

Apprentissage automatique

Séance 1

Bruno Bouzy

bruno.bouzy@parisdescartes.fr

www.mi.parisdescartes.fr/~bouzy

Objectif

- Aperçu de l'apprentissage automatique (AA)
- Techniques fondamentales et représentatives
- Problématiques:
 - Classification oui surtout
 - Optimisation oui un peu
 - Prediction oui un peu
 - Décision d'un (des) agent(s) situé(s) dans environnement: non (M2)

Techniques abordées

- Apprentissage symbolique
- Réseaux bayésiens
- Réseaux de neurones
- Machine a vecteurs support
- Arbres de décision
- Bagging Boosting
- Méthodes des différences temporelles
- Evolution artificielle
- Algorithmes « bandit »

Techniques représentatives

- « Machine Learning » top-cited papers (2011):
 - Quinlan 1986, induction of **decision trees** (~3000)
 - Cortes & vapnik 1995, **support vector networks** (~2000)
 - Sutton 1988, learning to predict by methods of **temporal differences**
 - Breiman 2001, random forests (**bagging**) (~1000)
 - Aha 1991, instance-based learning algorithms
 - Cooper Herskovits 1992, a **bayesian** method for the induction of probabilistic **networks** from data (~900)
 - Shapire 1999, improved **boosting** algo... (~800)
 - etc.

Techniques fondamentales

- Mise a niveau
 - Approximation polynomiale
 - Descente de gradient
 - Probabilités
- Réseaux de neurones
 - Perceptron
- Evolution artificielle
- Apprentissage symbolique
 - Induction dans l'espace des versions

Guide du cours

Probabilités



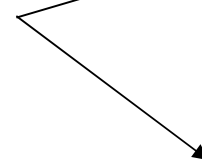
Réseaux bayésiens

Approximation polynomiale

Descente de gradient



Réseaux de neurones



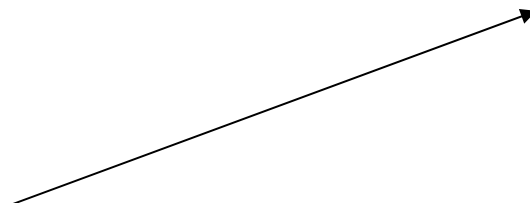
SVN

Arbres de décision



Bagging boosting

Apprentissage symbolique



Evolution artificielle

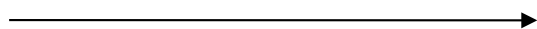
Différences temporelles

Algorithmes « bandit »

Apprentissage automatique

Apprentissage symbolique

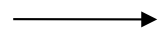
Probabilités



Réseaux bayésiens

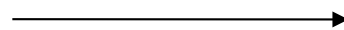
Approximation polynomiale

Descente de gradient



Réseaux de neurones

Arbres de décision



Bagging boosting

Apprentissage symbolique

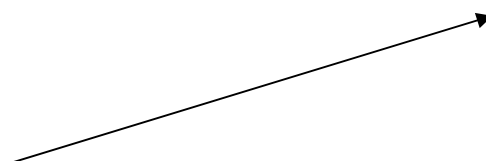
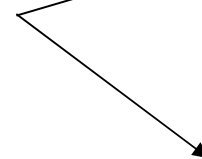
Evolution artificielle

Différences temporelles

Algorithmes « bandit »

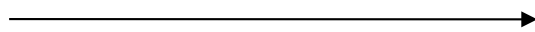
Apprentissage automatique

SVN



Apprentissage numérique

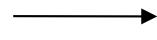
Probabilités



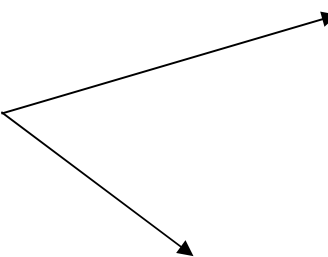
Réseaux bayésiens

Approximation polynomiale

Descente de gradient

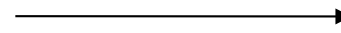


Réseaux de neurones



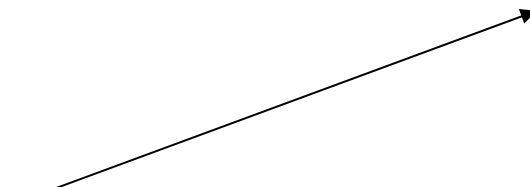
SVN

Arbres de décision



Bagging boosting

Apprentissage symbolique



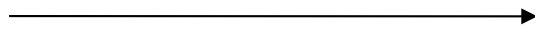
Evolution artificielle

Différences temporelles

Algorithmes « bandit »

