

EE-311—Apprentissage et intelligence artificielle

1. Introduction

Michael Liebling

<https://moodle.epfl.ch/course/view.php?id=16090>

21 février 2025 (compilé le 21 février 2025)



Organisation de la classe

Heures

Vendredi 8h00–10h00 BC01 (Cours)

Vendredi 10h00–midi BC07–08 (Exercices/Labo)

Fiche de cours

https://edu.epfl.ch/coursebook/fr/apprentissage-et-intelligence-artificielle-EE-311?cb_cycle=bama_cyclebachelor&cb_section=el

Quartier général du cours : site Moodle EE-311

Contient les liens vers les resources, interface de soumission, quiz

The screenshot shows the EPFL Moodle course page for "Apprentissage et intelligence artificielle". The top navigation bar includes the EPFL logo, moodle link, and user Michael Liebling. The main content area displays course information, a sidebar with contact lists, forums, announcements, and various course documents.

<https://moodle.epfl.ch/course/view.php?id=16090>

Michael Liebling

EE-311—Apprentissage machine / 1. Introduction

2 / 61

Obtenir de l'aide

Attention : Certains problèmes/exercices seront difficiles et/ou la méthodologie pour les résoudre ne sera pas directement disponible dans le matériel du cours.

Remèdes :

- Prévoyez suffisamment de temps pour faire vos devoirs
- Assistez aux cours en direct et prenez des notes sur le contenu du tableau noir : il y aura souvent des exemples et des conseils de départ pour la plupart des problèmes.
- Assistez aux sessions d'exercices/laboratoires en direct et prenez des notes : Les assistant·e·s donneront des points de départ
- Discutez avec les autres étudiant·e·s de votre groupe de travail
- Posez des questions sur le forum Ed de la classe (voir ci-dessous).



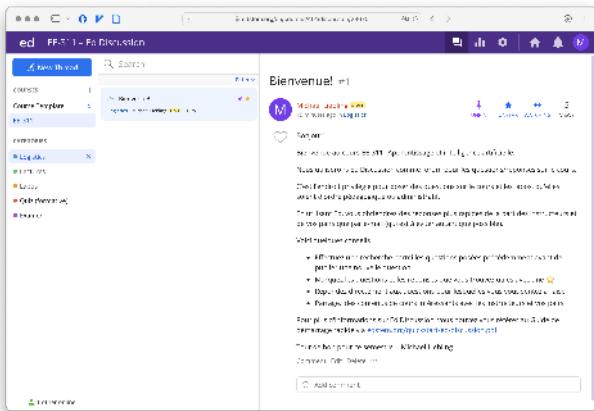
Michael Liebling

EE-311—Apprentissage machine / 1. Introduction

3 / 61

Forum de Discussion Ed

Ed Discussion (edstem.org) est un outil de communication qui a été développé pour une large applicabilité dans toutes les disciplines universitaires, avec de nombreuses caractéristiques qui se prêtent particulièrement bien à l'ingénierie.



Guide de démarrage rapide :

<https://edstem.org/quickstart/ed-discussion.pdf>

Forum : règles d'étiquette

- N'utilisez pas Ed comme substitut unique aux séances en classe. Bien que ce cours offre de nombreuses ressources en ligne, il ne s'agit PAS d'un cours à distance. Il vous sera difficile de résoudre les problèmes sans jamais venir aux séances de labos.
- Veuillez attendre la fin de la session de discussion pour demander de l'aide. Des indications pour démarrer sont généralement données pendant la session de labo.
- Abstenez-vous de publier des solutions partielles ou complètes des problèmes. Il est important que chacun·e ait la possibilité de trouver des solutions par soi-même.
- Vérifier si d'autres ont déjà posté la même question ou une question similaire
- Assurez vous que vos questions soient précises.

Guide rapide pour formuler ses questions

Titre préparez un bon titre (= court mais descriptif) pour votre question.
N'ajoutez pas de bruit (pas de mention “urgent”).

Raison : inciter les gens à lire votre question, faciliter vos révisions.

Catégorie sélectionnez une catégorie (Logistics, Lectures, Labos, Quiz, Examen) et une sous-catégorie (Lecture 01, Labo 05, Quiz 07, etc.)

Mise en forme Utilisez les possibilités de formattage offertes par Ed (\LaTeX , balises de code Python)

Multiples questions Créez des posts distincts pour les questions qui concernent un thème ou un problème distinct

Reproductibilité Qu'avez-vous entrepris avant d'être bloqué·e ? Quelles étapes doit-on suivre pour reproduire votre problème ?

But : permettre d'identifier la nature de votre problème et de vous fournir une réponse pertinente.

Courtoisie Les formes de politesse (“bonjour”, “au revoir”, “merci d'avance”) optionnelles lorsque vous posez la question **mais essentiel d'indiquer si la solution offerte a résolu le problème** : marquer $\checkmark \heartsuit$, commenter “merci” ou indiquer si besoin de clarifier

Poser des questions sur ed 1/2

The screenshot shows the 'ed Discussion' interface. On the left, there's a sidebar with 'Courses' (CS 101, ECON 102, MATH 201, ENGG 202, Playground selected), 'Categories' (General, Lectures, Tutorials, Problem Sets, Assignments, Midterm, Exam), and a status bar indicating '42 others online'. The main area is titled 'Quadratic equation' and shows a post by 'Anonymous' from 'Lectures - WI' 2 hours ago. The post asks, 'How do we solve $ax^2 + bx + c = 0$?'. Below it, Scott Maxwell (START) replies with the quadratic formula:
$$x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$$
. Emily Kwong (WI) adds a note about the graph being a parabola. At the bottom, there are buttons for 'Open a thread' and 'Read and respond to threads'.

Poser des questions sur ed 2/2

Post a question

Ask, with confidence.

Cancel **New Question** Post

Select **Type** ————— **Question** Post

Insert **Title** ————— Title

Select **Category** ————— Category General Logistics Sections **Assignments** Code Social

Comment vos connaissances seront-elles évaluées ?

Note finale (évaluation sommative) :

- Série d'exercices hebdomadaire notée portant sur le cours et le labo (en général, sous la forme d'un programme Python, rendu via Moodle) : 15%
- Examen final : 85% (format choix multiple sur papier, en présentiel)

Évaluation formative (= ne contribue pas à la note finale) :

- Quiz hebdomadaire sur la matière du cours, avec feedback, sur Moodle.
- Mock exam (mis à disposition en semaine 11)

Organisation des séances de labo et attentes

Les labos jouent un rôle essentiel pour l'apprentissage, pour illustrer et pour mettre en pratique les concepts discutés au cours. La partie labo comprend deux volets :

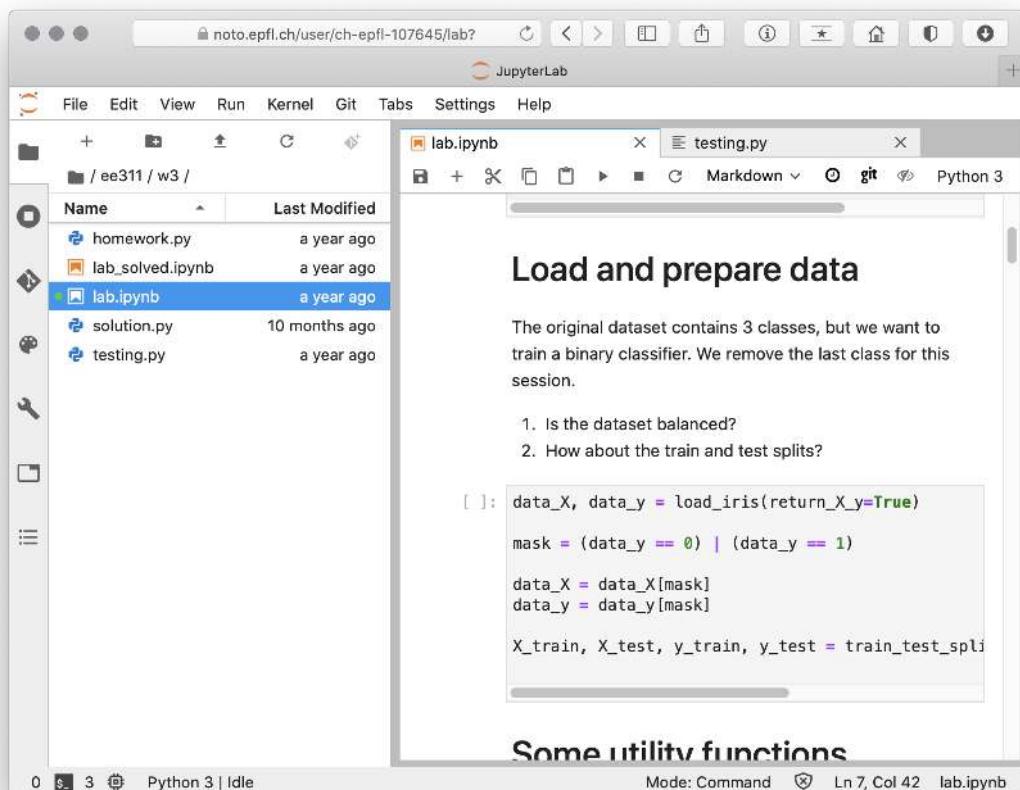
1. une séance interactive en classe avec des problèmes présentés progressivement par un·e assistant·e ; une version résolue du problème est fournie à la fin de la séance.
2. un programme à écrire que vous soumettrez individuellement ; une fois votre code écrit et vérifié, vous le soumettrez par Moodle.

scikit-learn : apprentissage machine en Python

<https://scikit-learn.org/stable/index.html> Scikit-learn est un package (collection d'outils, fonctions, données) Python pour l'apprentissage.

The screenshot shows the official scikit-learn website at <https://scikit-learn.org/stable/index.html>. The header includes the logo, navigation links for Install, User Guide, API, Examples, and More, and a search bar. The main content area features a blue banner with the text: "Simple and efficient tools for predictive data analysis", "Accessible to everybody, and reusable in various contexts", "Built on NumPy, SciPy, and matplotlib", and "Open source, commercially usable - BSD license". Below the banner, there are three main sections: "Classification" (Identifying which category an object belongs to), "Regression" (Predicting a continuous-valued attribute associated with an object), and "Clustering" (Automatic grouping of similar objects into sets). Each section includes a brief description, applications (e.g., Spam detection, image recognition for Classification; Drug response, Stock prices for Regression; Customer segmentation, Grouping experiment outcomes for Clustering), algorithms (e.g., SVM, nearest neighbors, random forest for Classification; SVR, nearest neighbors, random forest for Regression; k-Means, spectral clustering, mean-shift for Clustering), and a small image illustrating the concept. At the bottom, there are links for "Getting Started", "What's New in 0.22.1", and "GitHub".

