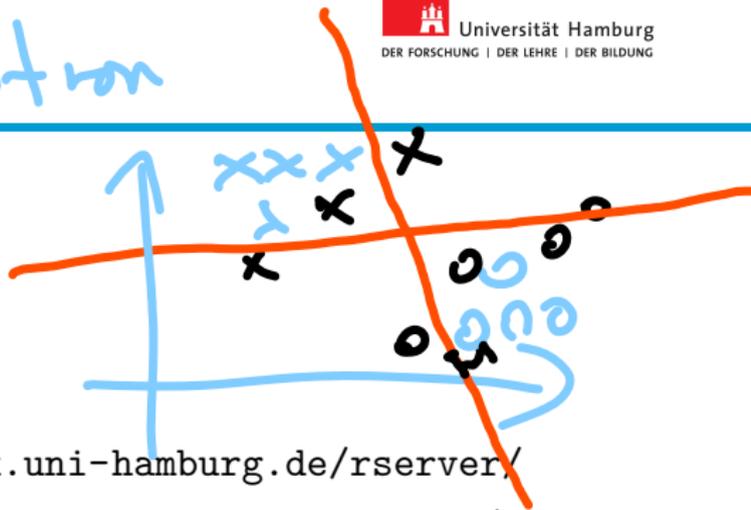
- fast and simple
- separates separable problems
- not very robust as a single classifier

Perceptron

Sneak Peek into the Future



Hands on

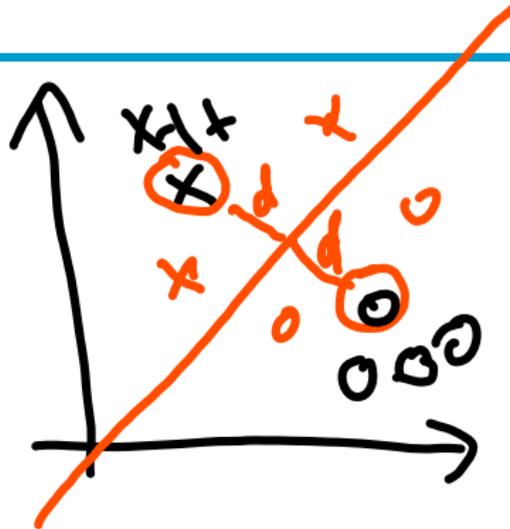


- `https://ltdemos.informatik.uni-hamburg.de/rserver/`
- `code: https://git.informatik.uni-hamburg.de/base.camp/ml-intro-talk/-/tree/master`

Support Vector Machine

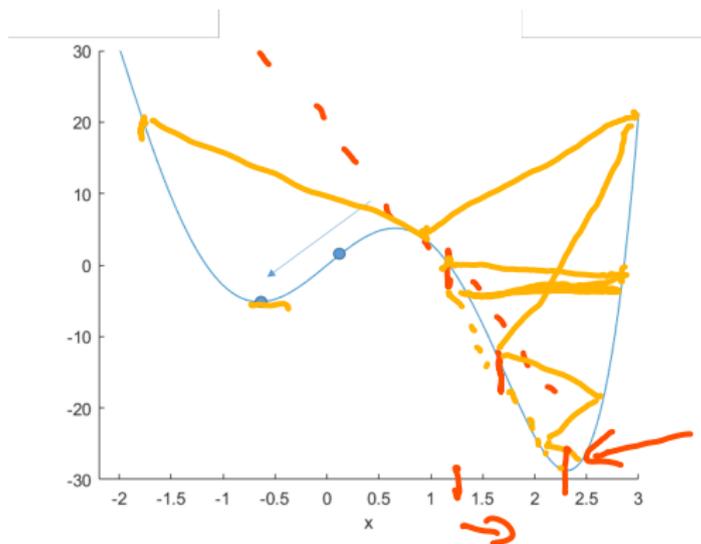
Support Vector Machine

- large-margin classifier
- very robust
- efficient selection of 'support vectors'
- hyperparameters for training



Support Vector Machine

Stochastic Gradient Descent



- get Loss / Error
- update \vec{w}

$$f'(1, 2) = -2$$

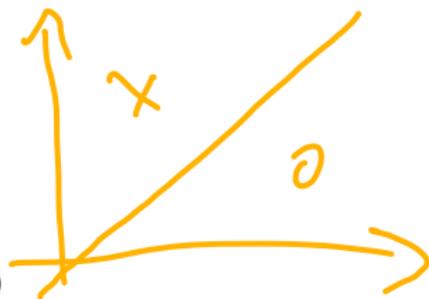
$$\vec{w} = \vec{w} - \eta f'(x)$$

η = learning rate

Support Vector Machine

Hyperplane

- in an n-dimensional space
- hyperplane has n-1 dimensions
- point for 1D, line for 2D, plane for 3D