Classification

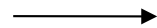
Probabilités



Réseaux bayésiens

Approximation polynomiale

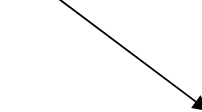
Descente de gradient



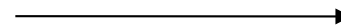
Réseaux de neurones



SVN



Arbres de décision



Bagging boosting

Apprentissage symbolique



Evolution artificielle

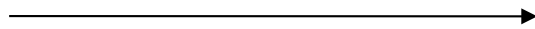
Différences temporelles

Algorithmes « bandit »

Apprentissage automatique

Prediction

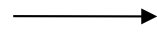
Probabilités



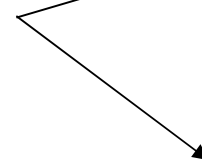
Réseaux bayésiens

Approximation polynomiale

Descente de gradient

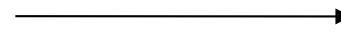


Réseaux de neurones



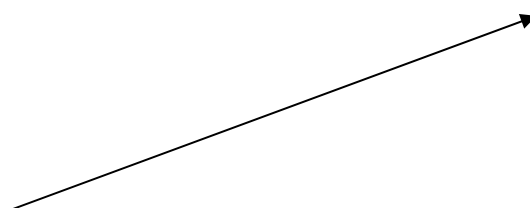
SVN

Arbres de décision



Bagging boosting

Apprentissage symbolique



Evolution artificielle

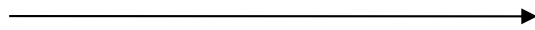
Différences temporelles

Algorithmes « bandit »

Apprentissage automatique

Optimisation

Probabilités



Réseaux bayésiens

Approximation polynomiale

Descente de gradient



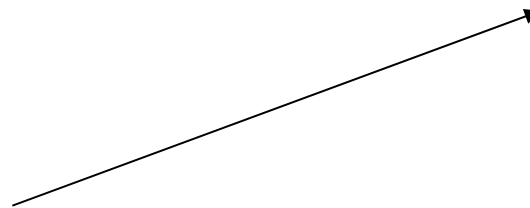
Réseaux de neurones

Arbres de décision



Bagging boosting

Apprentissage symbolique



SVN

Evolution artificielle

Différences temporelles

Algorithmes « bandit »

Apprentissage automatique

Apprentissage supervisé

Probabilités



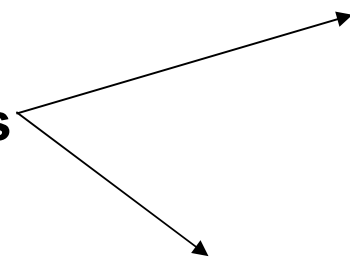
Réseaux bayésiens

Approximation polynomiale

Descente de gradient



Réseaux de neurones



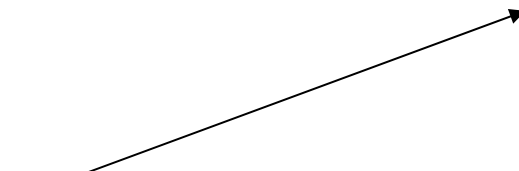
SVN

Arbres de décision



Bagging boosting

Apprentissage symbolique



Evolution artificielle

Différences temporelles

Algorithmes « bandit »