Jupyter Notebook et [noto.epfl.ch](#)



Introduction à l'apprentissage machine

But de ce premier cours : vous fournir un aperçu des bases de l'apprentissage machine et de la notation utilisée.

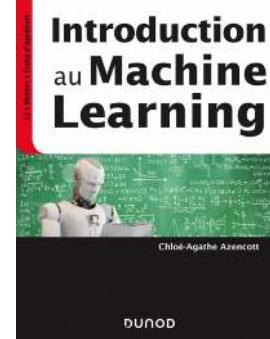
Contenu

- Introduction à l'apprentissage machine
 - définitions (apprentissage, données, modèle appris)
 - exemples et limites
- Types d'apprentissage : apprentissage supervisé et non-supervisé
- Exemples d'apprentissage supervisé
 - classification (par apprentissage supervisé)
 - régression (par apprentissage supervisé)
- Exemples d'apprentissage non-supervisé
 - clustering
 - réduction de dimension
- Estimation de densité
- Apprentissage semi-supervisé
- Apprentissage par renforcement

Ouvrage de référence et source

Ces transparents sont basés en grande partie sur le texte de Chloé-Agathe Azencott "Introduction au Machine Learning", Dunod, 2019

ISBN 978-210-080153-4



L'auteure a mis le texte (sans les exercices) à disposition ici :
http://cazencott.info/dotclear/public/lectures/IntroML_Azencott.pdf

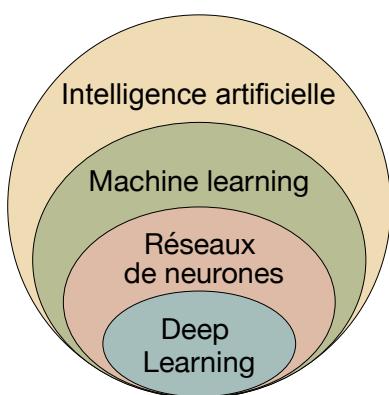
Avertissement : Bien que ces transparents partagent la notation mathématique, la structure de l'exposition (en partie), et certains exemples avec le livre, ils ne constituent qu'un complément et non un remplacement ou une source unique pour la couverture des matières du cours. À ce titre, ces transparents ne se substituent pas au texte.

Michael Liebling

EE-311—Apprentissage machine / 1. Introduction

14 / 61

Qu'est-ce que le machine learning ?



Apprentissage Définition générique (Fabien Benureau (2015)) :

L'apprentissage est une modification d'un comportement sur la base d'une expérience.

Définition applicable à :

- programme informatique
- robot
- animal de compagnie
- être humain



Programmation explicite des actions d'un robot : apprentissage automatique ?



Robot School. Programming For Kids - iOS / MacOS / Android

22,143 views • Jan 27, 2015

1 37 0 SHARE SAVE ...



Next is Great
42 subscribers

SUBSCRIBE

Robot School (en fait : jeu d'apprentissage de la *programmation*)
<https://www.youtube.com/watch?v=5rVBKXcv4h4>

<https://apps.apple.com/us/app/robot-school-programming-for-kids/>

Michael Liebling

EE-311—Apprentissage machine / 1. Introduction

16 / 61

Le robot qui racle (apprentissage automatique)



Le robot qui racle

90,509 views • Sep 27, 2019

1 94 SHARE SAVE ...



Valais Wallis
6.02K subscribers

SUBSCRIBE

Emmanuel Pignat, Sylvain Calinon, et al. Idiap

<https://www.youtube.com/watch?v=a6KpHR6jkSc>

Michael Liebling

EE-311—Apprentissage machine / 1. Introduction

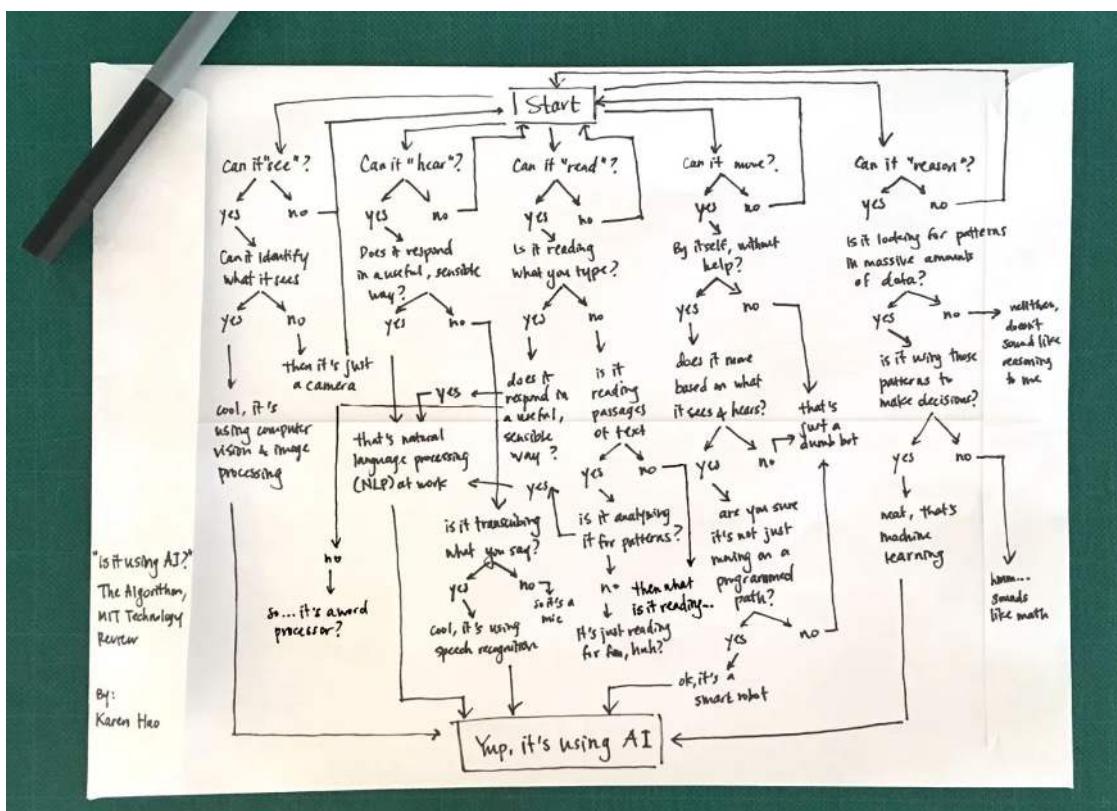
17 / 61

Apprentissage automatique (=machine learning)

Définition (Arthur Samuel, 1959) : l'apprentissage (pour un programme informatique) est automatique quand ce programme a la capacité d'apprendre sans que cette modification ne soit explicitement programmée.

Type de programme	Entrée	Sortie
programme classique	procédure + données	réponses
programme d'apprentissage automatique	données + réponses	procédure (pour obtenir des (nouvelles) réponses à partir de (nouvelles) données)

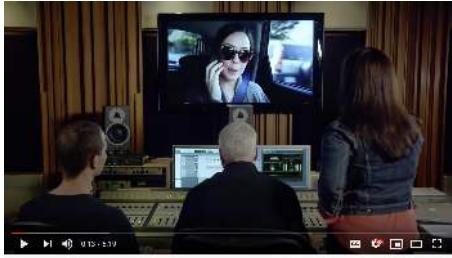
Est-ce que c'est de l'IA ?



Exemples d'applications de l'apprentissage machine automatique

Remplacement de voix

<https://www.youtube.com/watch?v=mexN6d8QF9o>



Signal Processing and Machine Learning

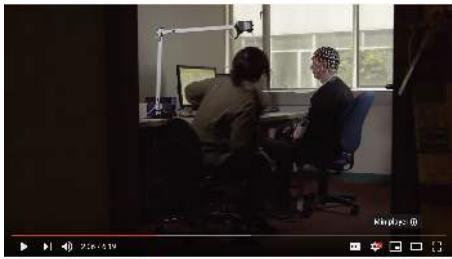
10:20 | 100% | 115 15 4K SHARE 84 RAVE ...

IEEE Signal Processing Society
IEEE-SPL YouTube Channel

SUBSCRIBE

Interface cerveau-machine

<https://youtu.be/mexN6d8QF9o?t=118>



Signal Processing and Machine Learning

10:20 | 100% | 115 15 4K SHARE 84 RAVE ...

IEEE Signal Processing Society
IEEE-SPL YouTube Channel

SUBSCRIBE

Michael Liebling

EE-311—Apprentissage machine / 1. Introduction

20 / 61

Traduction automatique

<https://youtu.be/mexN6d8QF9o?t=236>



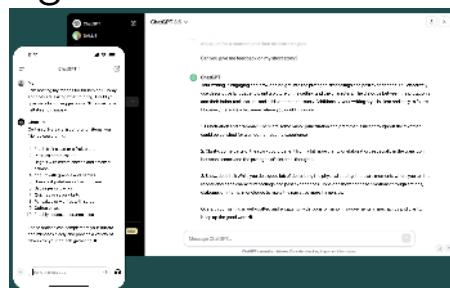
Signal Processing and Machine Learning

10:20 | 100% | 115 15 4K SHARE 84 RAVE ...

