LU3I026 - Science des données

Apprentissage et modèles

Olivier Schwander <olivier.schwander@sorbonne-universite.fr>

Sorbonne Université

2024-2025

Types d'apprentissage

Apprentissage supervisé

- Données
- Étiquettes
- Exemple: des images, et le nom de l'objet dans l'image
- Difficulté: obtenir les étiquettes...

Apprentissage non-supervisé

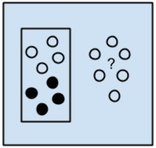
- Juste les données
- Pas d'étiquettes

Autres

- Semi-supervisé
- Renforcement
- Auto-supervisé

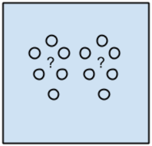
Types d'apprentissage

Supervisé



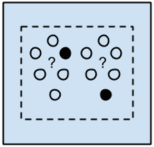
Supervised Learning Algorithms

Non-supervisé



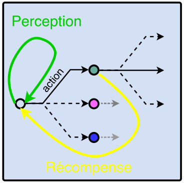
Unsupervised Learning Algorithms

Semi-supervisé



Semi-supervised Learning Algorithms

Renforcement

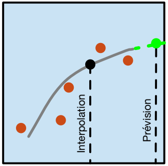


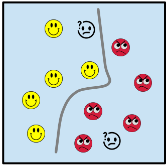
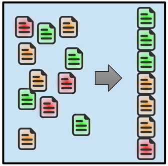
Première chose à déterminer

- conditionne les choix d'algorithmes et de méthodes
- et les mesures de performances

Tâches d'apprentissage

Régression | Classification | Ordonnancement



Autres tâches

- Souvent des cas particuliers des tâches précédentes
- Image: segmentation (classif), localisation (régression)
- Texte: génération (classif), dialogue (ordonnancement)

Deuxième chose à déterminer

- Conditionne les choix précis de modèles
- Permet de fixer le protocole expérimental
- Ensuite, on peut réfléchir aux modèles

Premier modèle: linéaire

Dépendance linéaire entre l'entrée et la sortie

- Dans un espace vectoriel de dimension d, soit un point $x \in \mathbb{R}^d$
- $f_w(x) = \sum_{i=1}^{d} w_i x_i$
- f est le *modèle*
- w est le vecteur de *paramètres*

Apprentissage

- À partir des données
- Trouver le meilleur w

Abus de langage

- Linéaire: $y = ax$
- Affine: $y = ax + b$
- **Attention** en TP !

Régression linéaire

- Avec **UN** point: $\mathbf{x} = (x_1, \ldots, x_d)$
- Une étiquette $y \in \mathbb{R}$
- Approcher l'étiquette y avec un modèle $f(x)$

Apprentissage supervisé

- À partir de N points $\{\mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(N)}\}$
- Et leur label $y_1, \ldots, y_N$
- Chercher le meilleur compromis

Classification linéaire

- Avec **UN** point: $\mathbf{x} = (x_1, \dots, x_d)$
- Une étiquette $y \in \{-1, +1\}$
- Retrouver l'étiquette y avec un modèle $\mathrm{signe}(f(x))$

avec $\mathrm{signe}(x) = +1$ si $x \leq 0$ et -1 sinon

Apprentissage supervisé

- À partir de N points $\{\mathbf{x}^{(1)}, \dots, \mathbf{x}^{(N)}\}$
- Et leur label $y_1, \dots, y_N$
- Chercher le meilleur compromis

Géométriquement

- Tracer les points
- Tracer la frontière de décision
- À quoi correspond la somme ?

Plus proche voisin

Nearest neighbor

Idée

Associer un nouveau point à l'exemple connu le plus proche
[Au tableau]

k-plus proches voisins

Algorithme

- Pour un nouveau point
- Calculer toutes les distances entre ce point et les points connus
- Rechercher les k plus proches voisins
- Aggréger: vote (classification), moyenne (régression)

[Au tableau]

Analyse

Complexité algorithmique

- Temps, espace

Performances

- Influence du k
- Score et généralisation

[Au tableau]

Modèle linéaire

Classifieur binaire: deux sorties $+1$ ou -1

Hyperplan

- en dimension d
- équation $\sum_{i=1}^{d} w_i x_i = 0$
- $(w_1, \ldots, w_d)$ vecteur normal

Décision (ou inférence, ou prédiction)

- De quel coté est-on par rapport à la droite ?
- Produit scalaire entre le point à classer et le vecteur normal: signe de $\sum_{i=1}^{d} w_i x_i$
- Si positif: $+1$
- Si négatif: -1

Comment apprendre ?

