

LECTURER: DR. TAI LE QUY

ANALYTICAL SOFTWARE AND FRAMEWORKS

TOPIC OUTLINE

Introduction to Analytical Software and Frameworks

1

Data Storage

2

Statistical Modeling Frameworks

3

Machine Learning and Artificial Intelligence Frameworks

4

Cloud Computing Platforms, On-Premise Solutions, Distributed Computing

5

Database Technology

6

UNIT 4

MACHINE LEARNING AND ARTIFICIAL INTELLIGENCE FRAMEWORKS

STUDY GOALS



- Understand concepts of machine learning and artificial intelligence
- Know modern frameworks for machine learning
- Get an overview of how to utilize machine learning algorithms in R and Python
- Get an overview of cloud-based machine learning providers and their portfolios



1. What is Machine Learning?
2. How does supervised machine learning differ from unsupervised learning?
3. How do features of available cloud-based machine learning frameworks differ?

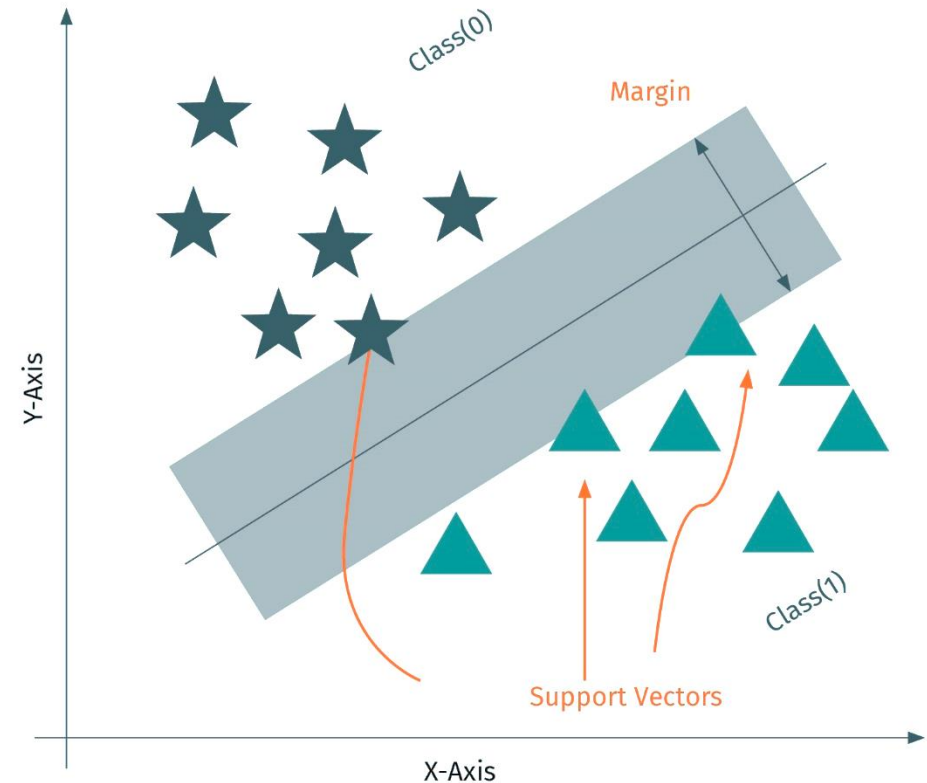
Machine learning is ...

- „[...] about ***machines making sense out of data in much the same way that humans do.*** [...] a type of artificial intelligence whereby an ***algorithm*** or method ***extracts patterns from data.***”¹
- “[...] a field of study that gives ***computers*** the ability to ***learn without being explicitly programmed.*** [...] the process by which a computer can ***work more accurately as it collects and learns from the data it is given.***”²

Three **main types** of machine learning

- **Supervised learning**
 - Classification
 - Regression
- **Unsupervised learning**
 - Clustering
- **Reinforcement learning**

