

Apprentissage par Adaptation de Domaine

Emilie Morvant

Laboratoire Hubert Curien
Université Jean Monnet



22 janvier 2016



Qui suis-je ?

- Maître de Conférence
 - à l'Université Jean Monnet, St-Etienne, France
 - au Laboratoire Hubert Curien
 - dans l'équipe d'apprentissage automatique
 - responsable de la Licence 2 d'informatique
- Domaine de recherche :
 - Domaine principal : Apprentissage Automatique (*Machine Learning*)
 - Théorie de l'apprentissage automatique statistique
 - Applications à des pbs de vision par ordinateur (*Computer Vision*)
- Comment me joindre :
 - Mail : emilie.morvant@univ-st-etienne.fr
 - Twitter : @EmilieMorvant
- Merci à R. Emonet et A. Habrard

Qui utilise l'apprentissage automatique ?

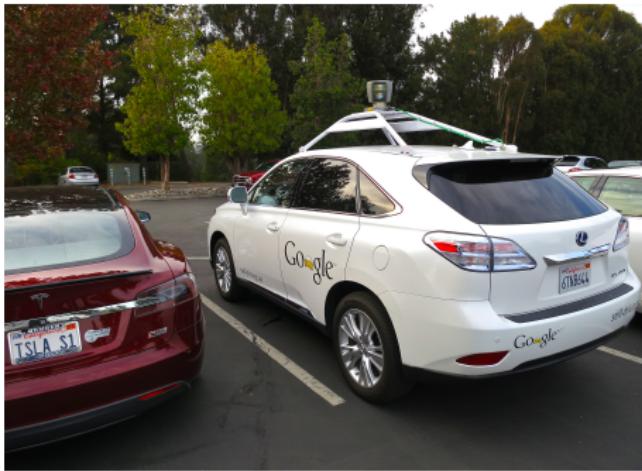
Google



NETFLIX

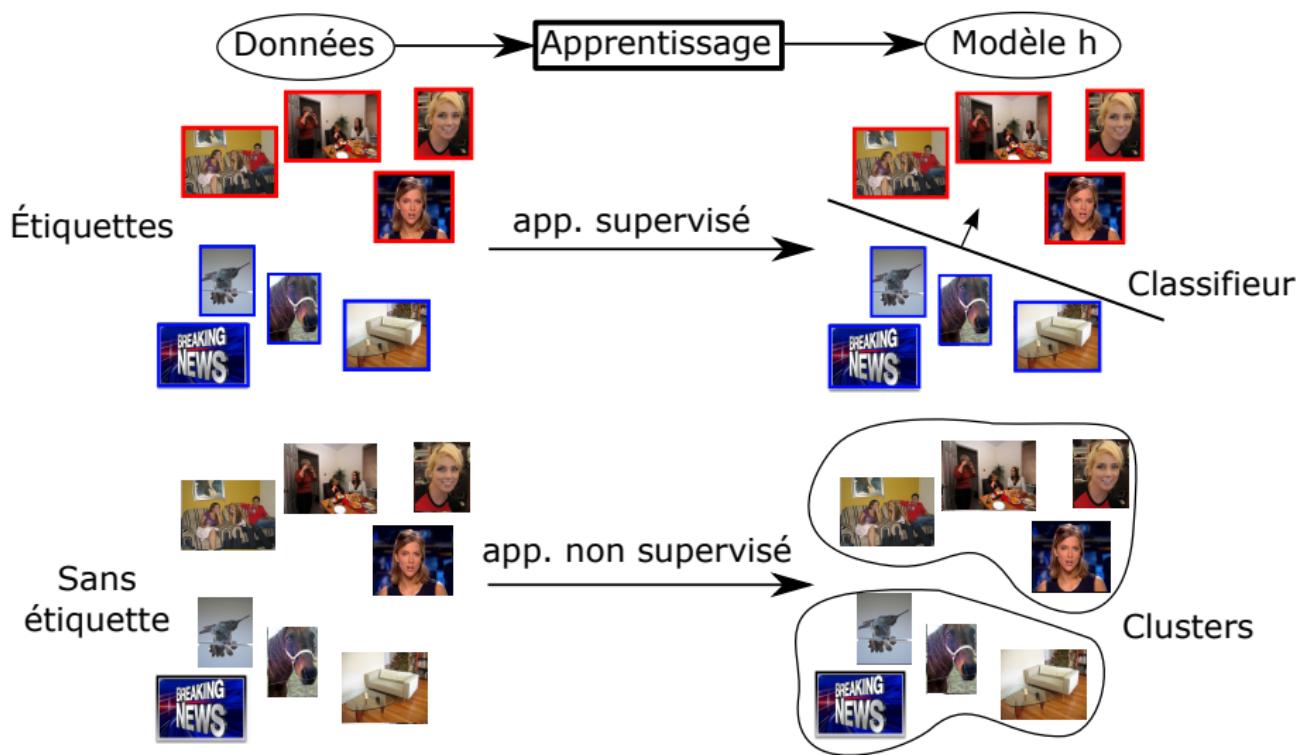
You Tube

Et le futur de l'apprentissage automatique ?



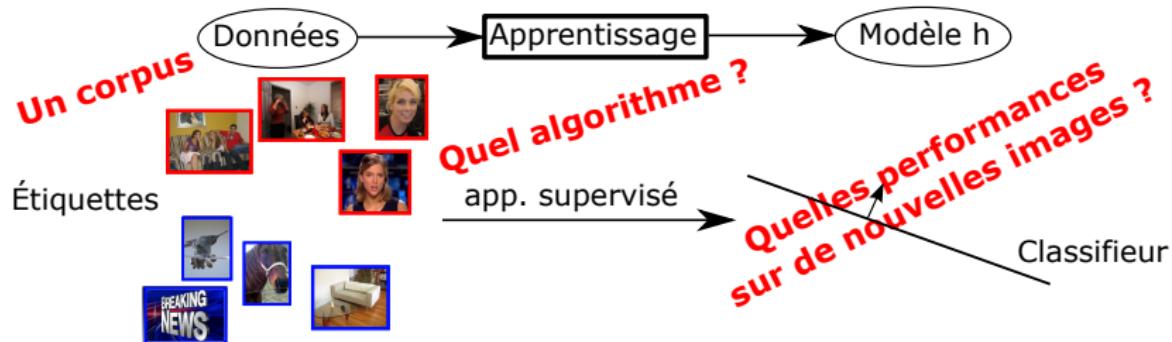
L'Apprentissage automatique, c'est quoi ?

Tâche : Y-a-t-il une Personne dans l'image/la vidéo ?



Comment garantir de bonnes performances ?

Tâche : Y-a-t-il une Personne dans l'image/la vidéo ?



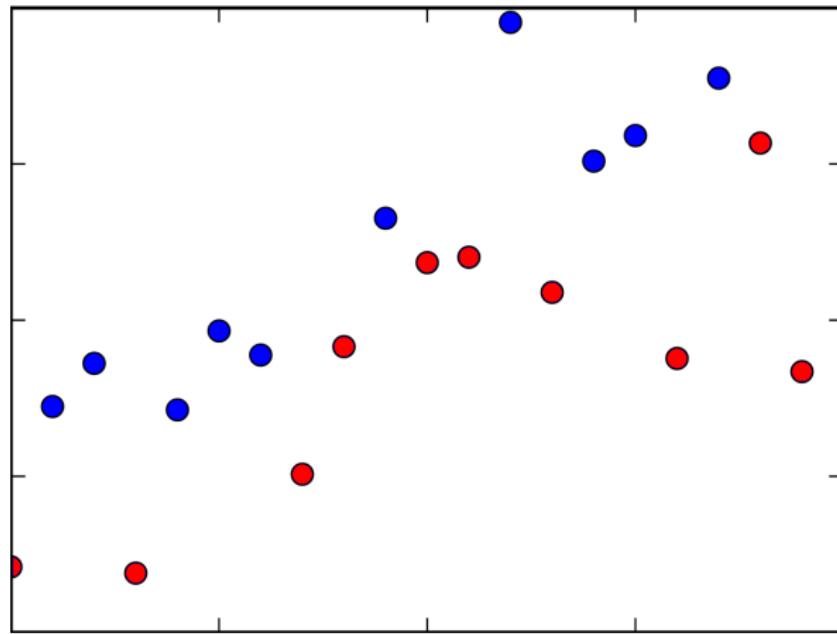
Comment apprendre le classifier $h : X \rightarrow Y$ pour qu'il se trompe le moins possible sur de nouvelles images ?

