

**Module Description, available in: EN, FR**

# Machine Learning

**General Information**

## Number of ECTS Credits

3

## Module code

FTP\_MachLe

## Valid for academic year

2023-24

## Last modification

2022-10-06

## Coordinator of the module

Christoph Würsch (OST, christoph.wuersch@ost.ch)

## Explanations regarding the language definitions for each location:

- Instruction is given in the language defined below for each location/each time the module is held.
- Documentation is available in the languages defined below. Where documents are in several languages, the percentage distribution is shown (100% = all the documentation).
- The examination is available 100% in the languages shown for each location/each time it is held.

	Lausanne			Lugano	Zurich		
<b>Instruction</b>	X E 100%	X F 100%	X E 100%		X E 100%		
<b>Documentation</b>	X E 100%				X E 100%		
<b>Examination</b>	X E 100%				X E 100%		

## Module Category

FTP Fundamental theoretical principles

## Lessons

2 lecture periods and 1 tutorial period per week

**Entry level competences**

## Prerequisites, previous knowledge

- **Math:** basic calculus / linear algebra / probability calculus (e.g., derivatives, matrix multiplication, normal distribution)
- **Statistics:** basic descriptive statistics (e.g., mean, variance, co-variance, histograms, box plots)
- **Programming:** good command of any structured programming language (e.g., Python, Matlab, R, Java, C, C++)
- **Analytics:** basic data analysis methods (data pre-processing, linear & logistic regression)

## Brief course description of module objectives and content

Machine learning (ML) emerged out of artificial intelligence and computer science as the academic discipline concerned with “giving computers the ability to learn without being explicitly programmed” (A. Samuel, 1959). Today, it is the methodological driver behind the mega-trend of digitalization. ML experts are highly sought after in industry and academia alike.

This course builds upon basic knowledge in math, programming and analytics/statistics as is typically gained in respective undergraduate courses of diverse engineering disciplines. From there, it teaches the foundations of modern machine learning techniques in a way that focuses on practical applicability to real-world problems. The complete process of building a learning system is considered:

- formulating the task at hand as a learning problem;
- extracting useful features from the available data;
- choosing and parameterizing a suitable learning algorithm.

Covered topics include cross-cutting concerns like ML system design and debugging (how to get intuition into learned models and results) as well as feature engineering; covered algorithms include (amongst others) Support Vector Machines (SVM) and ensemble methods.

## Aims, content, methods

### Learning objectives and acquired competencies

- Students **know** the **background and taxonomy** of machine learning methods
- On this basis, they **formulate** given problems as **learning tasks** and **select a proper learning method**
- Students **are able to convert** a data set into a proper **feature set** fitting for a task at hand
- They **evaluate** the chosen **approach** in a structured way using proper design of experiment
- Students **know how** to select models, and „**debug**“ features and learning algorithms if results do not fit expectations
- Students are able to leverage on the evaluation framework to **tune the parameters** of a given system and **optimize** its performances
- Students **have seen examples of different data** sources / problem types and **are able to acquire additional expert knowledge** from the scientific literature

### Contents of module with emphasis on teaching content

- **Introduction** (ca. 2 weeks): Convergence for participants with different backgrounds
- **Supervised learning** (ca. 7 weeks): Learn from labeled data  
*Cross-cutting topics:* such as feature engineering; ensemble learning; instance vs. model-based approaches, debugging ML systems  
*Algorithms:* e.g. kNN, decision tree, SVM, ensemble learning (bagging, boosting), graphical models (Bayesian networks), gradient based approaches  
*bias-variance tradeoff:* hyperparameter tuning, cross-validation, performance metrics
- **Unsupervised learning** (ca. 3 weeks): Learning without labels  
*Algorithms:* e.g., clustering, dimensionality reduction, anomaly detection, archetypal analysis
- **Special chapters** (ca. 2 weeks):  
*Algorithms:* e.g., reinforcement learning, recommender systems, hidden Markov / Gaussian mixture models

### Teaching and learning methods

Classroom teaching; programming exercises (e.g., in Python 3, Jupyter notebooks, Orange)