SVM Description for Linearly-Separable Data



MODERN MACHINE LEARNING FRAMEWORKS AND PACKAGES/LIBRARIES

Framework	Description
Apache Spark ML	A scalable stand-alone machine-learning library that can access multiple sources of data . It includes many high-quality algorithms such as Naïve Bayes, linear and logistic regression, and K-means . It is utilized in Java and interoperated with Python and R libraries.
PyTorch	An open-source framework with maximum flexibility and speed to build machine-learning algorithms for a wide range of applications . It provides a powerful multidimensional array to support many routines such as indexing and transposing. This framework is also embeddable through ports to iOS and Android backends.
Scikit-Learn	Built on NumPy, SciPy, and matplotlib . Mainly used when the data-analysis tasks require stand-alone computing on individual computers. Can be integrated via APIs to other languages to model workflows.
Tensorflow	An open-source End-to-End ML platform designed by Google to develop and deploy machine-learning models for different platforms. It has flexible tools and resources using data-flow graphs to train models . Runs on CPU, GPU, TPU , hence can be integrated on all platforms , e.g., like mobile, IoT, web browser, cloud.
Keras	An API on top of TensorFlow 2 designed “for human beings, not machines”. It follows best practices for reducing cognitive load by offering consistent & simple APIs, minimizing the number of actions required for common use cases, and providing clear error messages .

MODERN MACHINE LEARNING FRAMEWORKS AND PACKAGES/LIBRARIES

Framework	Description
Apple's Core ML	A framework that is designed to integrate machine-learning techniques into an iOS app . It can be personalized for domain-specific functionality and is optimized for on-device performance . It supports techniques such as image analysis, natural language processing, and decision trees .
Microsoft Azure ML	An interactive (drag and drop) framework and platform to build predictive analysis model on connected experimental datasets with a visual workspace . The developed models may be published as web services for custom apps.
Google ML, Vertex AI	A cloud-based unified AI platform to build, deploy, and scale ML models faster with pre-trained and custom tooling.
Amazon ML, SageMaker	A cloud-based machine-learning framework and platform designed by Amazon. It integrates multi-sourced data, builds models, and generates predictions . It also provides a visualization wizard and is mainly utilized for classification and regression tasks .

KERAS VS. TENSORFLOW VS. PYTORCH

	TensorFlow	Keras	PyTorch
API/Scope	<p>End-to-end platform for machine learning, originating from Google. Includes Keras as tf.keras for a user-friendly start. Covers full range for ML projects an every step of the ML workflow, from data management to distributed model training on clusters and deployment to many platforms (IoT devices, web browsers, mobile iOS and Android, web APIs in the cloud). Runs on CPU, GPU, or TPU.</p>	<p>High-level APIs focused on simplicity of use, enabling rapid development and prototyping for common use cases. Default backend is now TensorFlow 2 with all its capabilities.</p>	<p>Direct competitor of TensorFlow with similar capabilities and its own ecosystem. Originating from FAIR (Facebook AI Research). Covers ML projects from research prototyping to production deployment. Lower-level APIs. High-level library with scikit-learn compatibility via skorch from the PyTorch ecosystem. Runs on CPU, GPU, or TPU.</p>
Licensing	Open source	Open source	Open source

