Optimisation pour l'apprentissage automatique

Séance 1 : Introduction

Clément W. Royer

Master 2 Big Data

23/11/2021

About the instructors

Clément Royer

- Université Paris Dauphine-PSL;
- Séances de cours;
- clement.royer@dauphine.psl.eu

Radhia Bessi

- ENIT;
- Séances de TD/TP;
- radhia.bessi@enit.utm.tn.

Liens utiles

Page du cours

- https://www.lamsade.dauphine.fr/∼ croyer/coursTUN.html
- Notes de cours (mises à jour régulièrement) : https://www.lamsade.dauphine.fr/ croyer/ensdocs/TUN/Poly-TUN.pdf.

- Séances via Teams :
  - 1h30 cours (C. Royer);
  - 1h30 TD/TP (R. Bessi).
- Tous les supports sur la page web du cours.

À propos du cours

Un tour d'horizon

- Concepts de base en optimisation...
- ...importants dans le cadre de l'apprentissage.

Objectifs du cours

- Formaliser un problème d'optimisation;
- Reconnaître des classes de problèmes spécifiques, notamment ceux présents en sciences des données;
- Avoir une boîte à outils algorithmiques.

Plan (provisoire) du cours

- Introduction et bases de l'optimisation;
- Optimisation différentiable/Descente de gradient;
- Optimisation convexe/Accélération;
- Optimisation non lisse/Régularisation;
- Optimisation stochastique/Gradient stochastique;
- Optimisation avec contraintes/distribuée;
- ...

Sommaire

Vocabulaire

- Sciences des données (data science);
- Analyse de données (data analysis);
- Fouille de données (data mining);
- Apprentissage machine/profond (machine/deep learning);
- Intelligence artificielle (AI);
- Big Data;
- ...

Vocabulaire

- Sciences des données (data science);
- Analyse de données (data analysis);
- Fouille de données (data mining);
- Apprentissage machine/profond (machine/deep learning);
- Intelligence artificielle (AI);
- Big Data;
- ...

Ce dont nous allons parler

- Optimisation pour la science des données en général;
- Principes génériques.

L'apprentissage pour ce cours

Un ensemble de problèmes basés sur des données

- Extraction d'information à partir de la donnée : *statistiques*, *attributs principaux*, *structures*;
- Utilisation de cette information pour la prédiction du comportement de données futures.

L'apprentissage pour ce cours

Un ensemble de problèmes basés sur des données

- Extraction d'information à partir de la donnée : *statistiques*, *attributs principaux*, *structures*;
- Utilisation de cette information pour la prédiction du comportement de données futures.

Composantes de l'IA/de la science des données

- Statistiques;
- Informatique (gestion de la donnée, calcul parallèle, etc);
- **Optimisation** pour la modélisation des problèmes et leur résolution par des algorithmes.

Avertissement

Optimisation $\not\subset$ Apprentissage

- L'optimisation est un outil mathématique;
- Appliqué en économie, chimie, physique, sciences sociales...
- Utilisé dans d'autres branches des mathématiques appliquées : algèbre linéaire, EDPs, statistiques.

