CN=CPCN

Liva Ralaivola, François Denis, Christophe N. Magnan

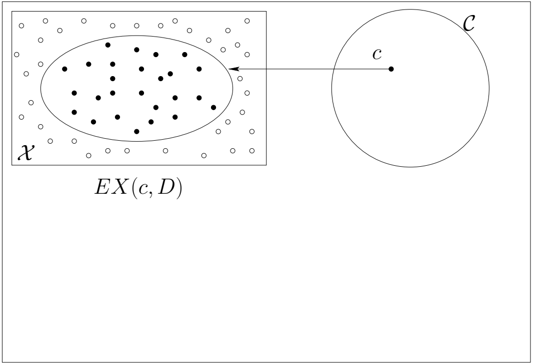
Laboratoire d'Informatique Fondamentale de Marseille, UMR CNRS 6166

CAp 2006, ICML 2006

Introduction

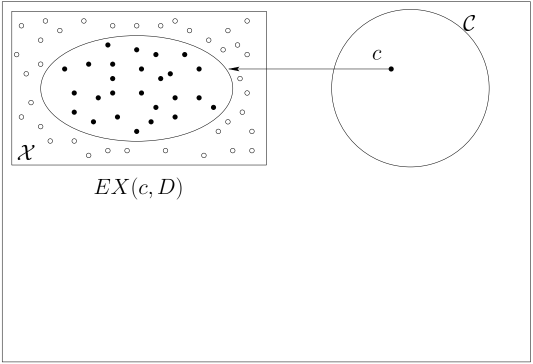
- Apprendre dans le modèle PAC
- Modèles d'apprentissage avec bruit de classification:
  - CN
  - CPCN
  - CCCN

Apprendre dans le modèle PAC



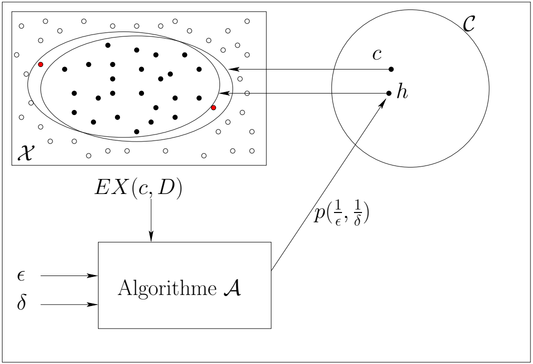
$$\text{Oracle } EX(c, D) \rightarrow \langle x, c(x) \rangle \text{ tel que } \begin{cases} x \in \mathcal{X} \text{ tiré selon } D \\ c(x) = 1 \text{ si } x \in c, \\ c(x) = 0 \text{ sinon} \end{cases}$$

Apprendre dans le modèle PAC



Objectif: $\forall c \in \mathcal{C}$, calculer à partir de $EX(c, D)$ une bonne approximation de c avec une forte confiance

Définition



$\mathcal{C}$ est efficacement PAC-apprenable ssi:
$\exists \mathcal{A}, \forall c \in \mathcal{C}, \forall D, \forall \epsilon, \delta > 0, \mathcal{A}(EX(c, D), \epsilon, \delta) \rightarrow h$
tel que $P(R(h) > \epsilon) < \delta$ avec $R(h) = P(h(x) \neq c(x))$.

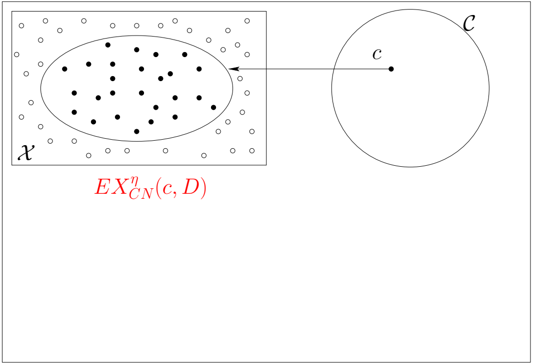
Bruit de classification

Apprentissage dans le cadre PAC avec différents modèles de bruit de classification