Apprentissage ?

Paramètres du modèle

- Caractérise les frontière de décision
- Permet de prendre la décision
- Ici les w_i

Apprentissage

- Choisir ces paramètres
- À partir des données
- Objectif: chercher les meilleurs paramètres

Moins de linéarité

Hyperplan linéaire

- Passe par l'origine
- C'est le $= 0$ dans l'équation

Modèle affine

- $\sum_{i=1}^{d} w_i x_i = b$
- b: **biais**

Comment choisir ce biais ?

- L'apprendre comme les w: paramètre
- Le fixer à l'avance: **hyper-paramètre**

Linéarité: suffisant ou pas ?

[Au tableau]

Hyperplan aléatoire

Paramètres

- Vecteur normal w
- Éventuellement biais b

"Apprentissage"

- Tirer aléatoirement un vecteur w (avec $||w||$ suffisament grand)
- Pas vraiment de l'apprentissage: ne dépend pas des données

Performances

- Est-ce que ça marche bien ?
- Comment le savoir ?

Perceptron de Rosenblatt

~~Blabla~~ Vocabulaire

- Neurone formel
- Poids synaptique
- Activation

Version maths/stats/info

- Hyperplan, vecteur normal, etc
- Activation: signe (fonction de Heaviside, notée h)

Apprentissage

- Cette fois, un algorithme !
- Principe: mise à jour successives des paramètres en fonction des erreurs

Règle de Rosenblatt

Algorithme

- Tirer aléatoirement w
- **Tant qu'**il y a des erreurs:
  - **Pour** chaque exemple x et son étiquette y:
    - **Si** $y\langle w, x\rangle < 0$:
      $w \leftarrow w + yx$

Questions

- $y\langle w, x\rangle < 0$?
- *Tant qu'il y a des erreurs* ?
- Influence du w initial ?
- Impact de la mise à jour ?

Théorème de Novikoff

Convergence si et seulement si classes linéairement séparables

Règle de Widrow-Hoff

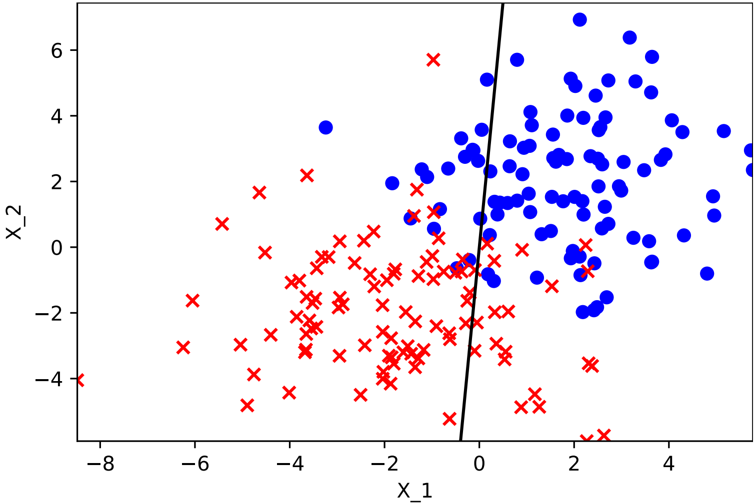
Algorithme

- Tirer aléatoirement w
- **Jusqu'à** convergence ou un nombre d'itérations maximum:
  - **Pour** chaque exemple x et son étiquette y:
    $\hat{y} = h(\langle w, x \rangle)$
    $w \leftarrow w + \alpha(y - \hat{y})x$