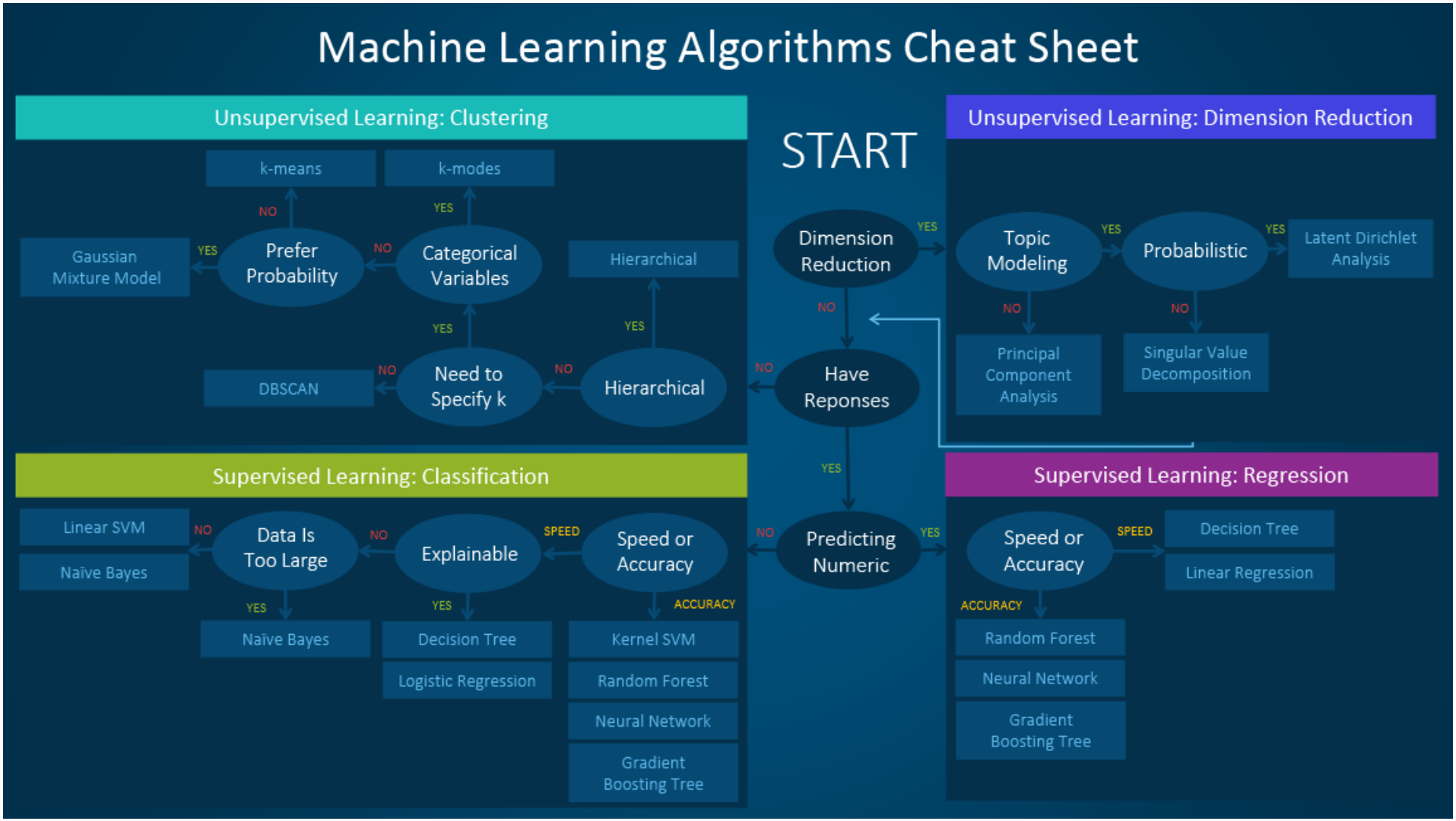
Avertissement

Optimisation $\not\subset$ Apprentissage

- L'optimisation est un outil mathématique;
- Appliqué en économie, chimie, physique, sciences sociales...
- Utilisé dans d'autres branches des mathématiques appliquées : algèbre linéaire, EDPs, statistiques.

Apprentissage $\not\subset$ Optimisation

- L'optimisation s'applique à un certain problème;
- Certaines dimensions de l'apprentissage (nettoyage de la donnée, *hardware*,...) sont orthogonales à l'optimisation.



Source: https://blogs.sas.com/content/subconsciousmusings/2017/04/12/ machine-learning-algorithm-use/

Historique

Optimisation numérique

- Montée en puissance en 1970-1980;
- Succès des algorithmes en ingénierie (chimique, aéronautique, etc).
- *Pratique standard en calcul scientifique:* utiliser une méthode de points intérieurs (basée sur Newton, développée dans les années 2000s).

Optimisation pour l'IA

- Problèmes basés sur de grands volumes de données;
- Les méthodes standard en optimisation ne sont pas les plus efficaces!

Pratique classique en IA: Utiliser une approche de gradient stochastique avec momentum (1950s + article théorique de 1983).

Ce qui a changé

Contexte de données massives/Big Data

- Les calculs usuels (fonction, dérivées) sont très coûteux car ils accèdent à **toute la donnée**.
- La précision souhaitée n'est pas forcément très grande en raison du bruit sur les données.

Ce qui a changé

Contexte de données massives/Big Data

- Les calculs usuels (fonction, dérivées) sont très coûteux car ils accèdent à **toute la donnée**.
- La précision souhaitée n'est pas forcément très grande en raison du bruit sur les données.

Communauté de l'optimisation pour l'apprentissage

- Le problème d'optimisation est souvent un moyen plus qu'une fin;
- Propriétés statistiques des solutions;
- Théorie et pratique différentes de la communauté d'optimisation "classique".

Sommaire

Un problème d'apprentissage supervisé

Au départ : Jeu de données $\{(\boldsymbol{x}_1, y_1), \ldots, (\boldsymbol{x}_n, y_n)\}$.

- $\boldsymbol{x}_i$ est un vecteur d'attributs dans $\mathbb{R}^d$;
- y_i est un label binaire égal à 1 ou -1.

Un problème d'apprentissage supervisé

Au départ : Jeu de données $\{(\boldsymbol{x}_1, y_1), \ldots, (\boldsymbol{x}_n, y_n)\}$.

- $\boldsymbol{x}_i$ est un vecteur d'attributs dans $\mathbb{R}^d$;
- y_i est un label binaire égal à 1 ou -1.