Bruit de classification uniforme [Angluin & Laird, 1988]

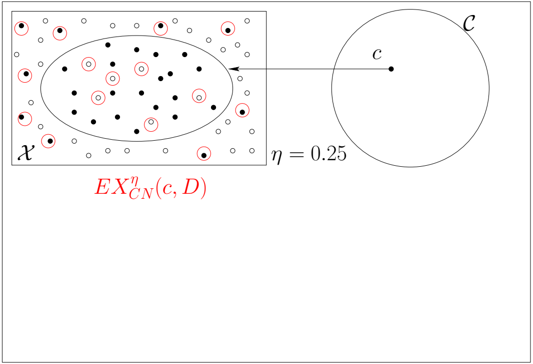
Bruit de classification uniforme: CN

Bruit de classification uniforme [Angluin & Laird, 1988]



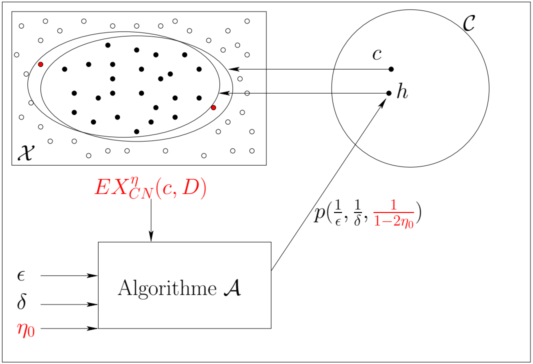
$$EX_{CN}^{\eta}(c, D) \rightarrow \langle x, c^{\eta}(x) \rangle \text{ t.q. } \begin{cases} c^{\eta}(x) = c(x) \text{ avec prob. } 1 - \eta \\ c^{\eta}(x) = 1 - c(x) \text{ avec prob. } \eta \end{cases}$$

Bruit de classification uniforme [Angluin & Laird, 1988]



$$EX^{\eta}_{CN}(c, D) \rightarrow \langle x, c^{\eta}(x) \rangle \text{ t.q. } \begin{cases} c^{\eta}(x) = c(x) \text{ avec prob. } 1 - \eta \\ c^{\eta}(x) = 1 - c(x) \text{ avec prob. } \eta \end{cases}$$

Définition

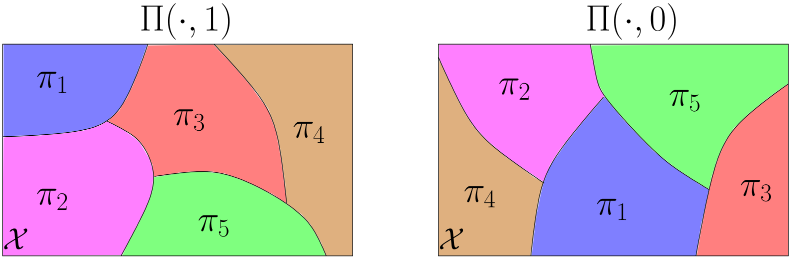


$\mathcal{C}$ est efficacement PAC-apprenable avec CN ssi:
$\exists \mathcal{A}, \forall c \in \mathcal{C}, \forall D, \forall \epsilon, \delta > 0, \forall \eta_0 \in [0; 0.5[, \forall \eta < \eta_0,$
$\mathcal{A}(EX^{\eta}_{CN}(c, D), \epsilon, \delta, \eta_0) \rightarrow h$ tel que $P(R(h) > \epsilon) < \delta$.