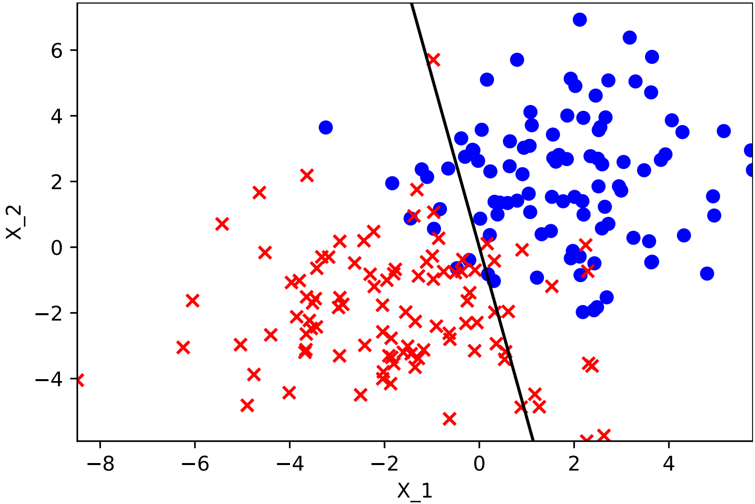
Questions

- Points communs et différences ?
- α ?
- Nombre d'itération maximum ?
- D'où ça vient ? Voir prochain cours