### Literature

- Aurélien Géron: "*Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow*", Concepts, Tools, and Techniques to Build Intelligent Systems, O'Reilley Media, 2022
- T. Mitchell, "*Machine Learning*", 1997
- C. M. Bishop, "*Pattern Recognition and Machine Learning*", 2006
- Simon Rogers, Mark Girolami: "*A First Course in Machine Learning*", ISBN-13: 978-0367574642, ?Chapman and Hall/CRC; 2. Edition, 2016
- G. James et al., "*An Introduction to Statistical Learning*", 2014

## Assessment

### Certification requirements

Module does not use certification requirements

### Basic principle for exams

**As a rule, all the standard final exams for modules and also all resit exams are to be in written form**

#### Standard final exam for a module and written resit exam

**Kind of exam**

written

**Duration of exam**

120 minutes

**Permissible aids**

*Aids permitted as specified below:*

**Permissible electronic aids**

The permitted electronic aids will be defined by the lecturer in the first week of the semester.

**Other permissible aids**

1 A4 page (front and back) of handwritten notes (no book, no slides, no further notes)

#### Special case: Resit exam as oral exam

**Kind of exam**

oral

**Duration of exam**

30 minutes

**Permissible aids**

No aids permitted

Description du module, disponible en: EN, FR

## Machine Learning

### Informations générales

Nombre de crédits ECTS

3

Code du module

FTP\_MachLe

Valable pour l'année académique

2023-24

Dernière modification

2022-10-06

Coordinateur/coordinatrice du module

Christoph Würsch (OST, christoph.wuersch@ost.ch)

Explication des définitions de langue par lieu :

- Les cours se dérouleront dans la langue définie ci-dessous par lieu/exécution.
- Les documents sont disponibles dans les langues définies ci-dessous. Pour le multilinguisme, voir la répartition en pourcentage (100% = documents complets)
- L'examen est disponible à 100% dans chaque langue sélectionnée pour chaque lieu/exécution.

	Lausanne			Lugano	Zurich		
Leçons	X E 100%	X F 100%	X E 100%		X E 100%		
Documentation	X E 100%				X E 100%		
Examen	X E 100%				X E 100%		

Catégorie de module

FTP bases théoriques élargies

Leçons

2 leçons et 1 leçon de pratique par semaine

### Compétences préalables

Connaissances préalables, compétences initiales

- **Mathématiques:** calcul de base / algèbre linéaire / calcul des probabilités (c.-à-d. dérivées, multiplication de matrices, distribution normale)
- **Statistiques:** statistiques descriptives de base (c.-à-d. moyennes, variances et covariances, histogrammes, boîtes à moustaches)
- **Programmation:** bonne maîtrise d'un langage de programmation structuré (c.-à-d. Python, Matlab, R, Java, C, C++)
- **Analyse:** méthodes d'analyse des données élémentaires (prétraitement des données, régression linéaire & logistique)

## Brève description du contenu et des objectifs

L'apprentissage automatique ou Machine Learning (ML) est le fruit de l'intelligence artificielle et de la science informatique, et se définit comme une discipline universitaire visant à « donner aux ordinateurs la faculté d'apprendre sans avoir été explicitement programmés pour le faire ». (A. Samuel, 1959). A l'heure actuelle, l'apprentissage automatique est le moteur méthodologique des grandes tendances de fond de la digitalisation. Les experts en ML sont très recherchés aussi bien dans le milieu universitaire que dans le monde industriel.

Ce cours s'appuie sur des connaissances de base en mathématiques, programmation et analyses/statistiques telles qu'elles sont dispensées dans les différents cursus bachelor d'ingénieurs. Partant de là, il enseigne les bases des techniques modernes de l'apprentissage automatique par une méthode centrée sur la mise en pratique et l'application à des problèmes de la vie réelle. Le processus de création d'un système d'apprentissage est abordé dans sa globalité:

- formulation de la tâche à l'étude sous forme de problème d'apprentissage ;
- extraction de caractéristiques intéressantes à partir des données disponibles ;
- sélection et paramétrage d'un algorithme d'apprentissage convenant à la situation.