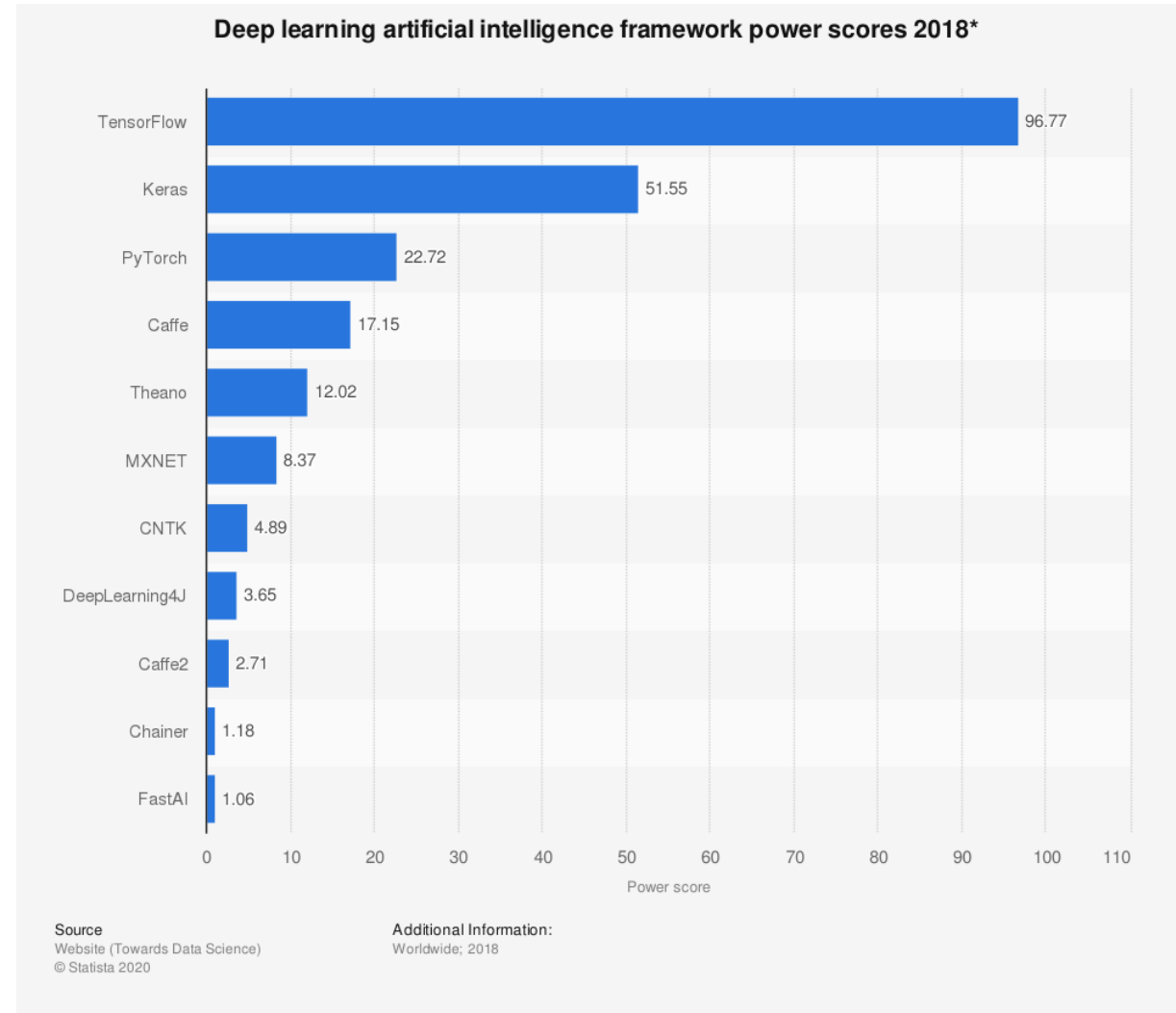
CLOUD-BASED MACHINE LEARNING/MACHINE LEARNING AS A SERVICE (MLaaS)

	Amazon ML, Amazon SageMaker	Microsoft Azure ML	Google ML, Vertex AI
Common Features (in different shapes)	Pre-trained models, services, APIs , e.g., computer vision, language, conversation; Development of own ML models ; Integration into other cloud services of that vendor; AutoML (training models without code and minimal expertise); MLOps (manage data and models at scale), One-click-deployments ; Support for responsible/fair ML models ; Distributed training ; Model versioning ; Experiment tracking ; Hosted collaborative and/or local notebooks		
Differences	SageMaker Studio IDE and Rstudio on SageMaker with built-in debugger and profiler, SageMaker Edge Manager for optimizing, securing, monitoring, maintaining edge devices	ML Studio , a GUI-based cloud IDE for “ drag-and-drop machine learning ”, Native ML pipelines (Azure ML pipeline), PyTorch Enterprise support	Jupyter notebook based Vertex AI workbench with managed or user-managed notebooks (Deep Learning VM instances), Vertex AI pipelines

- **Many packages for machine learning** are available **in the R ecosystem** provided by the “Comprehensive R Archive Network” (**CRAN**), e.g., the popular **tidymodels**.
- **TensorFlow** and **Keras** may be used through **wrapper packages** (**tensorflow**, **keras**, **kerasR**) that allow their usage from R code.
- **PyTorch** is **available** through the **torch** package without dependencies to Python.