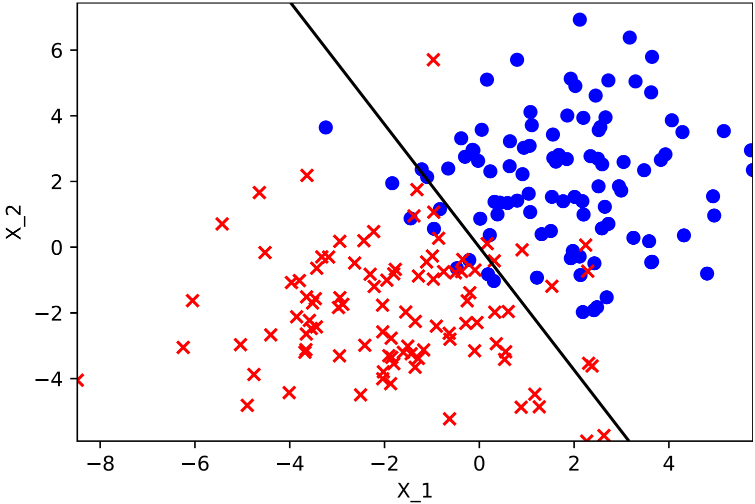
Démo



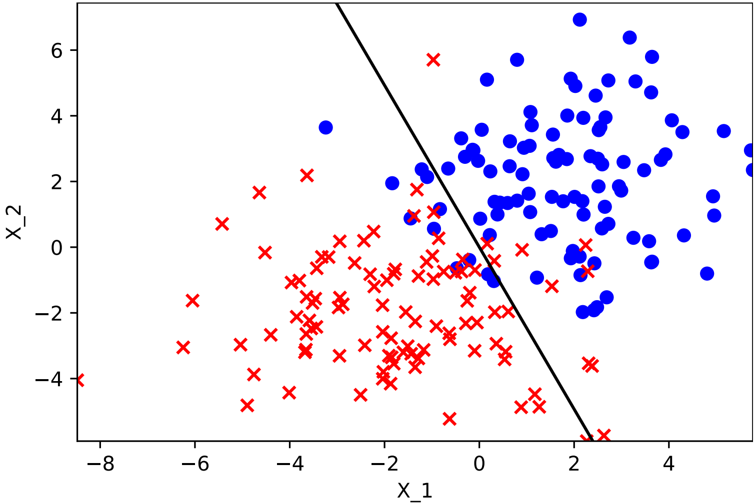
Démo



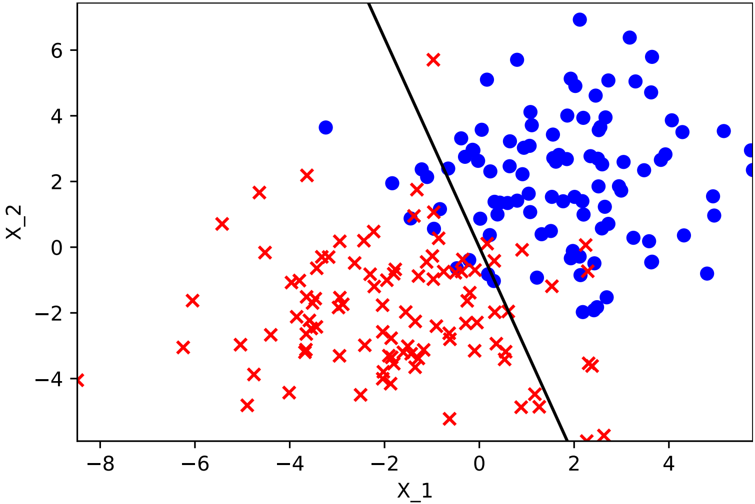
Démo



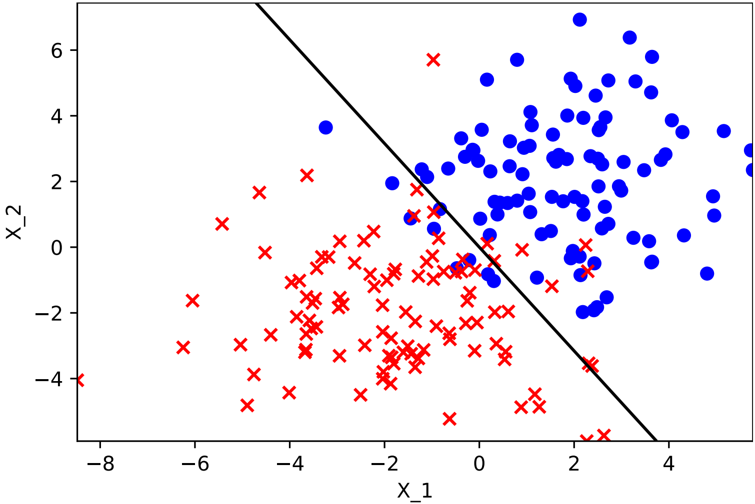
Démo



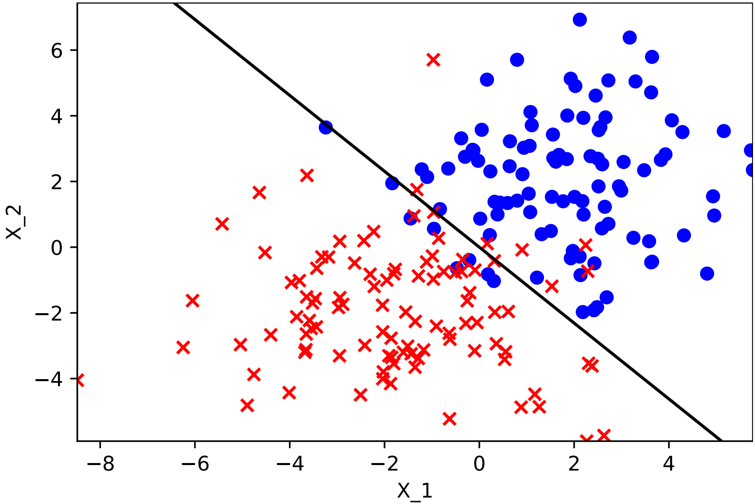
Démo



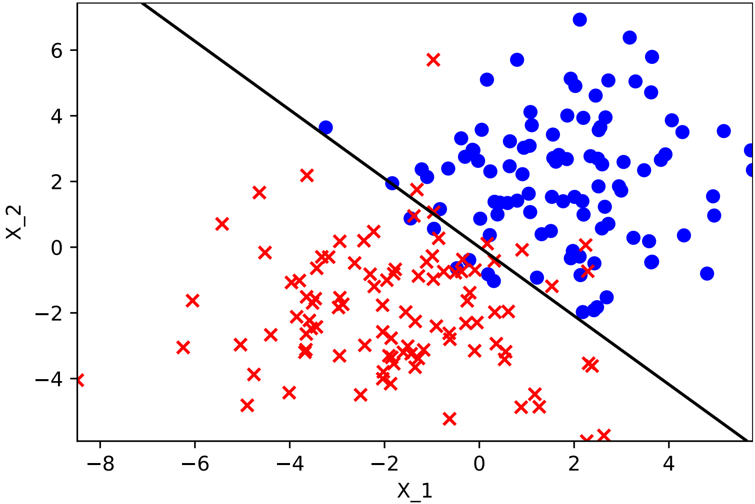
Démo



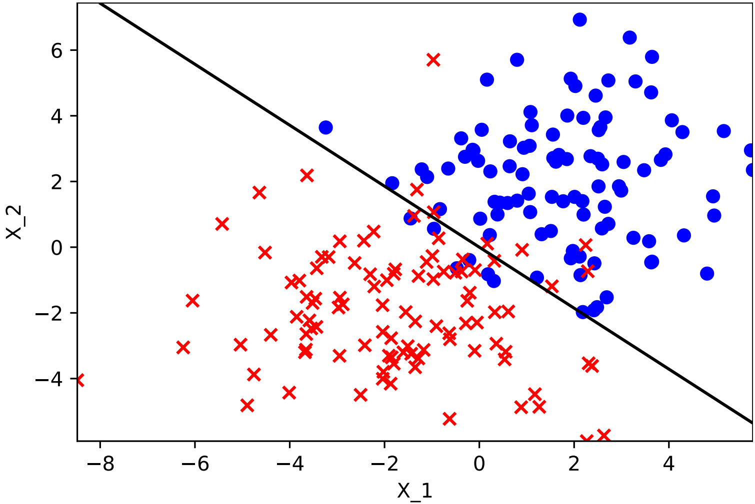
Démo



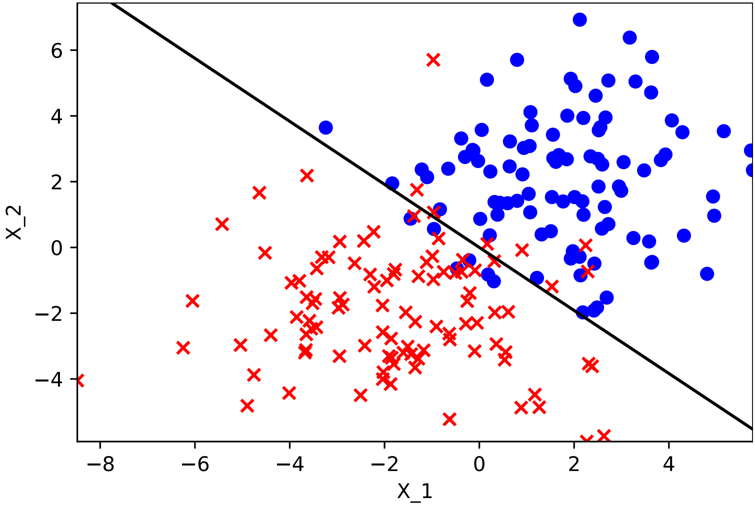
Démo



Démo



Démo



Et le biais ?

Astuce

- Départ: hyperplan en dimension d: $\sum_{i=1}^{d} w_i x_i = b$
- $\sum_{i=1}^{d} w_i x_i - b = 0$
- $\sum_{i=1}^{d} w_i x_i + b(-1) = 0$
- Nouveau $x = (x_1, \ldots, x_d, -1)$
- Nouveau $w = (w_1, \ldots, w_d, b)$
- Arrivée: hyperplan en dimension $d+1$: $\sum_{i=1}^{d+1} w_i x_i = 0$

Plus qu'une astuce

- Changement de la représentation des données
- Avant: modèle complexe, on ne sais pas quoi faire du biais
- Après: transformation des exemples, modèle simple et algorithmes déjà vus

Conclusion

Modèles et algorithme

- Modèle: forme de la frontière
- Apprentissage: trouver les paramètres à partir des données
- Hyper-paramètres: choisi à la main

Plein de choix

- Comment les faire ?

Évaluation

- La semaine prochaine

Objectif: généraliser