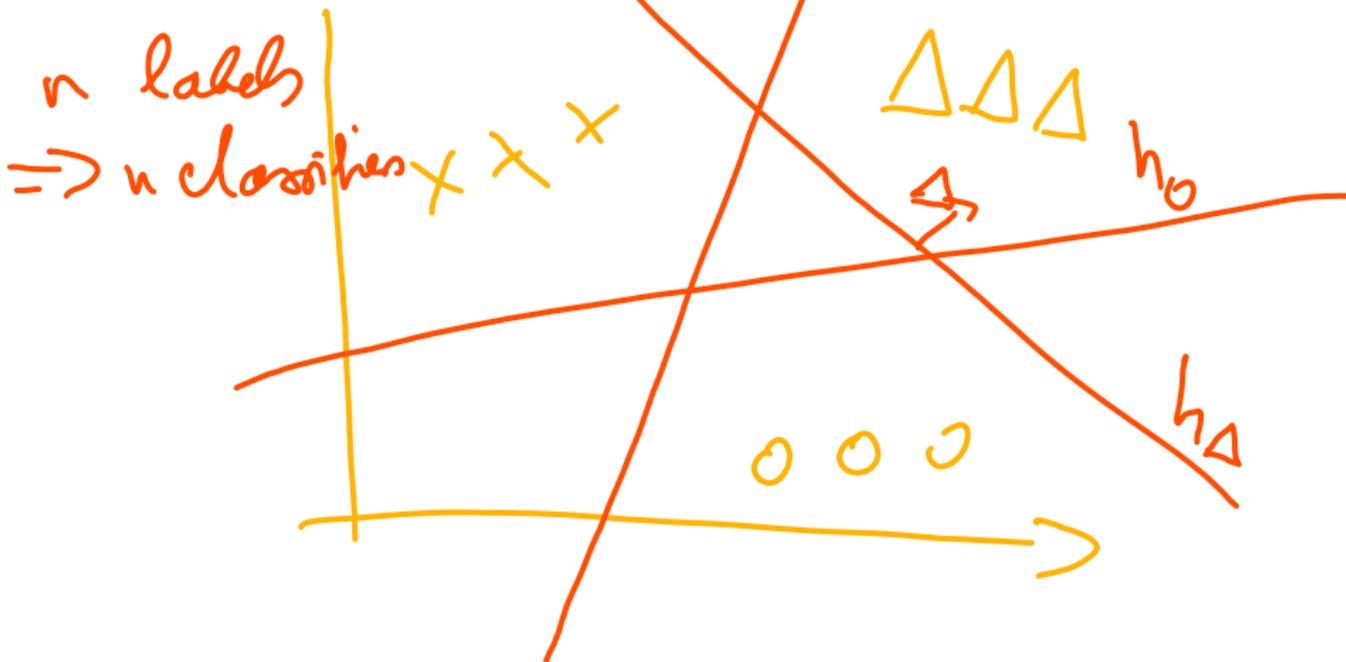
Support Vector Machine

Multiclass SVM

- one-vs-all
- or: one-vs-one

Support Vector Machine

Multiclass SVM-1:all

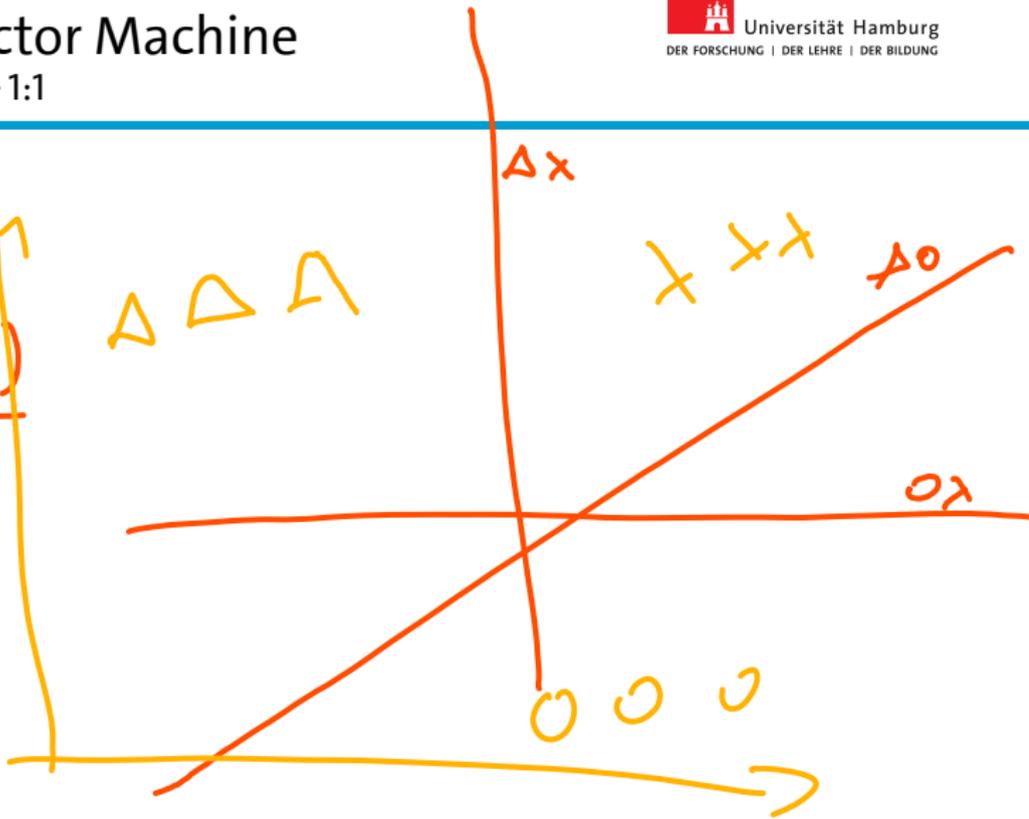


Support Vector Machine

Multiclass SVM-1:1



n labels
 $\rightarrow \frac{n \cdot (n-1)}{2}$
Classes



Hands on

- `https://ltdemos.informatik.uni-hamburg.de/rserver/`
- `code: https://git.informatik.uni-hamburg.de/base.camp/ml-intro-talk/-/tree/master`

Evaluation

Evaluation

Confusion Matrix

- 2 classes: +, -

		<u>gold</u>	
		+	-
<u>predicted</u>	+	TP	FP
	-	FN	TN

Evaluation

Accuracy

$$Accuracy = \frac{TP + TN}{TP + FP + FN + TN}$$

- common measure used for evaluation
- can have problems with imbalanced datasets

Evaluation

Bouncer classification



```
function isRelevant( document, query ) {  
  
}
```

Evaluation

Bouncer classification



```
function isRelevant( document, query) {  
  
return false;  