Solution : minimiser l'erreur empirique mesurée sur les données d'apprentissage

⇒ Requiert des garanties ⇒ Borne en généralisation

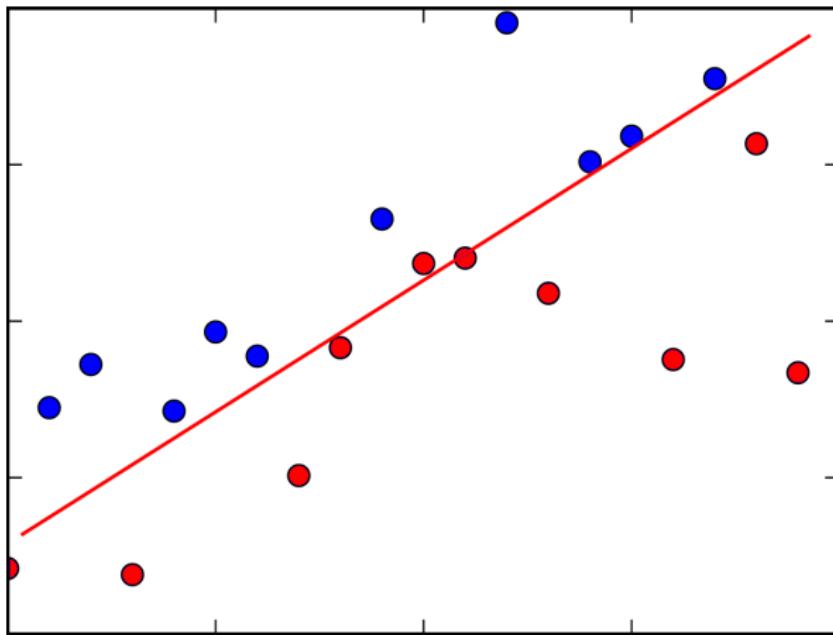
$$\text{erreur réelle} \leq \text{erreur empirique} + f(\text{complexité, nb de données})$$

Pourquoi la “complexité” du modèle ?



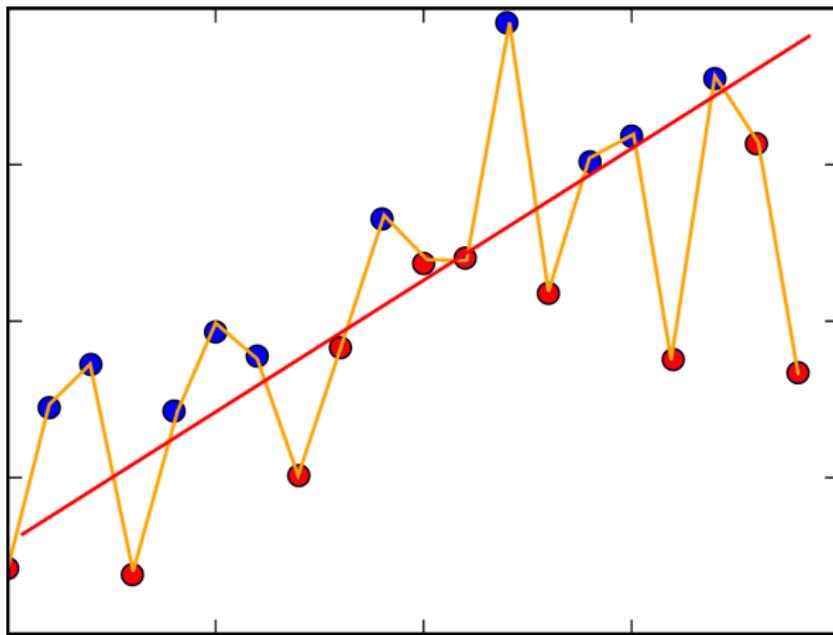
Graphiques de R. Emonet

Pourquoi la “complexité” du modèle ?