Bruit de classification constant par morceaux [Decatur, 1997]

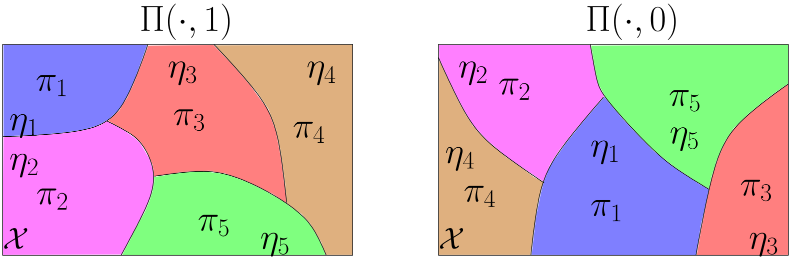
Bruit de classification constant par morceaux: CPCN

Bruit de classification constant par morceaux [Decatur, 1997]



$\Pi = \{\pi_1, ..., \pi_k\}$: partition de $\mathcal{X}$x$\{0, 1\}$

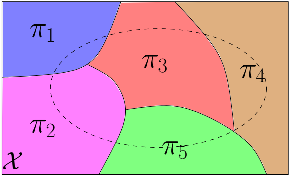
Bruit de classification constant par morceaux [Decatur, 1997]



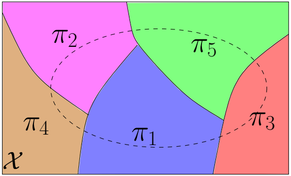
$\Pi = \{\pi_1, ..., \pi_k\}$: partition de $\mathcal{X}$x$\{0,1\}$
$\overrightarrow{\eta} = [\eta_1, ..., \eta_k]$, $\eta_i \in [0,1]\ \forall i$

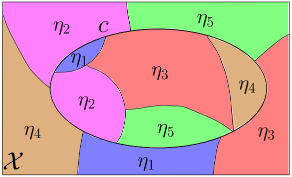
Bruit de classification constant par morceaux [Decatur, 1997]

$\Pi(\cdot, 1)$

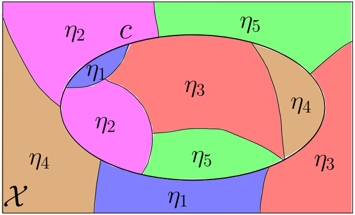


$\Pi(\cdot, 0)$



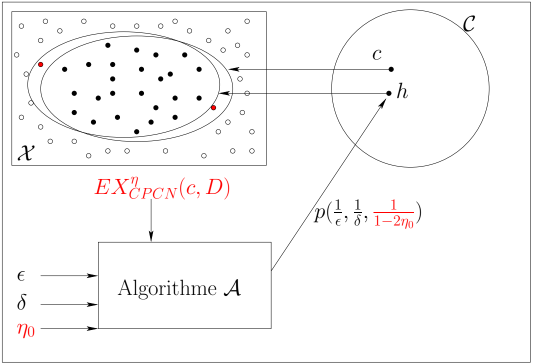


Bruit de classification constant par morceaux [Decatur, 1997]



$$EX^{\eta}_{CPCN} \rightarrow \langle x, c^{\eta}(x)\rangle t.q. \begin{cases} c^{\eta}(x) = c(x) \text{ avec prob. } 1 - \eta_i \\ c^{\eta}(x) = 1 - c(x) \text{ avec prob. } \eta_i \\ \text{avec } i \text{ tq } \pi_i(\langle x, c(x)\rangle) = 1 \end{cases}$$

Définition



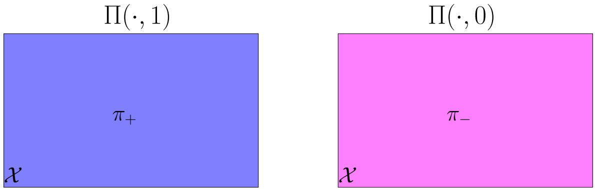
$\mathcal{C}$ est efficacement PAC-apprenable avec CPCN ssi:

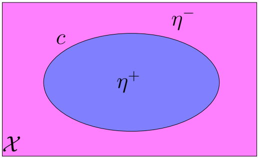
$$\exists \mathcal{A}, \forall c \in \mathcal{C}, \forall D, \forall \epsilon, \delta > 0, \forall \Pi = \{\pi_1, ..., \pi_k\},$$
$$\forall \eta_0 \in [0; 0.5]\ \forall \eta = [\eta_1 ... \eta_k] \text{ tel que } \eta_i \leq \eta_0\ \forall i,$$
$$\mathcal{A}(EX^{\eta}_{CPCN}(c, D), \epsilon, \delta, \eta_0) \rightarrow h \text{ tel que } P(R(h) > \epsilon) < \delta.$$

Bruit de classification conditionnel à chaque classe [Denis, 1998]

Bruit de classification conditionnel à chaque classe: CCCN
Cas particulier du bruit CPCN

Bruit de classification conditionnel à chaque classe [Denis, 1998]





$$\pi_+(\langle x, y\rangle) = y$$
$$\pi_-(\langle x, y\rangle) = 1 - y$$

$CN = CPCN$

Démonstration en deux étapes

- $CN = CCCN$
- $CCCN = CPCN$

et donc $CN = CCCN = CPCN$

Remarque: $CN \supseteq CCCN \supseteq CPCN$ est trivial

$CN \subseteq CCCN$

Soit $\mathcal{C}$ une classe de concepts CN-apprenable, soit $\mathcal{A}$ un algorithme qui CN-apprend $\mathcal{C}$.

Montrons que l'on peut construire un algorithme $\mathcal{A}'$ qui CCCN-apprend $\mathcal{C}$ à partir d'un oracle $EX_{CCCN}^{[\eta^+\eta^-]}$.

$CN \subseteq CCCN$

Idée:

- Rebruiter les données issues de $EX_{CCCN}^{[\eta^+,\eta^-]}$ pour plusieurs valeurs d'un paramètre de rebruitage ρ
- Apprendre avec $\mathcal{A}$ une hypothèse h sur chaque nouveau jeu de données rebruité
- Sélectionner une hypothèse h en appliquant le principe de minimisation du risque empirique

$CN \subseteq CCCN$

Idée:

- Rebruiter les données issues de $EX_{CCCN}^{[\eta^+,\eta^-]}$ pour plusieurs valeurs d'un paramètre de rebruitage ρ
- Apprendre avec $\mathcal{A}$ une hypothèse h sur chaque nouveau jeu de données rebruité
- Sélectionner une hypothèse h en appliquant le principe de minimisation du risque empirique

$\Rightarrow$ sous des conditions précises, on est assuré de sélectionner une hypothèse h telle que $P(R(h) > \epsilon) < \delta \quad \forall \epsilon, \delta > 0$

Rebruitage des données

Considérons $\eta^- > \eta^+$. Le processus suivant:

$$\begin{tikzcd}
 & & \langle x, 1\rangle \\
 & \langle x, 1\rangle \arrow[ur, "1-\rho"] \arrow[dr, "\rho"'] & \\
 {} \arrow[ur] \arrow[dr] & & \langle x, 0\rangle \\
 & \langle x, 0\rangle \arrow[r] & \langle x, 0\rangle
\end{tikzcd}$$

simule un appel à un oracle $EX_{CCCN}^{[\bar{\eta}^+ \bar{\eta}^-]}(c, D)$ tel que:

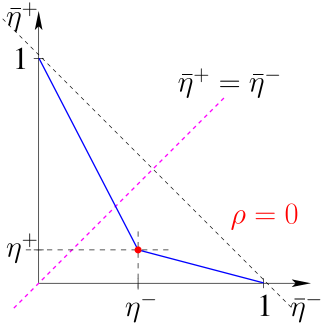
$$\begin{cases} \bar{\eta}^+ = (1-\rho)\eta^+ + \rho \\ \bar{\eta}^- = (1-\rho)\eta^- \end{cases}$$