Apprentissage automatique

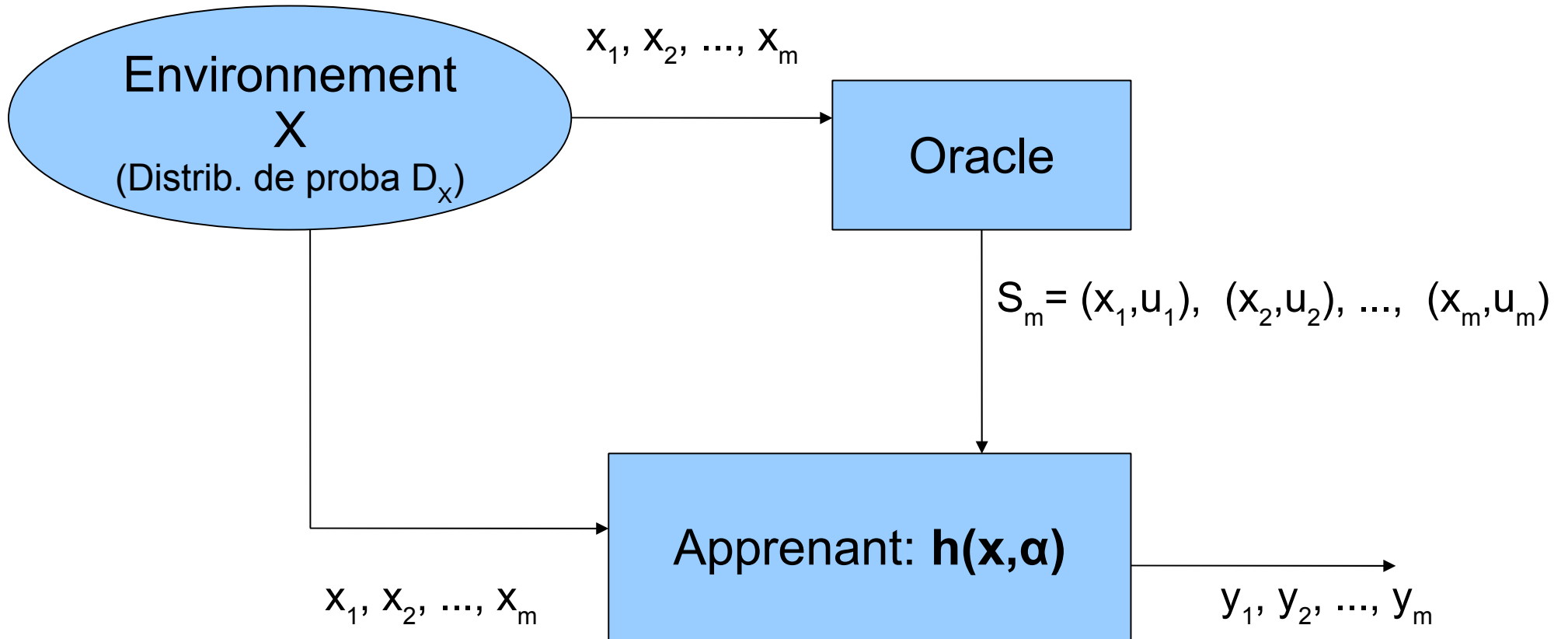
Apprentissage supervisé ou non ?

- Supervisé
 - Un oracle classe les exemples
 - L'apprenant apprend à classer comme l'oracle
 - Non supervisé
 - L'apprenant apprend par lui-même
- à classer
- à décider d'une action

Induction

- Langage courant:
 - Expliquer
 - Prédire, classifier
 - Faire simple
- Jargon de l'apprentissage automatique:
 - **Engendrer des hypothèses**
 - ...plus ou moins simples
 - ...pour classifier, expliquer, décider

Induction



Induction

- Exemples non classifiés
 - x_1, x_2, \dots, x_m .
- 1 oracle classe les exemples
 - $(x_1, u_1), (x_2, u_2), \dots, (x_m, u_m)$.
- L'apprenant apprend (phase d'apprentissage)
 - En recherchant une hypothèse $h(\cdot, \alpha)$ permettant de classier au mieux les exemples
- L'apprenant classe (phase de test)
 - les exemples nouveaux en utilisant l' hypothèse apprise