IEEE Signal Processing Society
IEEE-SPL YouTube Channel

SUBSCRIBE

Agents conversationnels



Cartes de fidélité et de rabais



Algorithme classiques (pas d'apprentissage nécessaire) :

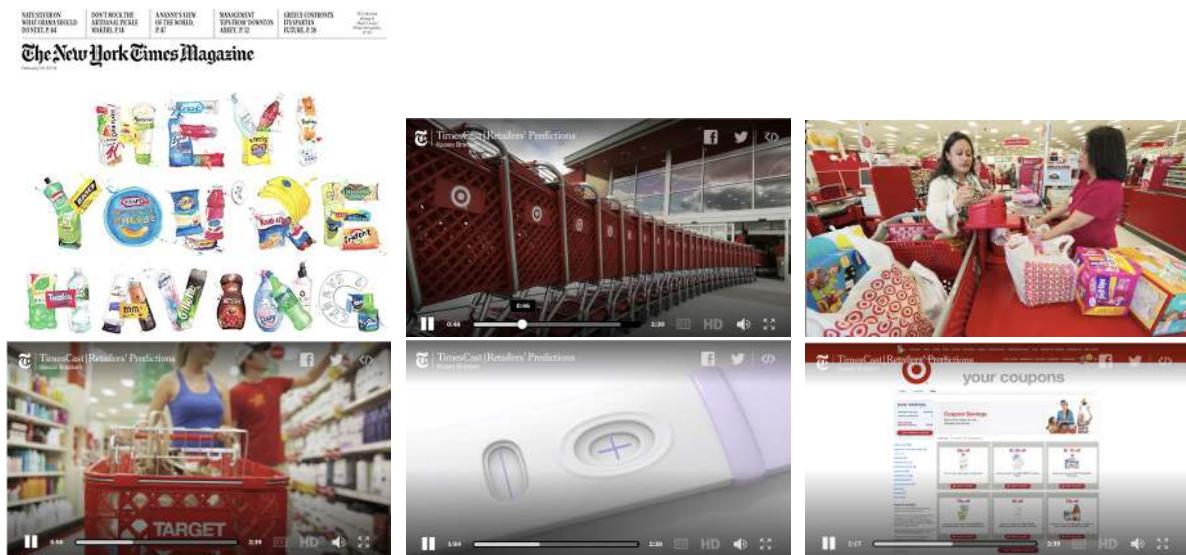
Liste des achats effectués dans le passé par la cliente ou le client → calcul du total dépensé (évtl. par catégorie d'achat) → offre d'un rabais individuel proportionnel au total dépensé

Michael Liebling

EE-311—Apprentissage machine / 1. Introduction

21 / 61

Prédiction des habitudes d'achat futures sur la base d'achats précédents (avec des effets inattendus)



<https://www.nytimes.com/video/magazine/100000001367956/timescast--retailers-predictions.html>

Charles Duhigg "How Companies Learn Your Secrets," The New York Times Magazine, Feb 16, 2012

<https://www.nytimes.com/2012/02/19/magazine/shopping-habits.html>
Michael Liebling EE-311—Apprentissage machine / 1. Introduction

22 / 61

Prédiction d'achats (suite)

Apprentissage automatique

Données+réponses : chronologie d'achat pour de nombreux clients (achats passés et futurs)

Modèle (obtenu par entraînement via un algorithme d'apprentissage sur les données ci-dessus) : permet de suggérer de la marchandise ciblée (sur la base de la chronologie d'achat passés d'une personne) et par la même de prédire des changements de vie majeurs imminents (naissance, mise en ménage, ...)

Ingédients fondamentaux du machine learning

1. données (à partir desquelles l'algorithme apprend)
+
2. *algorithme d'apprentissage* : procédure qu'on applique à ces données
=
3. un *modèle appris* : distribué à l'utilisateur·trice qui en fait usage (plus besoin des données).

Entraînement = faire tourner un algorithme d'apprentissage sur un jeu de données

Algorithme d'apprentissage ≠ modèle appris

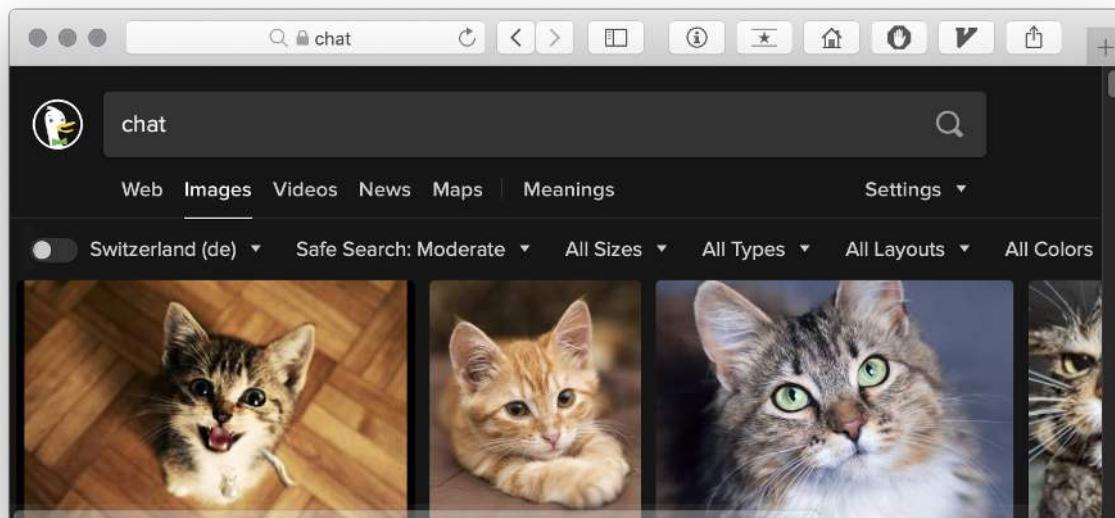
Note : le modèle appris est un algorithme classique !

Michael Liebling

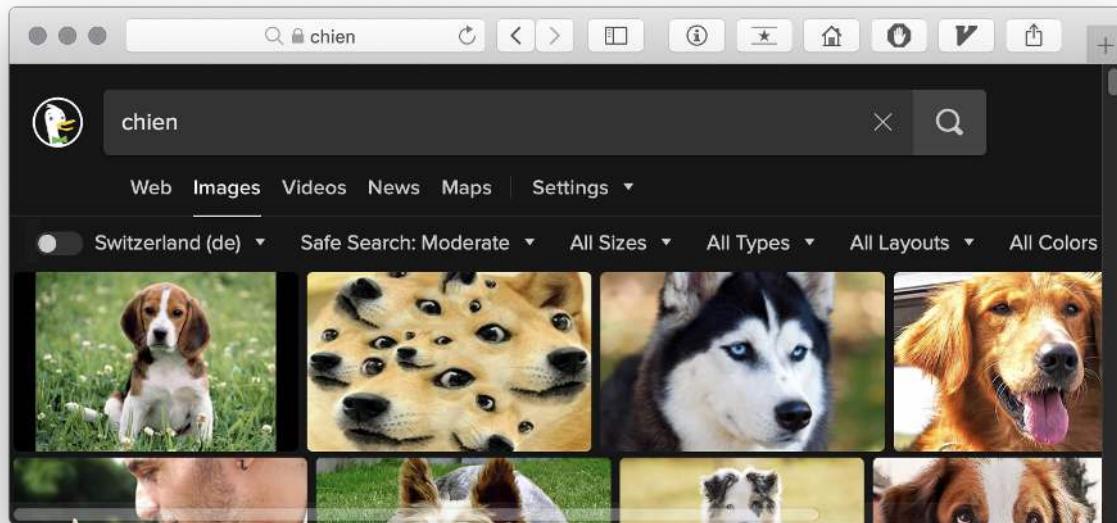
EE-311—Apprentissage machine / 1. Introduction