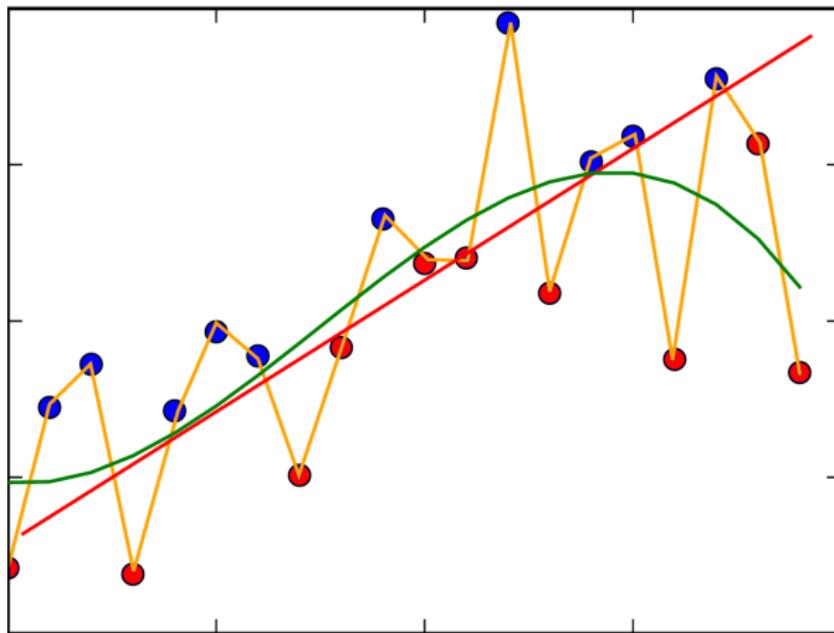
Graphiques de R. Emonet

Pourquoi la “complexité” du modèle ?



Graphiques de R. Emonet

Pourquoi la “complexité” du modèle ?



Graphiques de R. Emonet

Et l'adaptation de domaine dans tout ça ?

Quand a-t-on besoin d'adaptation de domaine (DA) ?

Lorsque l'on veut adapter un modèle d'une tâche **source** vers une tâche **cible**

Exemple



Personne pas de Personne

Corpus de **Photos** étiquetées

Domaine **source**

Y-a-t'il une Personne ?

Corpus de **Videos** non étiquetées

Domaine **cible**

Problématique

Comment apprendre, à partir d'une distribution,
un classifieur performant sur une distribution différente ?

Un autre exemple d'adaptation de domaine



critiques de livres

??? The end of the series.

This book was written to provoke those who wanted Adams to continue the trilogy but I loved it. Author settled down on a bob fearing planet where he has acquired the prestigious...

[Read more](#)

Published on Mar 19 2002 by dan

??? Mostly Harmless is Underrated

I think most of the reviews for this book downplay it seriously. While the ending is kind of disappointing, the book overall is wonderful.

[Read more](#)

Published on Jan 22 2002 by A Big Adams Fan

??? Please pretend this book was never written.

I have long been a fan of the Hitchhiker's series as they are comic genius. The book Mostly Harmless, however, should never have come about. It is frustration at its peak. [Read more](#)

Published on Jan 14 2002 by Paul Nomad

??? Kinda like horror movies...

...in that the last one usually isn't all that appealing. I liked it fine, with some of Adams's wit, but it was a bit disappointing. [Read more](#)

Published on Nov 4 2001 by Kristopher Vincent

??? A Terrible End to A Great Series

The ending for this book was so bad that I vowed never to read another Douglas Adams book. Adams was obviously sick and tired of the series and used this book to kill it off with...

[Read more](#)

Published on Oct 17 2001 by David A. Lessau



critiques de film

-1 An insult to Douglas Adams'

memory

I agree entirely with "darkgenius" comments. This movie is a travesty of the book and the TV series; a cutesy version totally lacking in the wit and satire of the original. [Read more](#)

Published 5 months ago by John W Beare

+1 Don't Panic!

If you haven't listened to the BBC radio-play, this isn't bad! Purists, no doubt, will dispute my verdict but the fact of the matter is THGTTG (see title) does have Douglas Adams'...