Les espaces et ensembles

- X : espace des exemples x
- H : espace des hypothèses h
- F : espace des fonctions cibles f
- S : échantillon (sample)
- S_+ : ensemble des exemples positifs
- S_- : ensemble des exemples négatifs $S = S_+ + S_-$
- A : ensemble d'apprentissage
- T : ensemble de test $S = A + T$

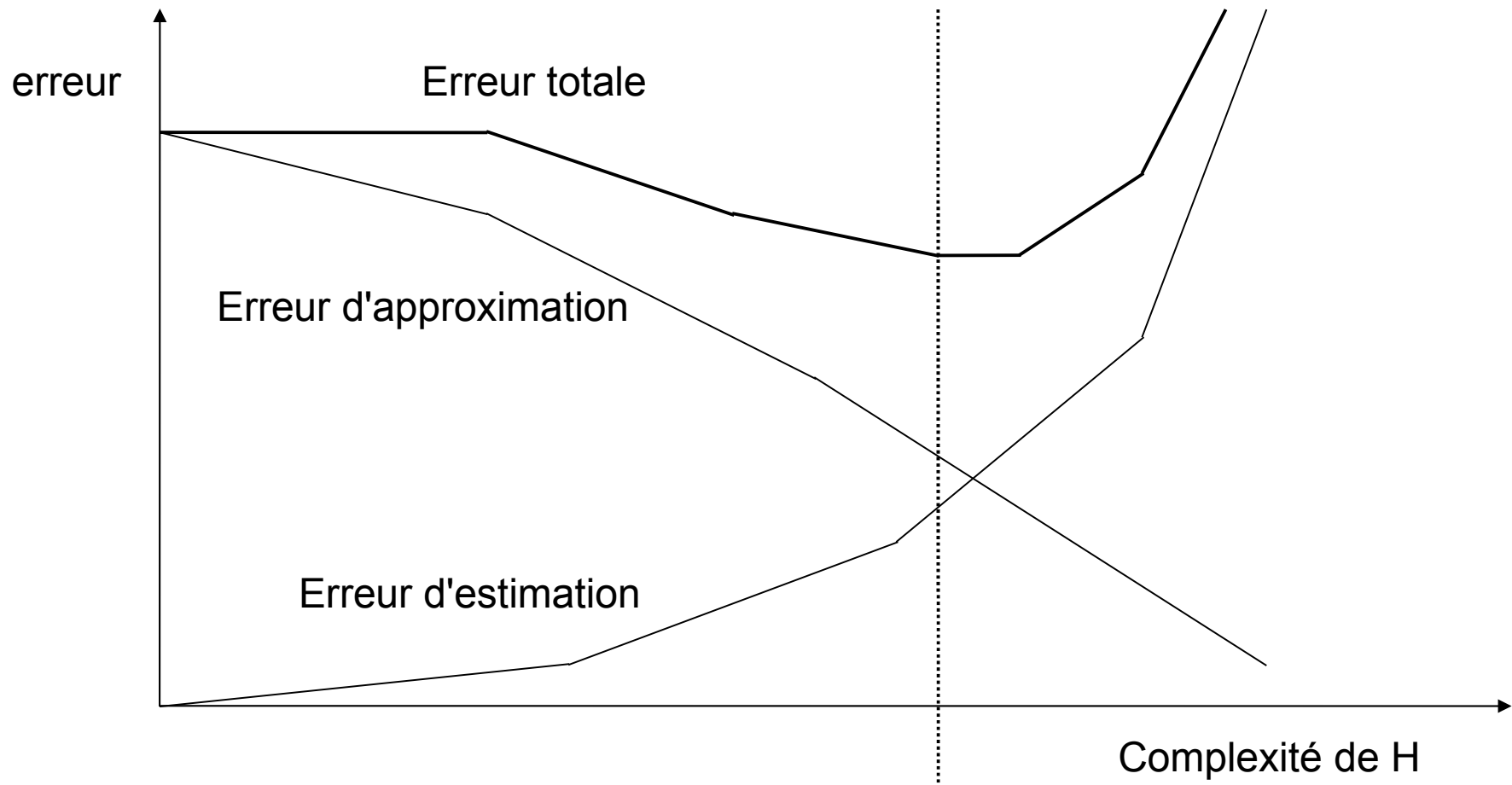
Biais et variance

- Apprentissage d'une hypothèse h dans H .
- h^* : hypothèse optimale dans H .
- $h - h^*$: « **variance** » ou « erreur d'estimation ».
- Exemples classifiés idéalement avec une fonction cible optimale f^* dans $F \neq H$.
- $h^* - f^*$: « **biais** » ou « erreur d'approximation ».
- $h - f^*$: erreur totale.