24 / 61

Les succès du machine learning : trouve-moi une image de chat



Les succès du machine learning : trouve-moi une image de chien

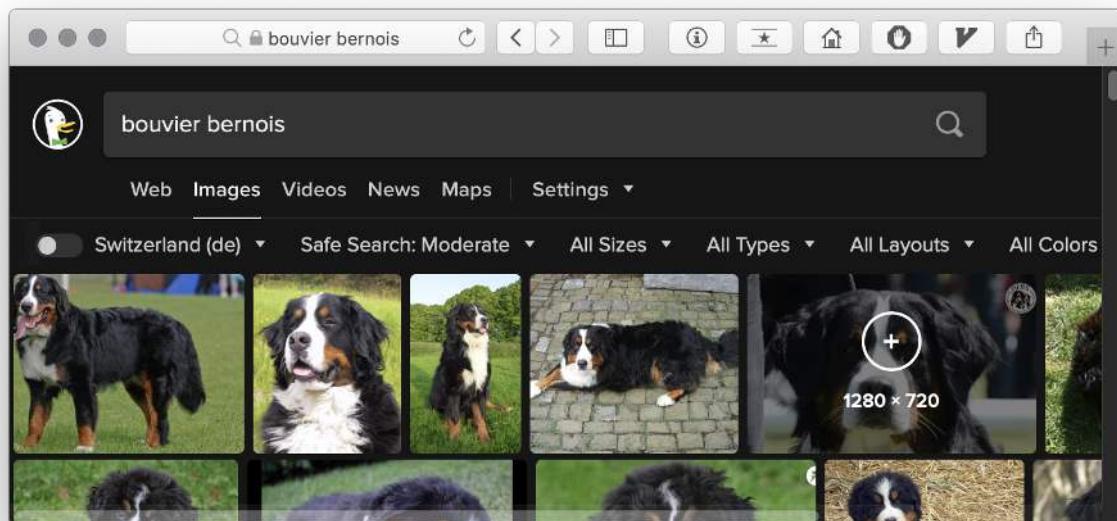


Michael Liebling

EE-311—Apprentissage machine / 1. Introduction

26 / 61

Les succès du machine learning : trouve-moi une image de bouvier bernois

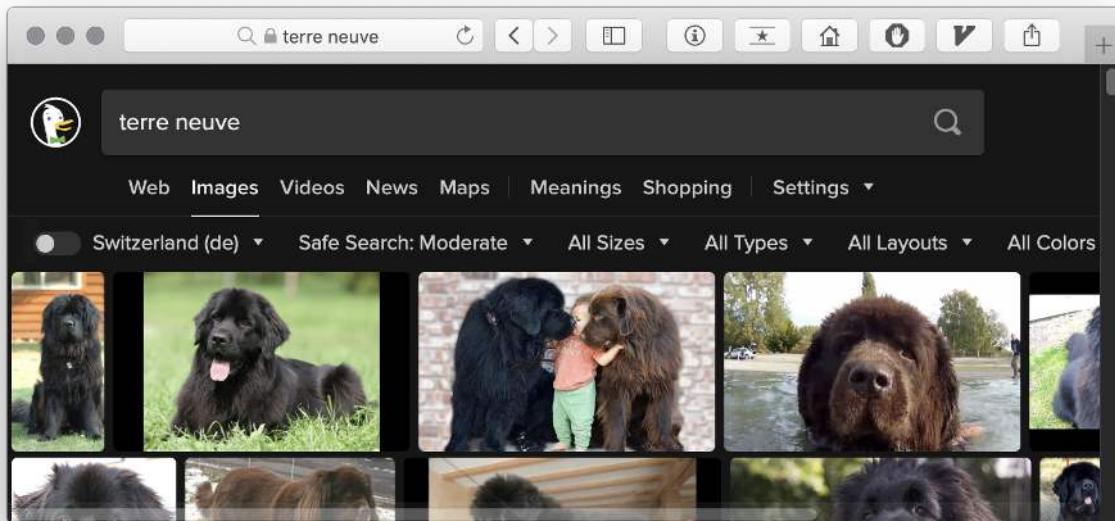


Michael Liebling

EE-311—Apprentissage machine / 1. Introduction

27 / 61

Les succès du machine learning : trouve-moi une image de terre-neuve

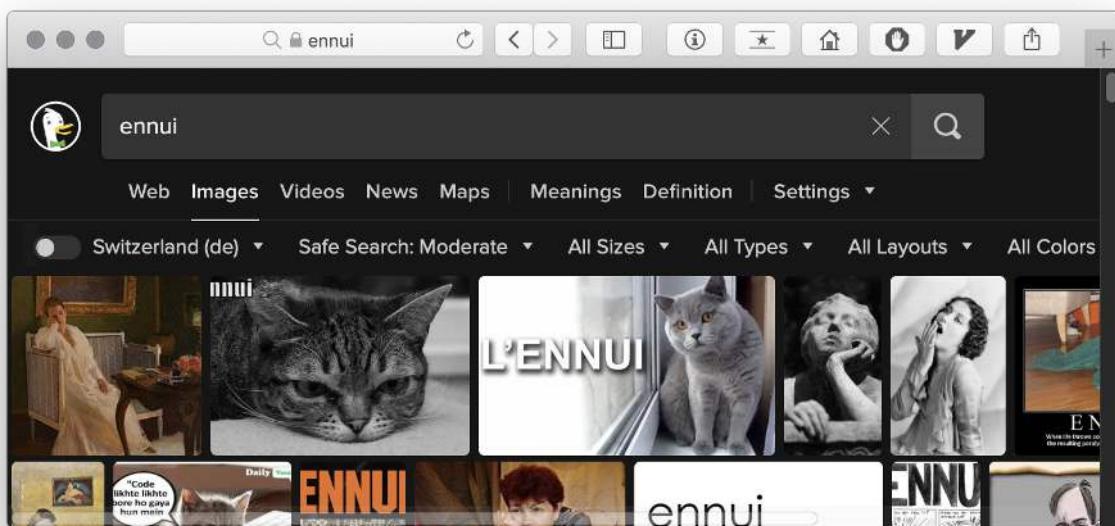


Michael Liebling

EE-311—Apprentissage machine / 1. Introduction

28 / 61

Les succès du machine learning : trouve-moi une image de... [introduire un concept abstrait ou obscur]



Prenons garde de ne pas naïvement extrapoler à de nouveaux domaines les bonnes performances du machine learning obtenus dans des domaines très restreints qui bénéficient de données annotées gigantesques

Michael Liebling

EE-311—Apprentissage machine / 1. Introduction

29 / 61

Les données, nerf de la guerre

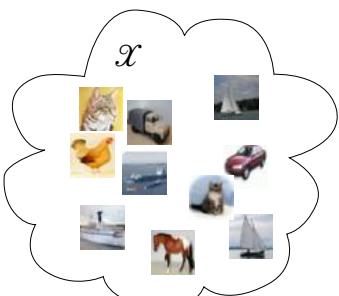
les données, nous les voulons (à choix) :

- en suffisance, en grande quantité
- annotées (cataloguées, organisées, étiquetées,...)
- accessibles (numérisées, accessibles via internet)
- anonymes, distribuables, dans le domaine public, propriétaires
- représentatives de l'étendue des situations rencontrées dans l'application envisagée, équilibrées

Types de problèmes de machine learning

- Apprentissage supervisé :
les données d'entraînement sont étiquetées
 - Apprentissage non supervisé :
les données d'entraînement ne sont pas étiquetées (= données non-étiquetées)
- ... et différentes combinaisons : apprentissage semi-supervisé.

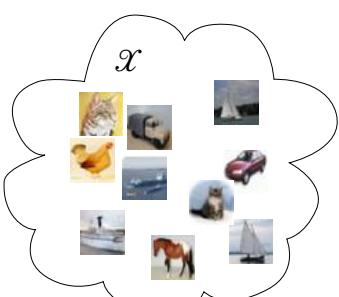
Apprentissage supervisé : espace des images



Espace des images
 $32 \times 32 \times 3$

(M. Liebling, images : CIFAR-10)

Apprentissage supervisé : observations

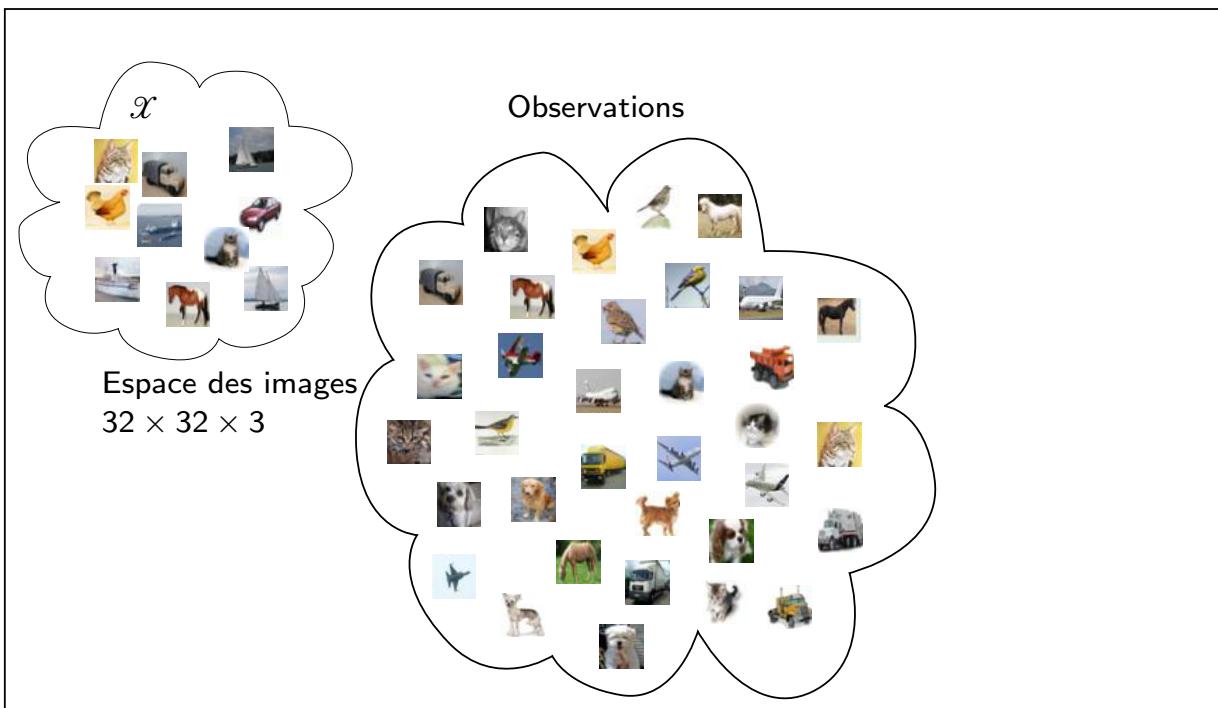


Espace des images
 $32 \times 32 \times 3$

Observations

(M. Liebling, images : CIFAR-10)

Apprentissage supervisé : n observations



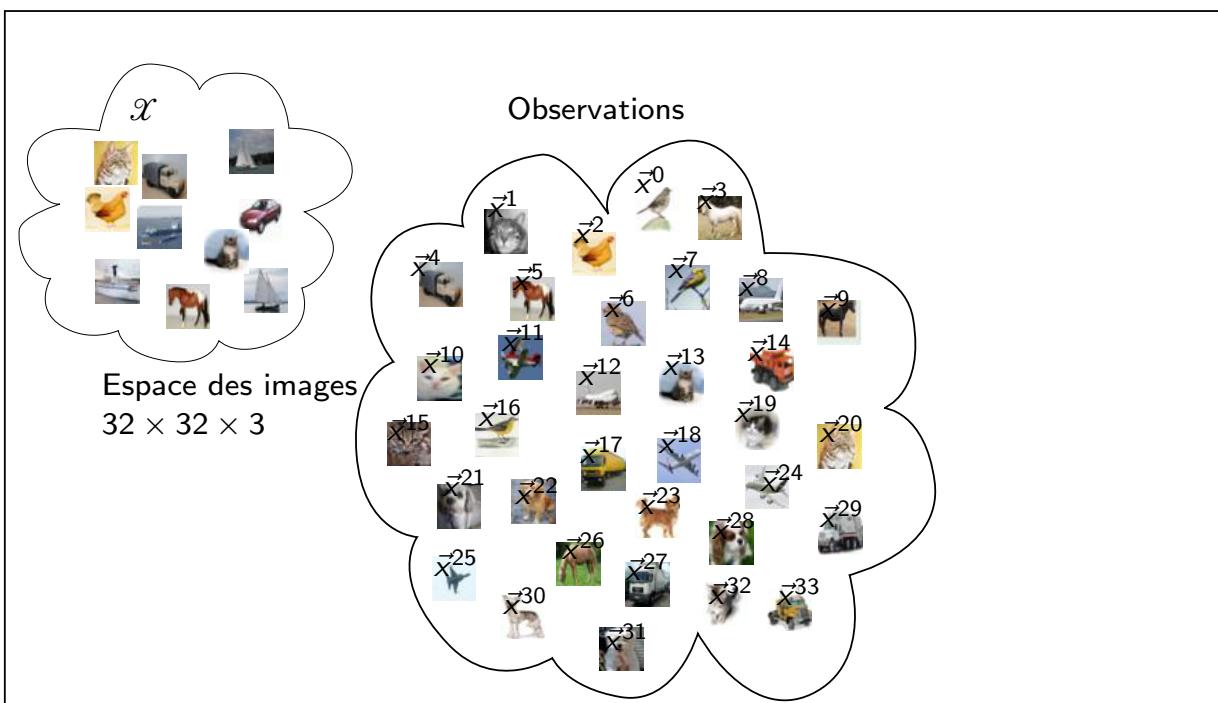
(M. Liebling, images : CIFAR-10)