Exemple : classification de documents

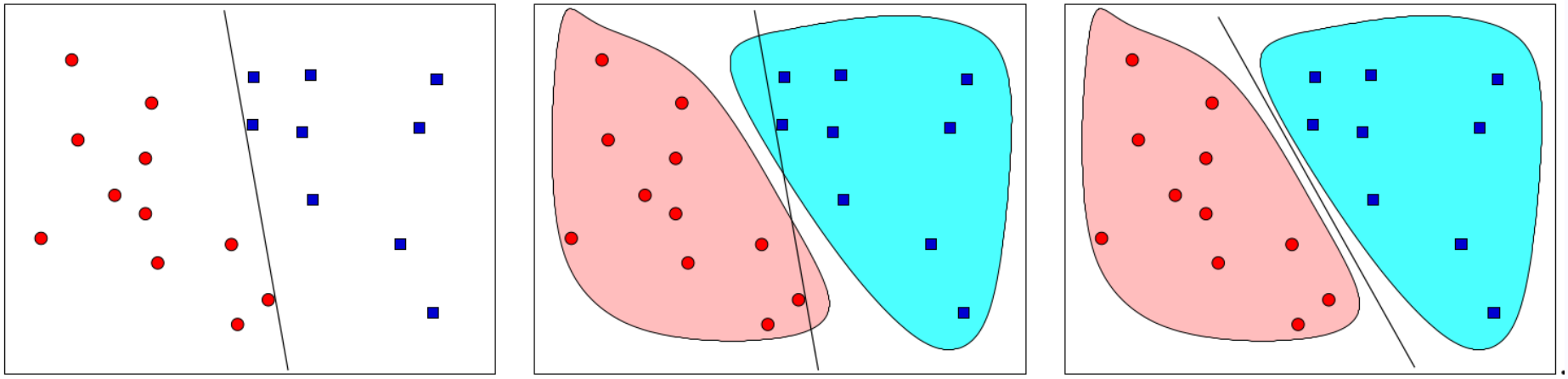
Soit un dictionnaire de d mots.

- $\boldsymbol{x}_i$ représente les mots contenus dans un document :

$$[\boldsymbol{x}_i]_j = \begin{cases} 1 & \text{si le mot } j \text{ est dans le document } i, \\ 0 & \text{sinon.} \end{cases}$$

- y_i égal à $+1$ si le document traite d'un certain sujet, égal à -1 sinon.

Différentes solutions



Source : S. J. Wright, Optimization Algorithms for Data Analysis, 2018.

- Points : $\boldsymbol{x}_i$, rouges/bleus : $y_i = 1/y_i = -1$;
- Nuages rouges/bleus : distribution des documents;
- Deux techniques de classification linéaires;
- Figure de droite : solution à marge maximale (SVM).

Deux points de vue sur le problème

Optimisation

- Le problème peut être modélisé comme un programme quadratique convexe, et résolu de manière efficace (nous verrons comment !);
- Toute solution qui sépare les données est valable.

Deux points de vue sur le problème

Optimisation

- Le problème peut être modélisé comme un programme quadratique convexe, et résolu de manière efficace (nous verrons comment !);
- Toute solution qui sépare les données est valable.

Apprentissage

- Le modèle doit s'appliquer à tous les documents suivant la même distribution que les $\{(\boldsymbol{x}_i, y_i)\} \Rightarrow$ généralisation;
- Les données peuvent être bruitées $\Rightarrow$ garanties statistiques sur les solutions.

Bilan de l'exemple

- Possible de définir des problèmes d'optimisation basés sur des données et de les résoudre efficacement;
- Potentiellement décorrélé du but originel : trouver un modèle sur la distribution des données.

Bilan de l'exemple

- Possible de définir des problèmes d'optimisation basés sur des données et de les résoudre efficacement;
- Potentiellement décorrélé du but originel : trouver un modèle sur la distribution des données.

Autres problématiques

- Quantité massive d'attributs *(tous les mots du dictionnaire)* ?
- Quantité massive de données *(articles Wikipedia)* ?
- Classification impossible par modèles linéaires ?

Bilan de l'exemple

- Possible de définir des problèmes d'optimisation basés sur des données et de les résoudre efficacement;
- Potentiellement décorrélé du but originel : trouver un modèle sur la distribution des données.

Autres problématiques

- Quantité massive d'attributs *(tous les mots du dictionnaire)* ?
  Réduction de dimension, recherche de parcimonie.
- Quantité massive de données *(articles Wikipedia)* ?
  Algorithmes stochastiques.
- Classification impossible par modèles linéaires ?
  Optimisation non linéaire.

Dans ce cours

Du point de vue apprentissage

- Exemples typiques de problèmes d'optimisation;
- Spécificités des formulations issues de l'apprentissage (somme finie, régularisation).

Du point de vue optimisation

- Présentation des algorithmes principaux;
- Étude théorique et illustrations pratiques.

La suite pour aujourd'hui

Cours

- Définition d'un problème d'optimisation;
- Conditions d'optimalité et convexité.

Séance de TD

- Optimalité;
- Convexité.