Compromis biais-variance

- Erreur totale =
 - erreur d'estimation + erreur d'approximation
 - variance + biais
- Plus H est complexe...
 - plus H se rapproche de F , plus le biais diminue.
 - plus 2 hypothèses donnent des résultats différents, plus la variance augmente.
- H de complexité intermédiaire pour minimiser l'erreur totale

Compromis biais-variance



Compromis biais-variance

- Reconnaître les hommes et les femmes...
 - Modèle simple: la taille pour classifier
 - Les hommes sont plus grands que les femmes.
 - biais important: hommes petits et femmes grandes (erreur d'approximation)
 - variance faible: seuil de taille précisément calculé
 - Modèle plus complexe: taille, longueur cheveux, poids, timbre de la voix, pilosité.
 - biais moins important,
 - variance plus importante.

Induction

- Environnement X :

Engendre des **exemples** x_i tirés indépendamment suivant une distribution D_x .

- Oracle:

Pour chaque x_i , fournit une **étiquette** u_i dans U (avec une distribution de probabilité $F(u|x)$ inconnue)

- Apprenant:

Construit une **hypothèse** h dans H telle que:

$h(x_i) = u_i$ pour chaque x_i .

Induction

- **Perte** (loss) de décider sur x_i avec h :

$L(u_i, h(x_i))$ coût de la décision $h(x_i)$.

- **Risque réel** d'une hypothèse h :

$$R_{\text{réel}}(h) = \int_{X \times U} L(u_i, h(x_i)) dF(x, u)$$

- **Principe inductif**: minimiser $R_{\text{réel}}(h)$
- Problème: F est inconnue, $R_{\text{réel}}$ inconnu
- $h^* = \operatorname{argmin}_h (R_{\text{réel}}(h))$

Induction

- **Risque empirique** d'une hypothèse h sur un échantillon S :

$$R_{\text{emp}}(h, S) = 1/m \sum_{i=1, m} L(u_i, h(x_i))$$

Perte moyenne mesurée sur l'échantillon S

- Minimiser $R_{\text{emp}}(h, S)$.
- Empirical Risk Minimization (ERM)
- $h^{\wedge}_S = \operatorname{argmin}_h (R_{\text{emp}}(h, S))$

Risque empirique et risque réel

$R_{\text{réel}}(h)$ inconnu, f^* inconnu, $R_{\text{emp}}(h, S_m)$ connu

(0) $R_{\text{réel}}(f^*)=0$ (par définition)

(1) $R_{\text{réel}}(h^*)>0$ (biais) (par définition)

(2) $R_{\text{réel}}(h^*) < R_{\text{réel}}(h^*_S)$ (par définition)