MACHINE LEARNING IN PYTHON

- **Direct access to most popular machine-learning frameworks, APIs, libraries, and platforms, e.g., `scikit-learn`, TensorFlow, Keras, PyTorch**
- One main driver of Python's popularity





- Understand concepts of machine learning and artificial intelligence
- Know modern frameworks for machine learning
- Get an overview of how to utilize machine learning algorithms in R and Python
- Get an overview of cloud-based machine learning providers and their portfolios

SESSION 4

TRANSFER TASK

TRANSFER TASK

Open the provided RStudio/Jupyter notebooks and work through the examples to build and train a machine-learning model.

Please let me know if you got stuck so we can work through it together.

TRANSFER TASK

Gather in three groups and work through the following task!

Visit

1. Azure ML (<https://azure.microsoft.com/en-us/services/machine-learning/>)
2. Amazon SageMaker (<https://aws.amazon.com/pm/sagemaker/>)
3. Google Vertex AI (<https://cloud.google.com/vertex-ai>)

and research to explain what the respective vendor understands by **AutoML** and how features and tools support AutoML in its portfolio.

TRANSFER TASK
PRESENTATION OF THE RESULTS

Please present your
results.

The results will be
discussed in plenary.





1. Which of the following frameworks offers machine learning in the cloud with built-in algorithms?
 - a) Amazon ML/SageMaker
 - b) Microsoft Azure ML
 - c) Google ML/Vertex AI
 - d) Apache Mahout ML



2. Scikit-learn (sklearn) is the official machine learning package in Python and is built on ...
- a) the Matplotlib package.
 - b) the Scipy package.
 - c) the Numpy package.
 - d) all of the above.



3. **model.compile** is a Keras method, which can be called on a model instance to ...

- a) configure the model for training, e.g., define optimizer, loss function, and metrics to be evaluated.
- b) train the developed prediction model on the training set.
- c) evaluate the performance of the developed prediction model.
- d) set the architecture of the developed prediction model.

LIST OF SOURCES

- Allaire, J. J. (n.d.). *TensorFlow for R*. <https://tensorflow.rstudio.com/guide/keras/>
- Amazon Web Services, Inc. (2021a). *Amazon SageMaker—Machine Learning Solution*. <https://aws.amazon.com/pm/sagemaker/>
- Amazon Web Services, Inc. (2021b). *Machine Learning | AWS AI and ML | Amazon Web Services (AWS)*. <https://aws.amazon.com/machine-learning/>
- Cielen, D., Meysman, A., & Ali, M. (2016). *Introducing data science: Big data, machine learning, and more, using Python tools*. Manning Publications.
- Google Cloud (n.d.-a). *Vertex AI*. Google Cloud. <https://cloud.google.com/vertex-ai>
- Google Cloud (n.d.-b). *AI & Machine Learning Products*. Google Cloud. <https://cloud.google.com/products/ai>
- Hothorn, T. (2021). *CRAN Task View: Machine Learning & Statistical Learning*. <https://CRAN.R-project.org/view=MachineLearning/>
- Kaggle Inc. (n.d.). *State of Data Science and Machine Learning 2020*. <https://www.kaggle.com/kaggle-survey-2020>
- Keras Team (n.d.). *Keras: The Python deep learning API*. <https://keras.io/>
- Kirk, M. (2017). *Thoughtful machine learning with Python: A test-driven approach* (1st edition). O'Reilly.
- Microsoft (2021). *Azure Machine Learning—ML as a Service | Microsoft Azure*. <https://azure.microsoft.com/en-us/services/machine-learning/>
- PyTorch Team (n.d.). *PyTorch*. <https://www.pytorch.org>
- RStudio Blog (2020). *Introducing torch for R*. <https://rstudio.comhttps://www.rstudio.com/blog/torch/>
- Statista (2018). *AI deep learning frameworks ranking 2018*. <https://www-statista-com.pxz.iubh.de:8443/statistics/943038/ai-deep-learning-frameworks-ranking/>
- TensorFlow Team (n.d.). TensorFlow. *TensorFlow*. <https://www.tensorflow.org>
- Zhang, J. (2021). Comparing Google Cloud Platform, AWS and Azure. *Georgian Impact Blog*. <https://medium.com/georgian-impact-blog/comparing-google-cloud-platform-aws-and-azure-d4a52a3adbd2>

© 2021 IU Internationale Hochschule GmbH

This content is protected by copyright. All rights reserved.

This content may not be reproduced and/or electronically edited, duplicated, or distributed in any kind of form without written permission by the IU Internationale Hochschule GmbH.