}
```

Evaluation

Accuracy



- big dataset
- 10 relevant items
- 10,000 irrelevant items

		gold	
		+	-
predicted	+	0	0
	-	10	10,000

$$Accuracy = \frac{100000}{10000 + 10} = 0.999$$

Evaluation

Precision & Recall

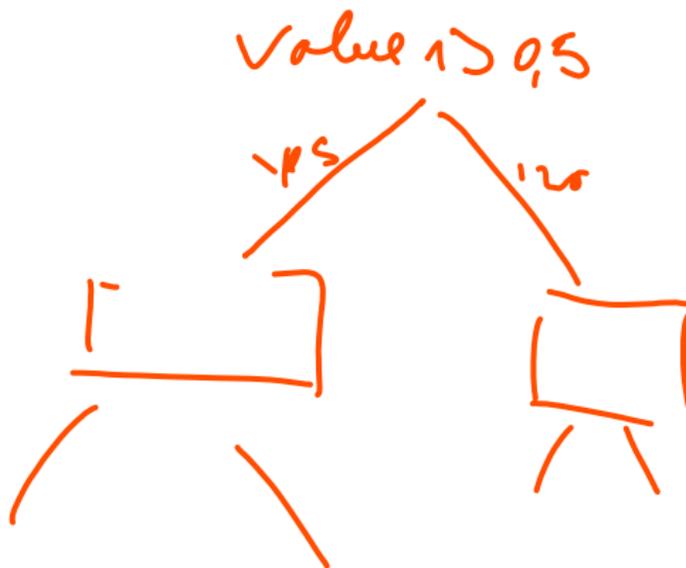


$$\text{Precision} = \frac{TP}{TP + FP} = 1 - \frac{3}{4}$$
$$\text{Recall} = \frac{TP}{TP + FN} = 1 - \frac{3}{5}$$
$$F\text{-score} = \frac{2 * Pr * R}{Pr + R} = 0,6\bar{6}$$

Decision for algorithm

Decision for algorithm

- Classification performance
- Online Learning possibility
- Training/classification time
- Availability of data
- Explainability



Outlook

- Feedback: <https://ep.mafiasi.de/p/basecamp-talks>
- Next week (23.04.2020):
Machine Learning 2 – Deep Learning
- Associated: 14.05.2020 - IBM Talk on Quantum Computing
- 28.05.2020: Otto Talk on Chat Bots

<https://www.inf.uni-hamburg.de/inst/basecamp/events/basecamp-talks.html>

Questions?