Michael Liebling

EE-311—Apprentissage machine / 1. Introduction

34 / 61

Apprentissage supervisé : observations $\{\vec{x}^i\}_{i=1,\dots,n}$



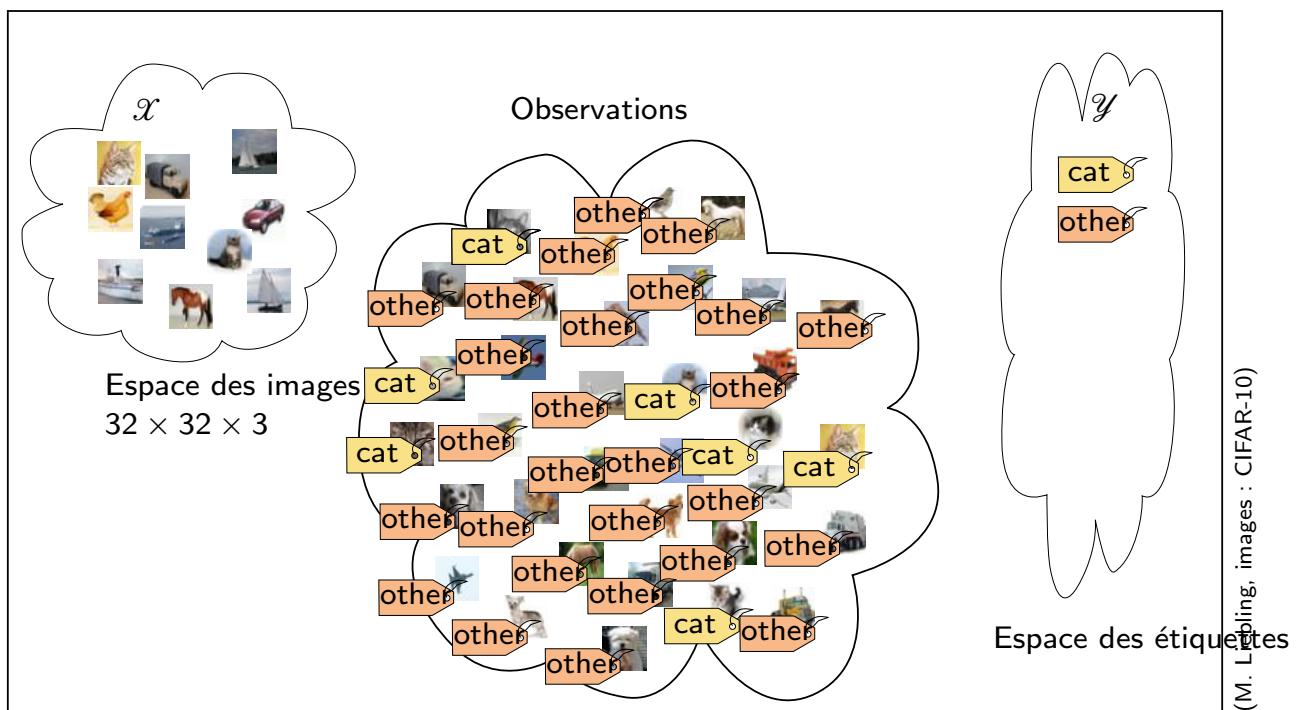
(M. Liebling, images : CIFAR-10)

Michael Liebling

EE-311—Apprentissage machine / 1. Introduction

35 / 61

Apprentissage supervisé : étiquetage (2 classes)

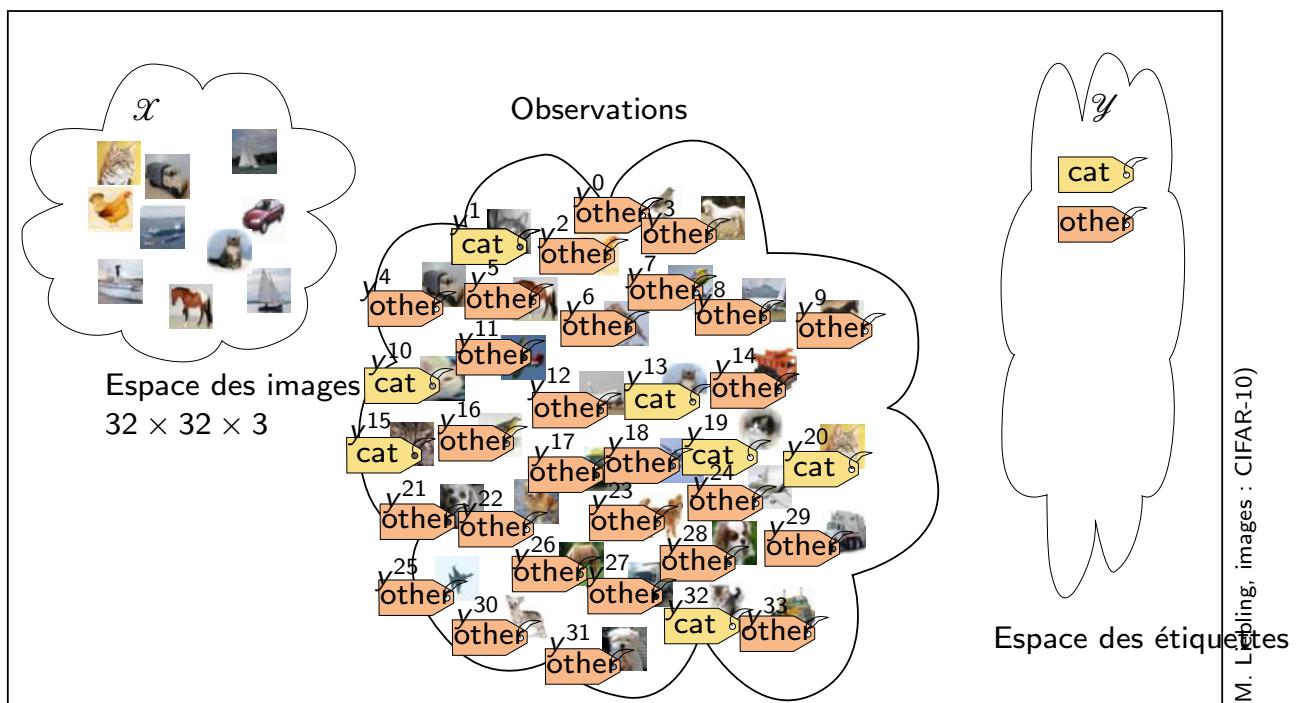


Michael Liebling

EE-311—Apprentissage machine / 1. Introduction

36 / 61

Apprentissage supervisé : étiquettes $\{y^i\}_{i=1,\dots,n}$



Michael Liebling

EE-311—Apprentissage machine / 1. Introduction

37 / 61

Apprentissage supervisé : déterminer un modèle à partir de données

Faire des prédictions à partir d'une liste d'exemples étiquetés

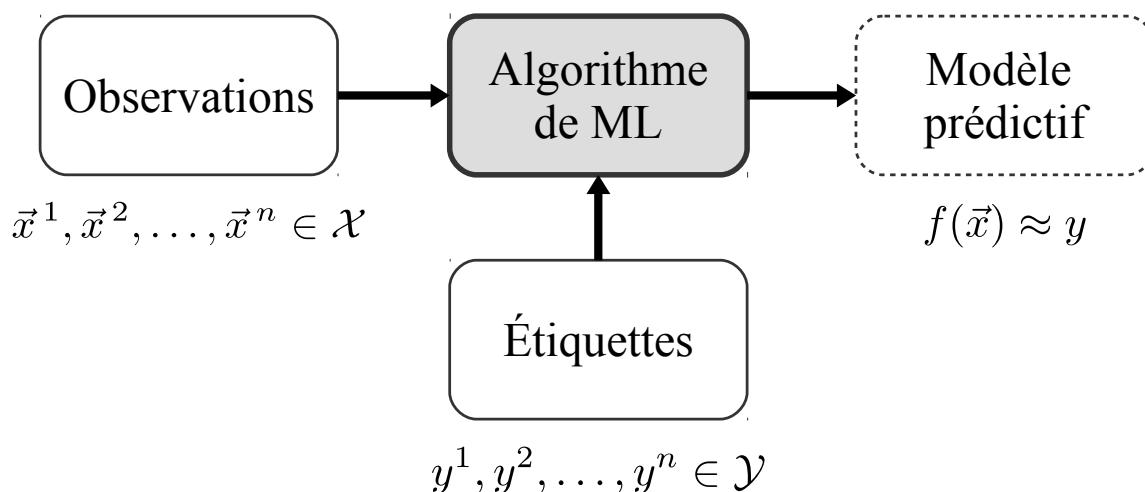


FIGURE 1.1 – Apprentissage supervisé.

Source : Azencott

Michael Liebling

EE-311—Apprentissage machine / 1. Introduction

38 / 61

Le résultat d'un algorithme d'apprentissage est un modèle

Ce modèle (qui prend typiquement la forme d'un algorithme classique qui a beaucoup de paramètres, *apris* durant la phase d'entraînement) peut ensuite être utilisé pour faire des prédictions sur des entrées jusque-là inconnues :

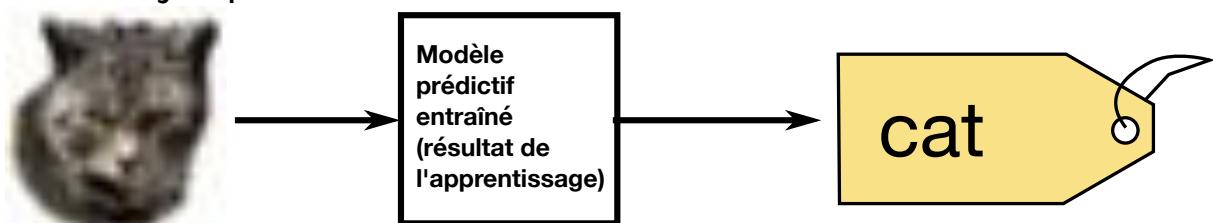


Image inconnue

étiquette prédictive

Note : les données d'entraînement ne sont plus nécessaires pour utiliser le modèle entraîné !