[Read more](#)

Published on Mar 13 2011 by Sid Matheson

+1 On Blu-ray, even better

I've seen this movie on TV and wanted to add it to my collection. I couldn't find it locally so when I saw it on Amazon and on Blu-ray, I picked it up. [Read more](#)

Published on April 18 2009 by J. W. Life

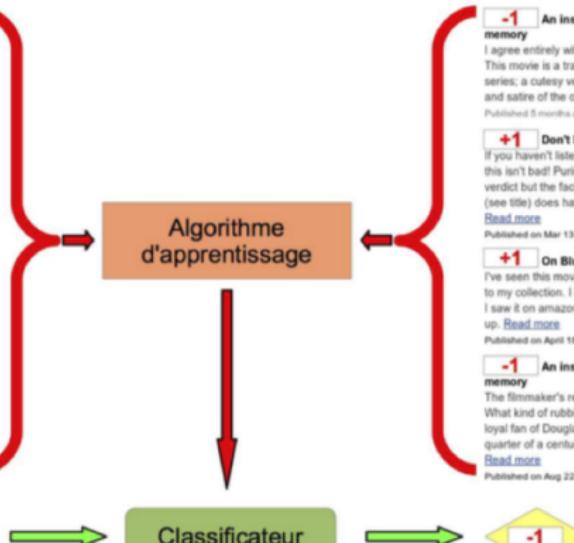
-1 An insult to Douglas Adams'

memory

The filmmaker's reverence for Adams' legacy? What kind of rubbish statement is that? As a loyal fan of Douglas Adams for more than a quarter of a century, I was appalled and...

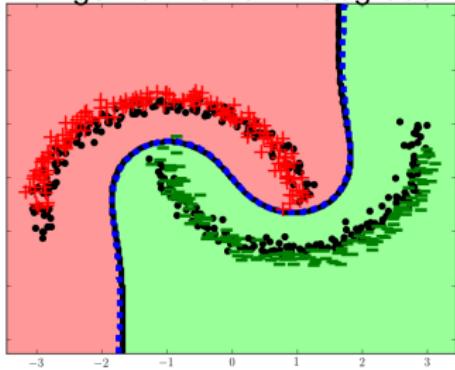
[Read more](#)

Published on Aug 22 2006 by Daniel Jolley

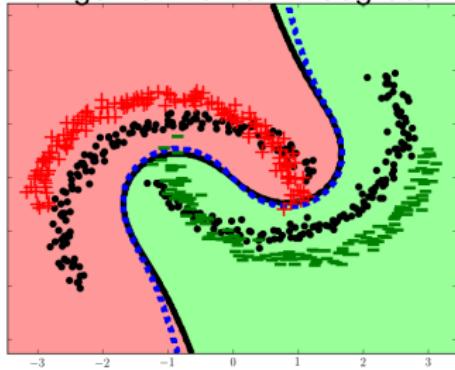


Un exemple jouet d'adaptation de domaine

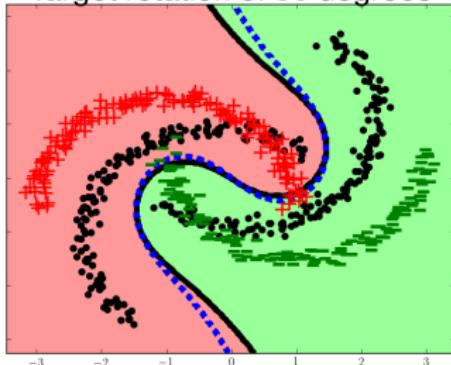
Target rotation of 10 degrees



Target rotation of 30 degrees

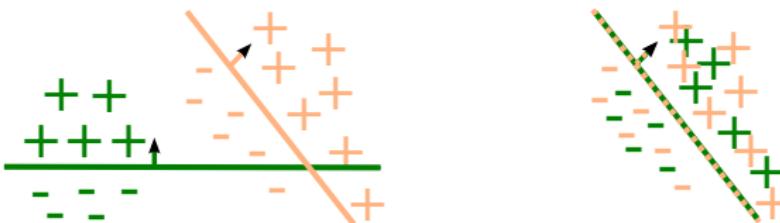


Target rotation of 50 degrees



Quelques problématiques liées à l'adaptation

- Estimation de la “distance” entre les distributions



- Dériver des garanties théorique via des bornes en généralisation

$$\text{erreur cible} \leq \text{erreur source} + \dots$$

- Caractériser si et quand l'adaptation est possible

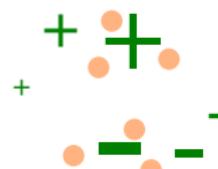


- Définir des algorithmes \Rightarrow Idée : Rapprocher les deux distributions

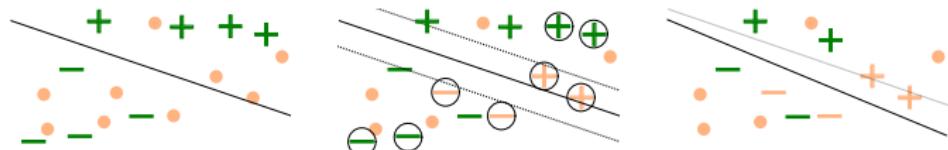
Quelques algorithmes d'adaptation de domaine