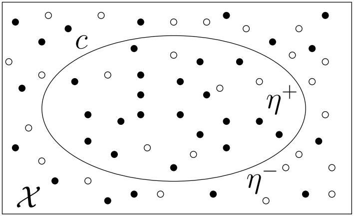
ρ_{opt}

Soit $\rho_{opt} = \frac{|\eta^+ - \eta^-|}{1+|\eta^+ - \eta^-|}$

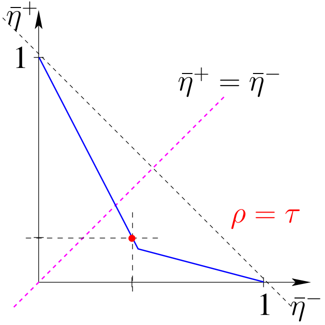
alors $\bar{\eta}^+ = \bar{\eta}^- = \eta_{opt} = \frac{max(\eta^+,\eta^-)}{1+|\eta^+ - \eta^-|}$

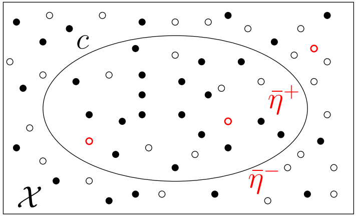
et $EX_{CCCN}^{[\bar{\eta}^+ \bar{\eta}^-]} \equiv EX_{CN}^{\eta_{opt}}$.


$\bar{\eta}^+$
1
$\bar{\eta}^+ = \bar{\eta}^-$
$\rho = 0$
η^+
η^-
1
$\bar{\eta}^-$