Apprentissage supervisé (définition plus formelle)

Définition (Apprentissage supervisé) : La branche du machine learning qui s'intéresse aux problèmes pouvant être formalisés de la façon suivante :

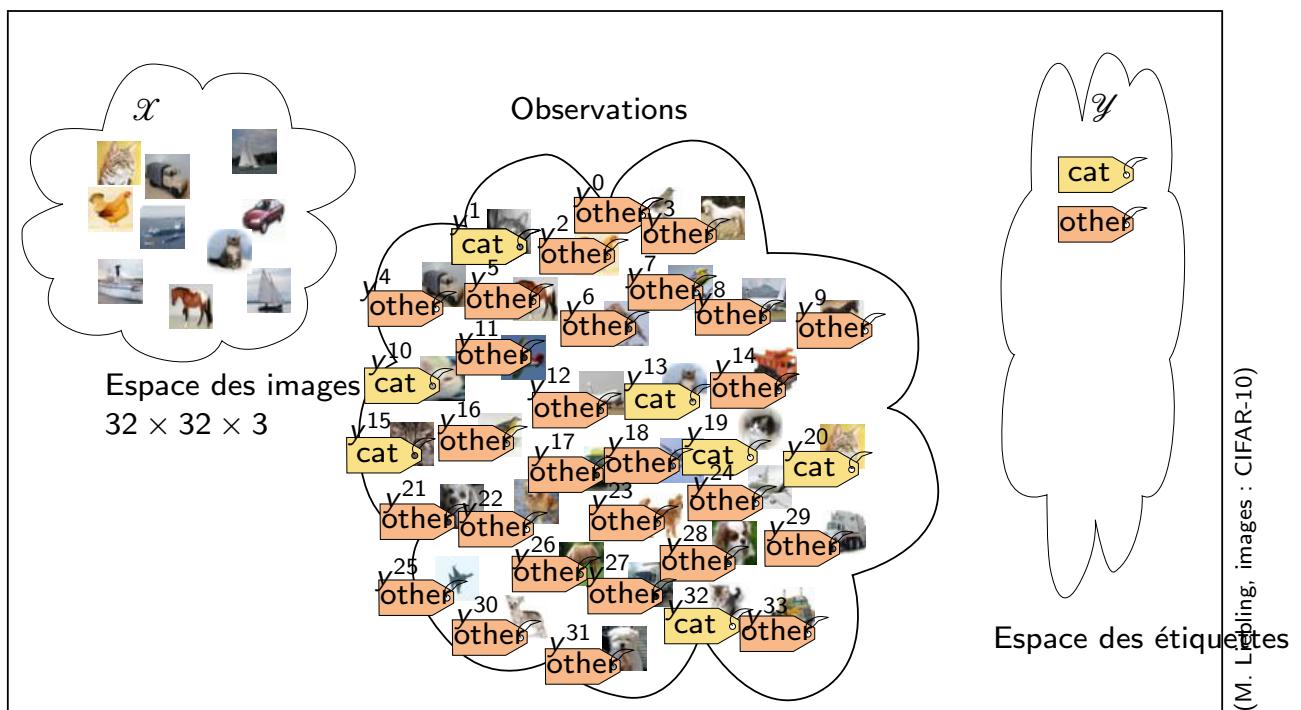
Soient n observations $\{\vec{x}^i\}_{i=1,\dots,n}$ décrites dans un espace \mathcal{X} , et leurs étiquettes $\{y^i\}_{i=1,\dots,n}$ décrites dans un espace \mathcal{Y} , on suppose que les étiquettes peuvent être obtenues à partir des observations grâce à une fonction $\phi : \mathcal{X} \rightarrow \mathcal{Y}$ fixe et inconnue : $y^i = \phi(\vec{x}) + \epsilon_i$, où ϵ_i est un bruit aléatoire. Il s'agit alors d'utiliser les données pour déterminer une fonction $f : \mathcal{X} \rightarrow \mathcal{Y}$ telle que, pour tout couple $(\vec{x}, \phi(\vec{x})) \in \mathcal{X} \times \mathcal{Y}$, on a $f(\vec{x}) \approx \phi(\vec{x})$

Classification binaire

Définition 1.2 (Classification binaire) Un problème d'apprentissage supervisé dans lequel l'espace des étiquettes est binaire, autrement dit $\mathcal{Y} = \{0, 1\}$ est appelé un problème de *classification binaire*.

\vec{x}	\mathcal{Y}
email	spam/not spam
tableau	Peint par Picasso/pas peint par Picasso
image	chat/pas chat

Apprentissage supervisé : étiquetage avec 2 classes

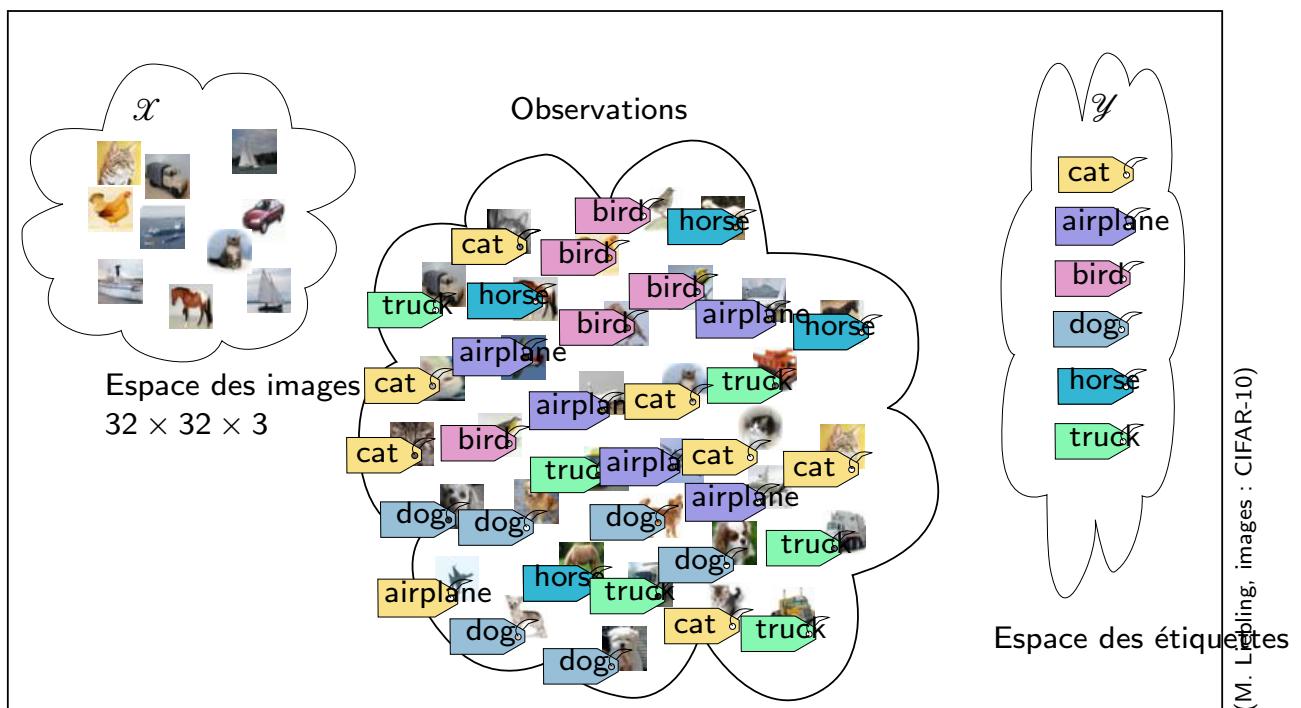


Classification multi-classe

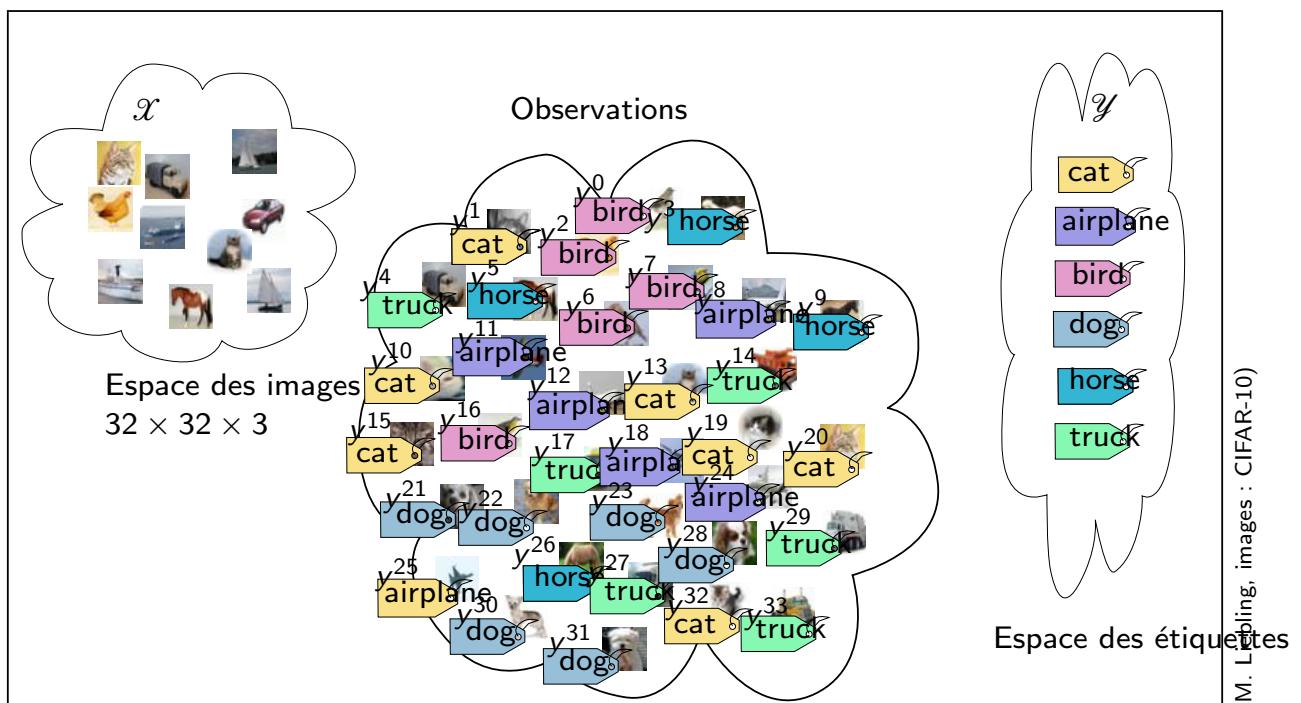
Définition : un problème d'apprentissage supervisé dans lequel l'espace des étiquettes est discret et fini, autrement dit $\mathcal{Y} = \{1, 2, \dots, C\}$ est appelé un problème de *classification multi-classe*. C est le nombre de classes.

\vec{x}	\mathcal{Y}
texte	français, anglais, italien, espagnol, mandarin, etc.
image de chiffre	0, 1, 2, 3, 4, 5, 6, 7, 8, 9
image de chien	labrador, bouvier bernois, chihuahua, etc.

