

# Exam Preparation and Summary of Topics

## Introduction to Artificial Intelligence S. 5 Bachelor WS21/22

### h\_da Darmstadt

Jonas Weßner

February 21, 2022

## Contents

<b>1</b>	<b>Machine Learning</b>	<b>1</b>
1.1	Definition . . . . .	1
1.2	Examples of ML tasks . . . . .	1
1.3	Categories of ML tasks . . . . .	1
1.3.1	Preproccession . . . . .	2
1.3.2	Error functions . . . . .	2
1.3.3	Metrics . . . . .	2
1.3.4	K-fold cross validation . . . . .	3
<b>2</b>	<b>Knowledge Representation</b>	<b>3</b>
2.1	Knowledge Graphs . . . . .	3
<b>3</b>	<b>Natural Language Processing (NLP)</b>	<b>3</b>

## 1 Machine Learning

### 1.1 Definition

Machine learning is the approach of generating a model based on inputs (training) and using it for making predictions (productive application) instead of programming instructions explicitly. It is therefore a data-driven approach to finding solutions to problems.

### 1.2 Examples of ML tasks

- Spam Filter: Classify emails as spam and not spam
- Stock market analysis: Make recommendations on buying or selling stocks
- Loyal Customer Detection: Given Data about a customer, classify him/her as loyal or non-loyal customer for the future

### 1.3 Categories of ML tasks

- Supervised Learning
  - Classification
  - Regression
- Unsupervised Learning
  - Clustering

- Feature selection
- Feature extraction
- Reinforcement Learning: Try and error mit nach Belohnungsprinzip

### 1.3.1 Preproccession

- Extracting relevant columns →
  - Faster learning process
  - better results
- Filling missing values
  - Deleting rows with missing values
  - Mean
  - Median
  - Use Algorithm that supports missing values

### 1.3.2 Error functions

- Mean absolute error (MAE)
- Mean squared error (MSE)
- Rooted MSE (RMSE)

### 1.3.3 Metrics

#### Accuracy

Accuracy answers the question "What is the probability that a predicition is correct?".

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

It is only good, if the real distribution of positive and negatives in the data is close to symmetric.

#### Precision

Precision answers the question "If we classify something as positive, how probable is it that it is actually positive?".

$$Precision = \frac{TP}{TP + FP}$$

#### Recall

Recall a.k.a. sensitivity answers the question "If a sample is positive, what is the probability we also label it as positive?".

$$Recall = \frac{TP}{TP + FN}$$

## F1 Score

The F1-score divides the true positives by the sum of the true positives and the mean of the false positives and false negatives. This a high F1-score requires the model to make not few false predictions in either direction. Therefore F1-score is better than accuracy if the real distribution of positive and negative values in the dataset is uneven.

$$F1 = 2 \cdot \frac{Precision \cdot Recall}{Precision + Recall} = \frac{TP}{TP + \frac{1}{2} \cdot (FP + FN)}$$

## Validation

### 1.3.4 K-fold cross validation

- split dataset into k subsets
- Perform k trainings on k-1 subsets and use the one remaining set for validation such that each subset has been used for validation exactly once
- calculate the mean of all k iterations

## 2 Knowledge Representation

### 2.1 Knowledge Graphs

A knowledge graph a.k.a. semantic net or ontology is a data structure linking different entities to each other. A common type of knowledge graph is one that consists of triples [subject, predicate, object]. Linked Open Data (LOD) refers to all publicly available data in the internet that can be found using URIs and HTTP.

RDF (Resource Description Framework) is a system for describing such data in nets of triples. SPARQL is a query language that can be used to retrieve data from a database which is using the RDF format. A popular free SPARQL server is Apache Jena Fuseki.

In the following we can see how to extract tuples (name, country) of city names and country names for which the city is the capital of the country and the country is located in Africa. `es` is the prefix for all classes used in this example. `cityname`, `isCapitalOf`, `country` and `isInContinent` are classes of predicates. All identifiers starting with a question mark are variables.

```
1 PREFIX ex: <http://example.com/exampleOntology#>
2 SELECT ?capital
3        ?country
4 WHERE
5 {
6     # match every city here, assign them to variable ?x and
7     # their name and the countries it is a capital for to
8     # variables ?capital and ?y
9     ?x ex:cityname      ?capital    ;
10      ex:isCapitalOf    ?y          .
11     # For all those countries which have capitals,
12     # filter those who are in located in Africa and extract their names
13     ?y ex:countryname   ?country    ;
14      ex:isInContinent   ex:Africa    .
15 }
```

## 3 Natural Language Processing (NLP)