Les sujets abordés englobent des concepts transversaux tels que la conception et le débogage de système ML (comment interpréter les modèles appris et leurs résultats) ainsi que l'ingénierie des caractéristiques (feature engineering) ; les algorithmes abordés comprennent (parmi d'autres) les machines à vecteurs de support (Support Vector Machines) ainsi que les méthodes d'apprentissage ensemblistes.

## Objectifs, contenus, méthodes

### Objectifs d'apprentissage, compétences à acquérir

- Les étudiants **connaissent le contexte et la taxonomie** des méthodes d'apprentissage automatique ;
- Sur cette base, ils **traduisent** des problèmes donnés **en tâches d'apprentissage et sélectionnent une méthode d'apprentissage adaptée** ;
- Les étudiants **sont capables de convertir** un set de données en un **set de caractéristiques** adapté à la tâche à l'étude ;
- Ils **évaluent l'approche choisie** de façon structurée en faisant appel à un design d'expérience adapté ;
- Les étudiants **savent comment** sélectionner des modèles, et «**déboguer**» des caractéristiques et des algorithmes d'apprentissage si les résultats ne répondent pas aux attentes ;
- Les étudiants sont capables d'utiliser les résultats d'évaluation pour **optimiser les paramètres** d'un système donné et **maximiser** ses performances ;
- Les étudiants **ont pris connaissance d'exemples de différentes sources de données** / types de problèmes et sont **capables de consulter la littérature scientifique pour acquérir des connaissances** spécialisées **additionnelles**.

### Contenu des modules avec pondération du contenu des cours

- **Introduction** (ca. 2 semaines): convergence des participants ayant suivi diverses formations
- **Apprentissage supervisé** (ca. 7 semaines) : méthodes de modélisation à partir de données labellisées  
*Sujets transversaux:* feature engineering ; ensemble Learning ; débogage des systèmes ML  
*Algorithmes:* p. ex., kNN, arbre décisionnel, SVM, apprentissage ensembliste, modèles graphiques (réseaux bayésiens), méthodes basées sur le gradient  
*biais-variance* : réglage des hyperparamètres, validation croisée, mesures de performance
- **Apprentissage non supervisé** (ca. 3 semaines) : méthodes de modélisation à partir de données non-labellisées  
*Algorithmes:* p.ex., algorithmes de regroupement, réduction de dimensionnalité, détection d'anomalie, analyse de prototype
- **Chapitres spéciaux** (ca. 2 semaines):  
*Algorithmes:* p.ex., apprentissage par renforcement, systèmes de recommandation, modèles de Markov / modèles de mélange gaussien

### Méthodes d'enseignement et d'apprentissage

Cours en classe; exercices de programmation (p.ex., Python 3, Jupyter notebooks, Orange)

### Bibliographie

- Aurélien Géron: "*Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow*", Concepts, Tools, and Techniques to Build Intelligent Systems, O'Reilly Media, 2022
- T. Mitchell, "*Machine Learning*", 1997
- C. M. Bishop, "*Pattern Recognition and Machine Learning*", 2006
- Simon Rogers, Mark Girolami: "*A First Course in Machine Learning*", ISBN-13: 978-0367574642, ? Chapman and Hall/CRC; 2. Edition, 2016
- G. James et al., "*An Introduction to Statistical Learning*", 2014

## Evaluation

### Conditions d'admission

Le module n'utilise pas de conditions d'admission.

### Principe pour les examens

**En règle générale, tous les examens de fin de module réguliers et les examens de rattrapage sont organisés sous la forme écrite**

### Examen de fin de module régulier et examen écrit de répétition

#### Type de l'examen

écrit

#### Durée de l'examen

120 minutes

#### Aides autorisées

*Les aides suivantes sont autorisées:*

#### Aides électroniques autorisées

Les aides électroniques autorisées seront définies par l'enseignant au cours de la première semaine du semestre.

#### Autres aides autorisées

1 page A4 (recto et verso) de notes manuscrites (pas de livre, pas de diapositive, pas d'autres notes)

### Cas spécial: examen de répétition oral

#### Type de l'examen

oral

#### Durée de l'examen

30 minutes

#### Aides autorisées

Sans aides