Apprentissage supervisé : étiquetage multi-classe



Apprentissage supervisé : étiquetage multi-classe



Régression

Définition : un problème d'apprentissage supervisé dans lequel l'espace des étiquettes est $\mathcal{Y} = \mathbb{R}$ est appelé un problème de *régession*.

\vec{x}	\mathcal{Y}
date	nombre d'utilisateurs d'un site internet en ligne
$\begin{pmatrix} \text{type de molécule 1} \\ \text{type de molécule 2} \end{pmatrix}$	affinité
$\begin{pmatrix} \text{type de semence (type de blé)} \\ \text{type de sol} \\ \text{ensoleillement} \end{pmatrix}$	rendement [masse/surface]

Apprentissage supervisé : déterminer un modèle à partir de données

Faire des prédictions à partir d'une liste d'exemples étiquetés

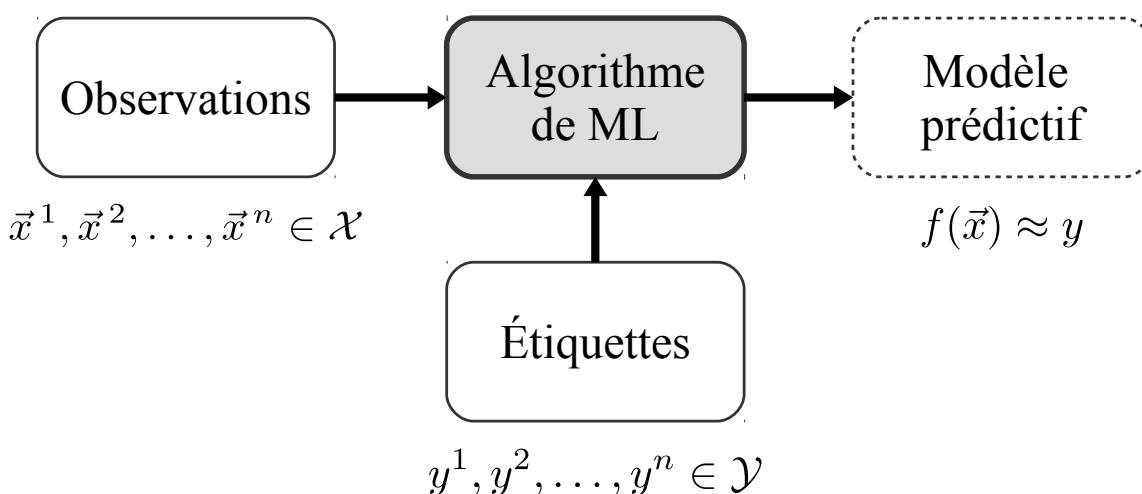


FIGURE 1.1 – Apprentissage supervisé.

Source : Azenctot

Régression linéaire

Modèles paramétriques On parle de modèle paramétrique lorsque le but de l'algorithme d'apprentissage est de trouver les valeurs optimales des paramètres d'un modèle dont on a défini la forme analytique.

Exemple de régression linéaire



Image : <https://actu.epfl.ch/image/76443/1296x728.jpg>

Livres au mètre : On veut entraîner un modèle pour prédire le poids des livres rangés sur une étagère de bibliothèque à partir de la longueur d'étagère qu'ils couvrent. En l'absence de formule physique que relierait la longueur et la masse, on décide de faire une régression linéaire sur la base de n observations $\{x^0, x^1, \dots, x^{n-1}\}$ (on mesure la longueur couverte par les livres dans différents endroits de la bibliothèque) et on détermine aussi les n étiquettes correspondantes $\{y^0, y^1, \dots, y^{n-1}\}$ (on sort les livres pour les étagères où l'on a mesuré la longueur et on les pèse : étiquette = masse).

Exemple de régression linéaire (suite)

Sur la base de ces observations étiquetées, le problème de régression revient à estimer les paramètres d'un modèle qu'il nous faudra postuler.

Régression linéaire : on définit le modèle par la fonction

$$f : x \mapsto \beta_0 + \beta_1 x$$

dont le résultat $f(x)$ pour une nouvelle mesure donnera une estimation de y (qu'on pourra utiliser plutôt que d'utiliser une balance).

Nombre de dimension pour les observations $\vec{x}^i = x^i$: $p = 1$

$$\vec{\beta} \in \mathbb{R}^{p+1}$$

Solution par minimisation de la somme des moindres carrés (on a n observations) :

$$\arg \min_{\vec{\beta} \in \mathbb{R}^{p+1}} \sum_i^n (y^i - \beta_0 - \beta_1 x^i)^2$$

Régression linéaire (suite)

Apprentissage supervisé

on définit :

$$X = \begin{pmatrix} 1 & x^1 \\ \vdots & \vdots \\ 1 & x^n \end{pmatrix}$$

On peut réécrire la somme des résidus quadratiques :

$$\text{RSS} = (\vec{y} - X \vec{\beta})^\top (\vec{y} - X \vec{\beta})$$

On peut minimiser cette forme quadratique en annulant son gradient :

$$\nabla_{\vec{\beta}} \text{RSS} = -2X^\top (\vec{y} - X \vec{\beta})$$

et on obtient :

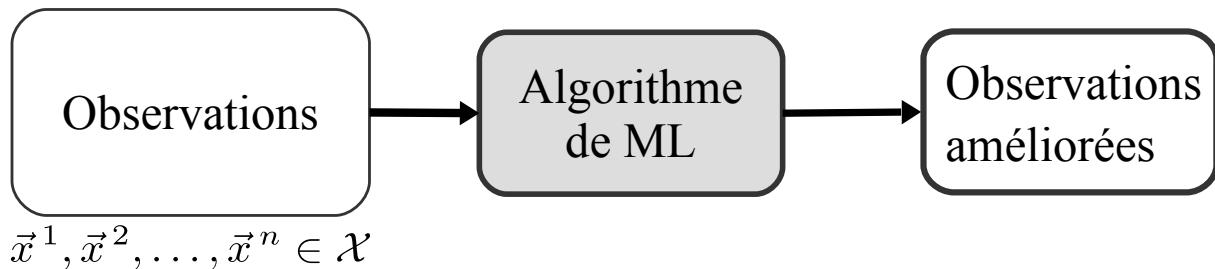
$$X^\top X \vec{\beta}^* = X^\top \vec{y}$$

Si X est de rang 2, on a :

$$\vec{\beta}^* = (X^\top X)^{-1} X^\top \vec{y}$$

Apprentissage non supervisé

Définition : On appelle apprentissage non supervisé la branche du machine learning qui s'intéresse aux problèmes pouvant être formalisés de la façon suivante : étant données n observations $\{\vec{x}^i\}_{i=1,\dots,n}$ décrites dans un espace \mathcal{X} , il s'agit d'apprendre une fonction sur \mathcal{X} qui vérifie certaines propriétés.



Azencott

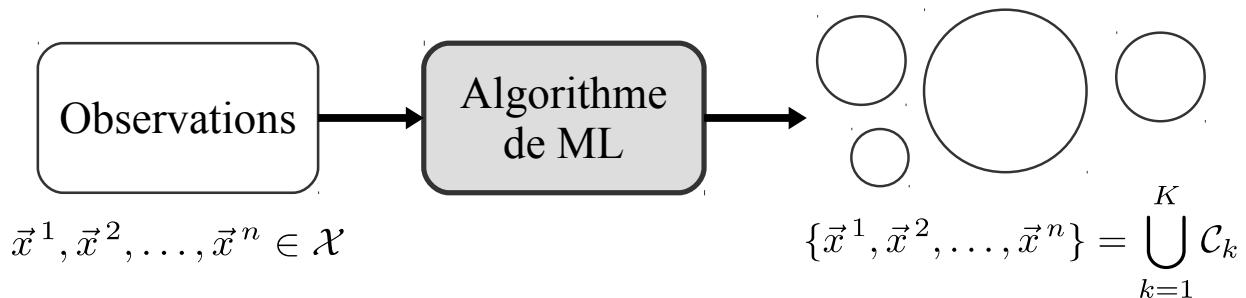
Michael Liebling

EE-311—Apprentissage machine / 1. Introduction

52 / 61

Partitionnement (Clustering)

Définition : on appelle *partitionnement* ou *clustering* un problème d'apprentissage non supervisé pouvant être formalisé comme la recherche d'une partition $\bigcup_{k=1}^K \mathcal{C}_k$ des n observations $\{\vec{x}^i\}_{i=1,\dots,n}$. Cette partition doit être pertinente au vu d'un ou plusieurs critères à préciser.

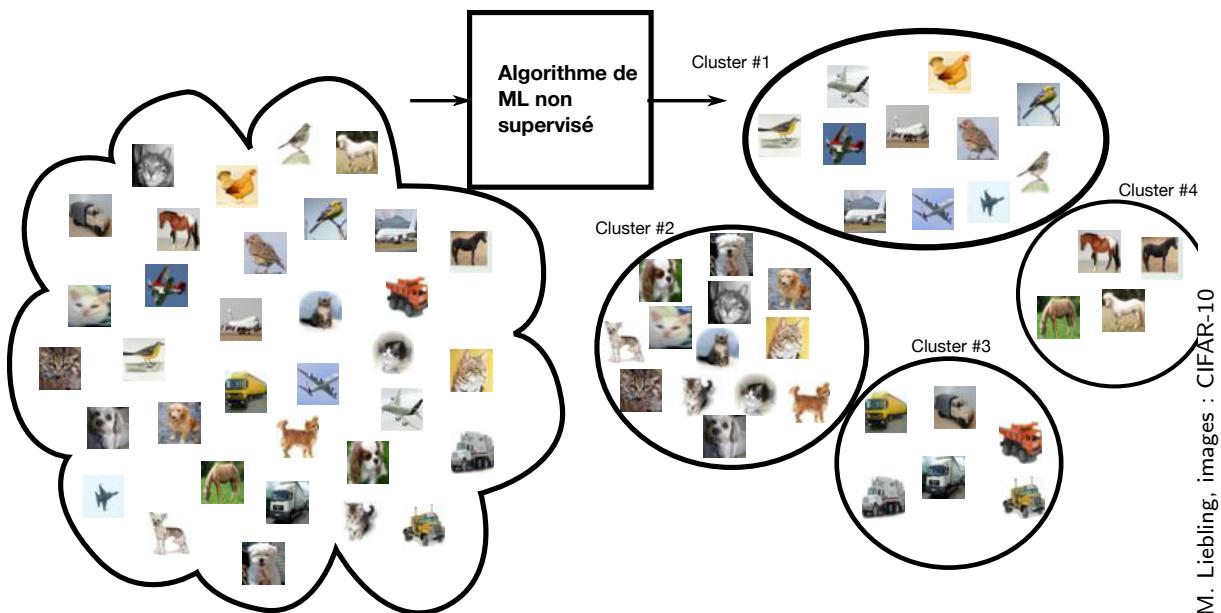


Azencott

FIGURE 1.3 – Partitionnement des données, ou clustering.