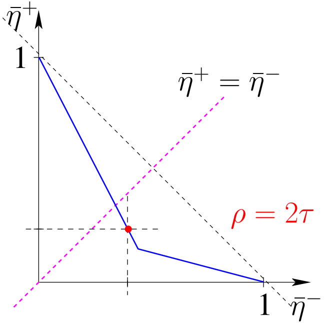

c
η^+
η^-
$\mathcal{X}$

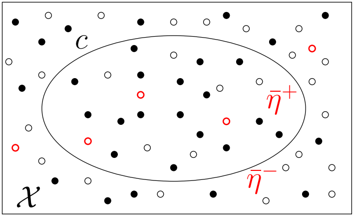
$$\mathcal{H} = \{h_0\}$$



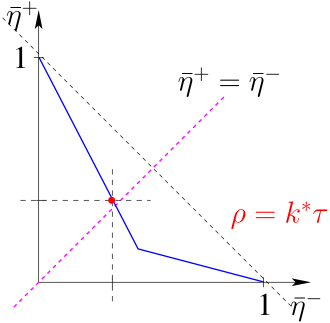


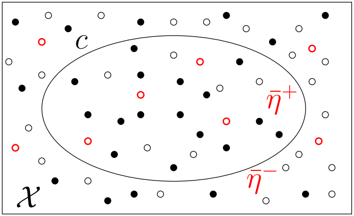
$$\mathcal{H} = \{h_0, h_1\}$$



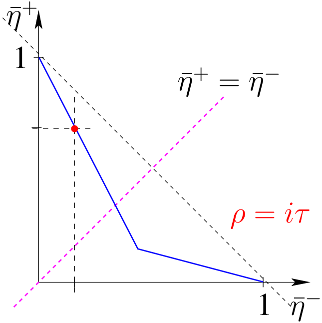


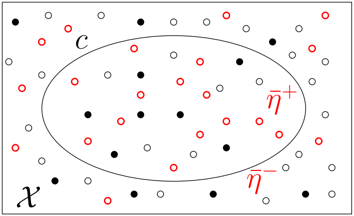
$$\mathcal{H} = \{h_0, h_1, h_2\}$$



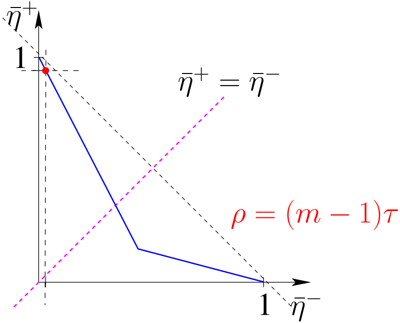


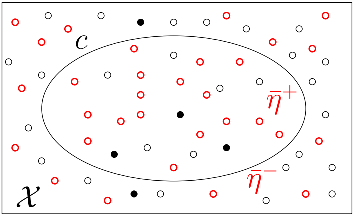
$$\mathcal{H} = \{h_0, h_1, h_2, ..., h_{k^*}\}$$



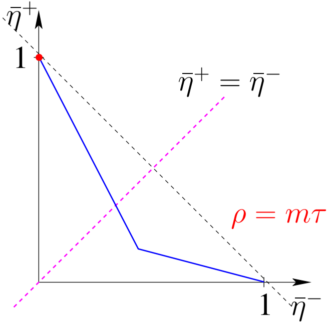


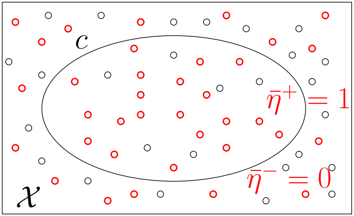
$$\mathcal{H} = \{h_0, h_1, h_2, ..., h_{k^*}, ..., h_i\}$$





$$\mathcal{H} = \{h_0, h_1, h_2, ..., h_{k^*}, ..., h_i, ..., h_{m-1}\}$$





$$\mathcal{H} = \{h_0, h_1, h_2, ..., h_{k^*}, ..., h_i, ..., h_{m-1}, h_m\}$$

$$CCCN \subseteq CPCN$$

$CCCN \subseteq CPCN$

Soit $\mathcal{C}$ une classe de concepts CCCN-apprenable, soit $\mathcal{A}$ un algorithme qui CCCN-apprend $\mathcal{C}$.

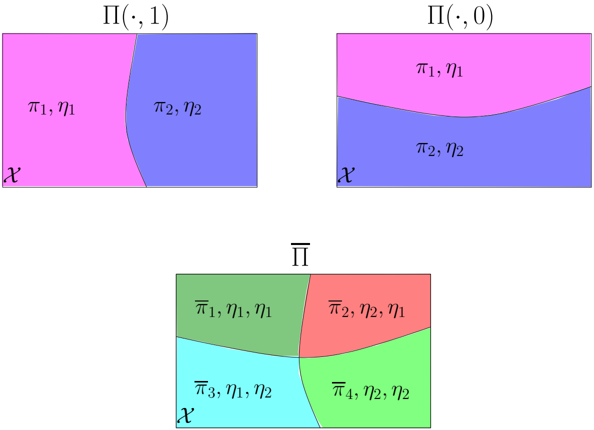
Montrons que l'on peut construire $\mathcal{A}'$ qui CPCN-apprend $\mathcal{C}$ à partir d'un oracle EX^{η}_{CPCN}.

$CCCN \subseteq CPCN$

Idée: créer une nouvelle partition $\bar{\Pi}$ de $\mathcal{X} \times \{0, 1\}$ à partir de Π

But: transformer un problème d'apprentissage CPCN en k problèmes d'apprentissage CCCN