Trois grands types d'approches

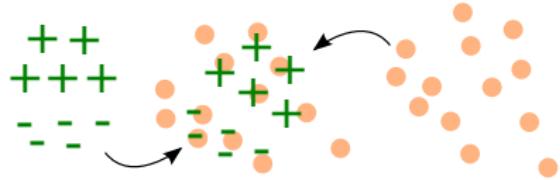
Repondérations des données



Auto-étiquetage



Recherche d'un espace de représentation



Un petit peu de théorie avec les mains... .

Une philosophie classique

$$\text{erreur cible} \leq \text{erreur source} + \text{Distance}(\text{source}, \text{cible}) + \lambda_{\text{etiquetage}}$$

Si les tâches sont reliées selon l'étiquetage $\Leftrightarrow \lambda_{\text{etiquetage}}$ est petit

Alors une solution pour adapter est de trouver un bon compris entre

- “rapprocher” les tâches selon $\text{Distance}(\text{source}, \text{cible})$
- commettre peu d'erreur pour la tâche source

Mais à quoi ressemble la théorie ?... [Germain et al. ICML'13]

Borne en généralisation PAC-Bayésienne pour l'adaptation de domaine

Pour toute distributions P_S et P_T sur $X \times Y$, pour tout ensemble de votants \mathcal{H} , pour toute distribution a priori π sur \mathcal{H} , pour tout $\delta \in (0, 1]$, pour tout $a > 0$ et $c > 0$, avec une probabilité d'au moins $1 - \delta$ sur le choix aléatoire de $S \times T \sim (P_S \times D_T)^m$, on a pour toute distribution a posteriori sur \mathcal{H} :

$$R_{P_T}(G_\rho) \leq c'R_S(G_\rho) + a' \text{dis}_\rho(S, T) + \left(\frac{c'}{c} + \frac{2a'}{a} \right) \frac{\text{KL}(\rho \parallel \pi) + \ln \frac{3}{\delta}}{m} + \nu_\rho + a' - 1$$

$$\text{où } \nu_\rho = R_{D_T}(G_\rho, G_{\rho_T^*}) + R_{D_S}(G_\rho, G_{\rho_T^*}) + R_{P_T}(G_{\rho_T^*}), \text{ } c' = \frac{c}{1-e^{-c}}, \text{ et } a' = \frac{2a}{1-e^{-2a}}$$

Comment dériver un algorithme ?

- En spécialisant la borne à des modèles particuliers (classificateurs linéaires)
- On obtient un algorithme qui vise à minimiser la partie droite de la borne

$$\min_w C R_S(G_w) + A \text{dis}_{\rho_w}(S, T) + \frac{1}{2} \|w\|^2$$

Take home message

L'adaptation de domaine

L'adaptation de domaine est un champs de recherche de l'apprentissage automatique et plus précisément de l'apprentissage par transfert

Objectif

Adapter un système d'apprentissage d'une tâche source pour laquelle on a des informations vers une tâche cible pour laquelle on a peu d'informations

Ce que dit la théorie

Si les deux tâches sont reliées alors on peut adapter en développant des algorithmes permettant de rapprocher les tâches tout en gardant de bonnes performances pour la tâche source

Mais que fait-on si l'on a plus de deux tâches ?

À partir de plusieurs tâches on peut apprendre

- un modèle pour une seule tâche cible

↪ **Adaptation de domaine multi-sources**

- un modèle pour plusieurs tâches cibles

↪ **Adaptation de domaine multi-cibles**

- un modèle pour TOUTES les tâches

↪ **Apprentissage multi-tâches**

Comme pour l'adaptation de domaine,
on désire tirer avantage des similarités entre les tâches !

Merci de votre attention.

@EmilieMorvant