Clustering : une méthode non supervisée

Observations sans étiquettes



Note : les clusters obtenus n'ont pas d'étiquette et les groupements peuvent être incongrus...

Michael Liebling

EE-311—Apprentissage machine / 1. Introduction

54 / 61

Réduction de dimension

Définition : on appelle *réduction de dimension* un problème d'apprentissage non supervisé pouvant être formalisé comme la recherche d'un espace \mathcal{Z} de dimension plus faible que l'espace \mathcal{X} dans lequel sont représentés n observations $\{\vec{x}^i\}_{i=1,\dots,n}$. Les projections $\{\vec{z}^i\}_{i=1,\dots,n}$ des données sur \mathcal{Z} doivent vérifier certaines propriétés à préciser.



$$\vec{x}^1, \vec{x}^2, \dots, \vec{x}^n \in \mathcal{X}$$

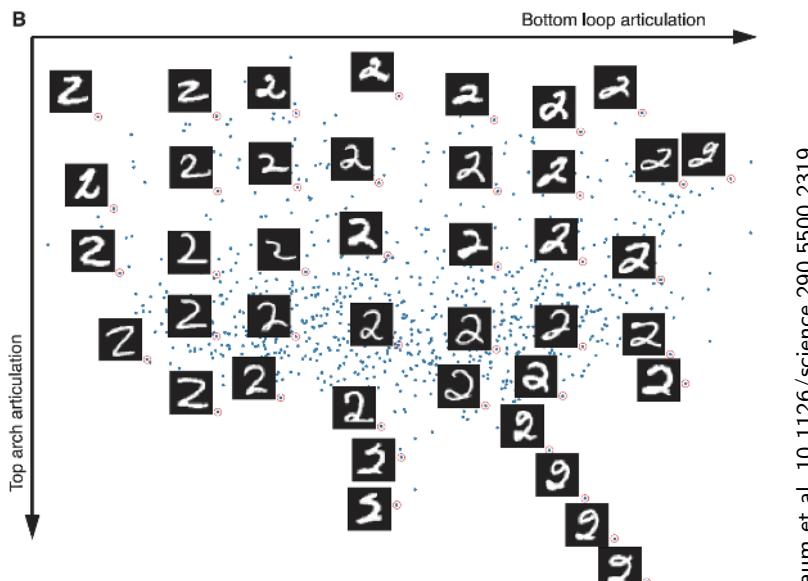
$$\vec{z}^1, \vec{z}^2, \dots, \vec{z}^n \in \mathcal{Z}$$

FIGURE 1.4 – Réduction de dimension.

Exemple de réduction de dimension 784 → 2

- chaque image est de taille 28×28 pixels (jeu de données MNIST)
- chaque images est représentée par un point dans un plan (via algorithme isomap)

dimensionality
istent with
cting from
and repre-
an extra



Tenenbaum et al. 10.1126/science.290.5500.2319

22 DECEMBER 2000 VOL 290 SCIENCE www.sciencemag.org

Michael Liebling

EE-311—Apprentissage machine / 1. Introduction

56 / 61

Exemple de réduction de dimension 3 → 2

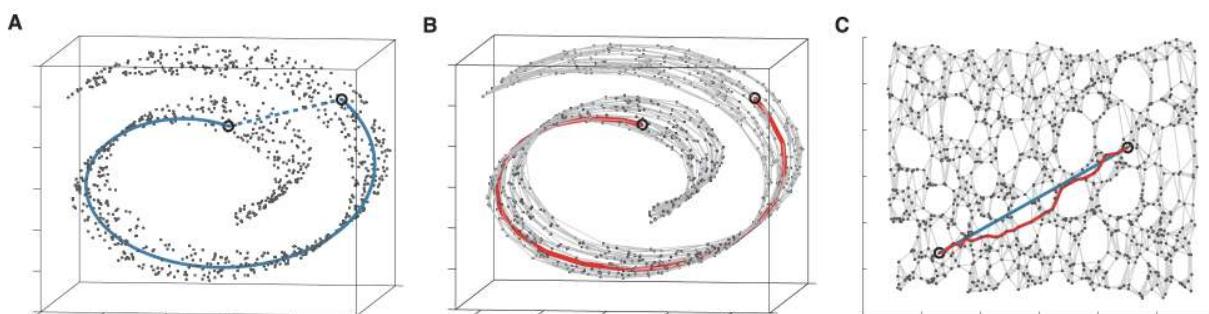


Fig. 3. The "Swiss roll" data set, illustrating how Isomap exploits geodesic paths for nonlinear dimensionality reduction. (A) For two arbitrary points (circled) on a nonlinear manifold, their Euclidean distance in the high-dimensional input space (length of dashed line) may not accurately reflect their intrinsic similarity, as measured by geodesic distance along the low-dimensional manifold (length of solid curve). (B) The neighborhood G constructed in step one of Isomap (with $K = 7$ and $N =$

1000 data points) allows an approximation (red segments) to the true geodesic path to be computed efficiently in step two, as the shortest path in G . (C) The two-dimensional embedding recovered by Isomap in step three, which best preserves the shortest path distances in the neighborhood graph (overlaid). Straight lines in the embedding (blue) now represent simpler and cleaner approximations to the true geodesic paths than do the corresponding graph paths (red).

www.sciencemag.org SCIENCE VOL 290 22 DECEMBER 2000

2321

Tenenbaum et al. 10.1126/science.290.5500.2319

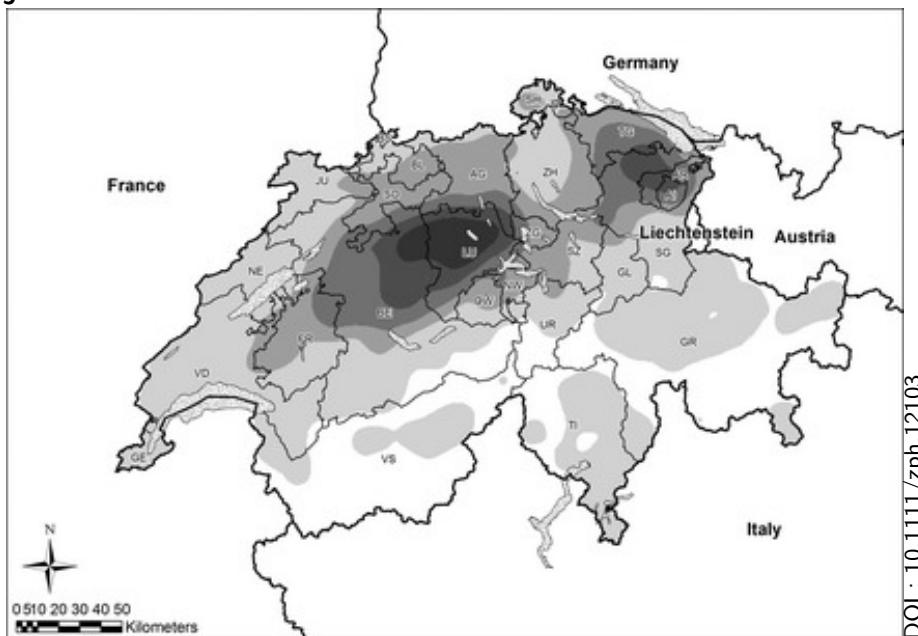
Michael Liebling

EE-311—Apprentissage machine / 1. Introduction

57 / 61

Estimation de densité

Problème d'estimation d'une loi de probabilité en supposant que le jeu de donnés en est un échantillon aléatoire.



Density estimation for pig farms (per km²) in Switzerland based on the agricultural census 2010 (Federal Office for Agriculture, Switzerland). Progressing from light grey to black, the classification is 0.001–0.15 ; 0.15–0.45 ; 0.45–0.9 ; 0.9–1.5 ; 1.5–2.42 pig farms per km².

Michael Liebling

EE-311—Apprentissage machine / 1. Introduction

58 / 61

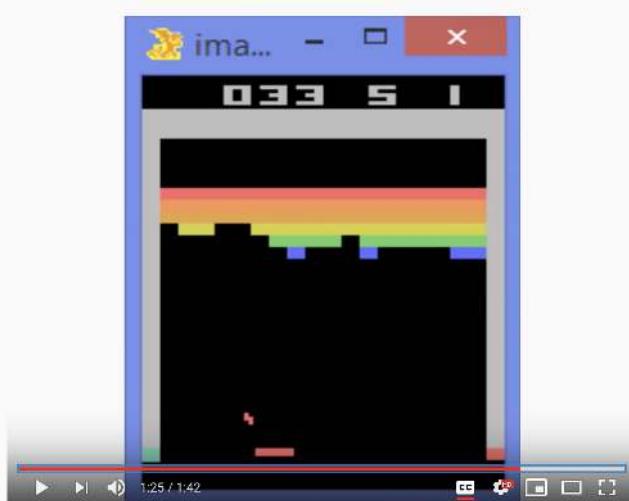
Autres types d'apprentissage

Apprentissage semi-supervisé

apprentissage à partir d'un jeu de données partiellement étiqueté

Apprentissage par renforcement

- système d'apprentissage interagit avec son environnement et accomplit des actions
- actions résultent en une récompense (par exemple +1, -1, 0)



Google DeepMind's Deep Q-learning playing Atari Breakout

Résumé du cours 1

- information sur le cours :
<https://moodle.epfl.ch/course/view.php?id=16090>
- utiliser Ed pour les communications
- formation de groupes de travail (informels)
- labos de programmation en Python3 sur la plateforme JupyterHub jupyterhub.noto.epfl.ch
- algorithme d'apprentissage automatique : algorithme qui apprend un modèle à partir d'exemples
- différencier supervisé (exemples annotés) et non-supervisé (pas d'annotation des données)
- type principaux de problèmes d'apprentissage (classification, régression, clustering, estimation de densité)
- Régression linéaire, solution via minimisation de la somme des moindres carrés

Guide de lecture pour ce cours

Chloé-Agathe Azencott “Introduction au Machine Learning”,
Dunod, 2019, ISBN 978-210-080153-4
Chapitre 1 : Présentation du machine learning