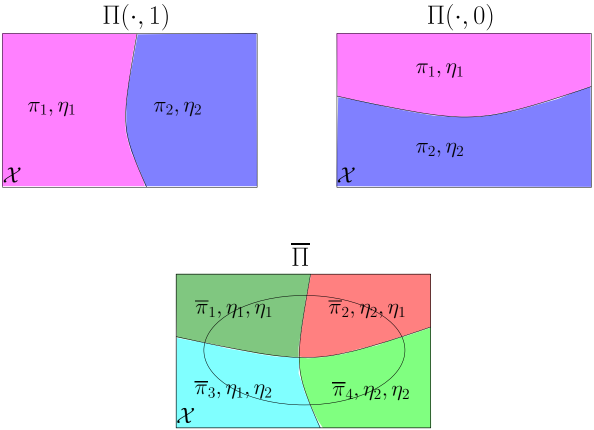
Repartitionnement de $\mathcal{X}$x$\{0,1\}$



$$\bar{\pi}_i = \pi_{uv} = \{x \in \mathcal{X} | \langle x, 1\rangle \in \pi_u, \langle x, 0\rangle \in \pi_v\}$$

et soient $\eta_i^+ = \eta_u$ et $\eta_i^- = \eta_v$

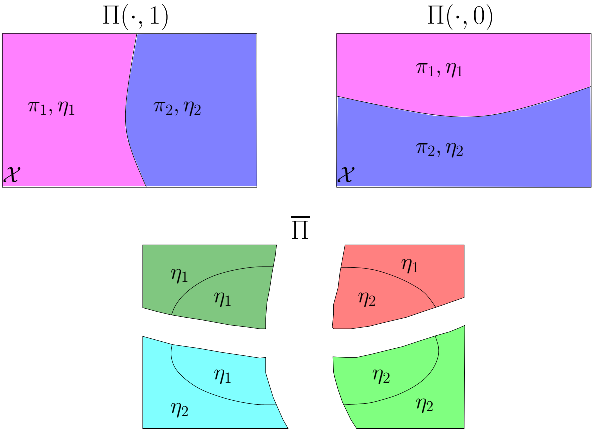
Repartitionnement de $\mathcal{X}$x$\{0, 1\}$



$$\bar{\pi}_i = \pi_{uv} = \{x \in \mathcal{X} | \langle x, 1\rangle \in \pi_u, \langle x, 0\rangle \in \pi_v\}$$

et soient $\eta_i^+ = \eta_u$ et $\eta_i^- = \eta_v$

Repartitionnement de $\mathcal{X}$x$\{0,1\}$



1 partition $CPCN \rightarrow k$ partitions $CCCN$

$CCCN \subseteq CPCN$

- Soit S un ensemble d'exemples tiré par EX^{η}_{CPCN}.
- $S \rightarrow k$ ensembles $S_i = S \cap \bar{\pi}_i$
- $\forall S_i$, apprendre avec $\mathcal{A}$ une hypothèse h_i sur $\bar{\pi}_i$
- Avec les hypothèses h_i, re-étiqueter les exemples d'un ensemble de données, et apprendre une hypothèse h sur $\mathcal{X}$ à partir du nouveau jeu de données

$CCCN \subseteq CPCN$

- Soit S un ensemble d'exemples tiré par EX^{η}_{CPCN}.
- $S \rightarrow k$ ensembles $S_i = S \cap \bar{\pi}_i$
- $\forall S_i$, apprendre avec $\mathcal{A}$ une hypothèse h_i sur $\bar{\pi}_i$
- Avec les hypothèses h_i, re-étiqueter les exemples d'un ensemble de données, et apprendre une hypothèse h sur $\mathcal{X}$ à partir du nouveau jeu de données

Si S est de taille suffisante (fixée), alors on est assuré que $R(h) < \epsilon$ avec une confiance $\delta \quad \forall \epsilon, \delta > 0$

Discussion

Bornes sur les bruits

- Cette démonstration tient lorsque les bruits sont tous inférieurs à 0.5 ou tous supérieurs à 0.5
- Que se passe-t-il lorsque les bruits ne respectent plus cette condition?
- Est-il possible de supprimer la contrainte de la borne η_0 sur les bruits?

Perspectives

- Qu'en-est-il de l'apprentissage PAC avec bruit de classification?
- Etendre ce type de résultat au cadre de l'apprentissage statistique.