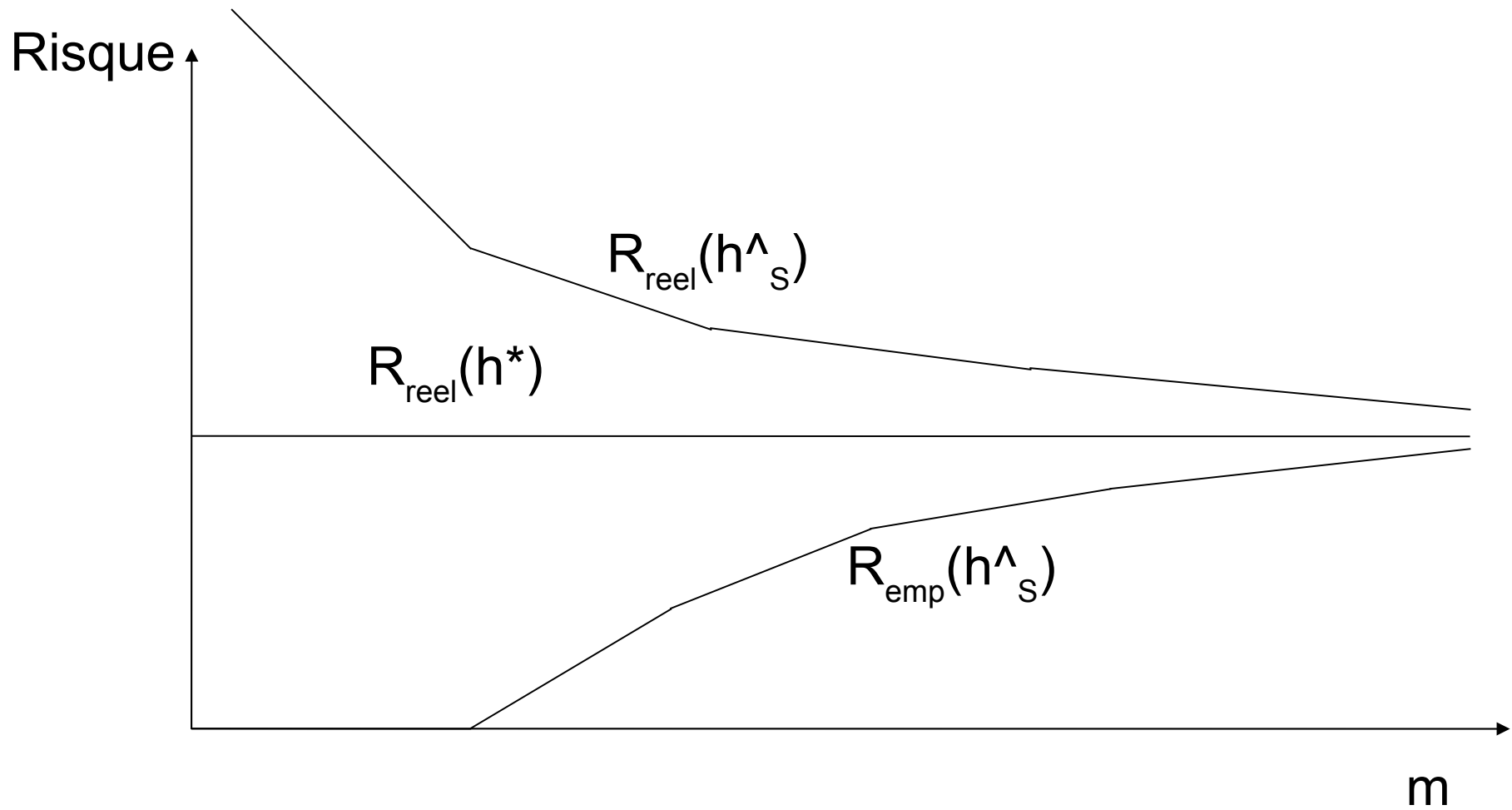
(3) $R_{\text{emp}}(h^*_S) < R_{\text{emp}}(h^*)$

(4) $R_{\text{emp}}(h^*_S)$ augmente si $m \rightarrow \infty$ (plus S grand, plus difficile de minimiser)

(5) $R_{\text{réel}}(h^*_S)$ diminue si $m \rightarrow \infty$ (plus S grand, plus monde réel couvert)

(6) $\lim_{m \rightarrow \infty} (R_{\text{emp}}(h^*_S)) < R_{\text{réel}}(h^*) < \lim_{m \rightarrow \infty} (R_{\text{réel}}(h^*_S))$? (oui si pertinence)

Risque empirique et risque réel



Apprentissage automatique

Références

- [1] Antoine Cornuéjols & Laurent Miclet, « Apprentissage artificiel, concepts et algorithmes », (préface de Tom Mitchell), Eyrolles.
- [2] Stuart Russell & Peter Norvig, « Artificial Intelligence: a modern approach ».
- [3] Machine Learning.