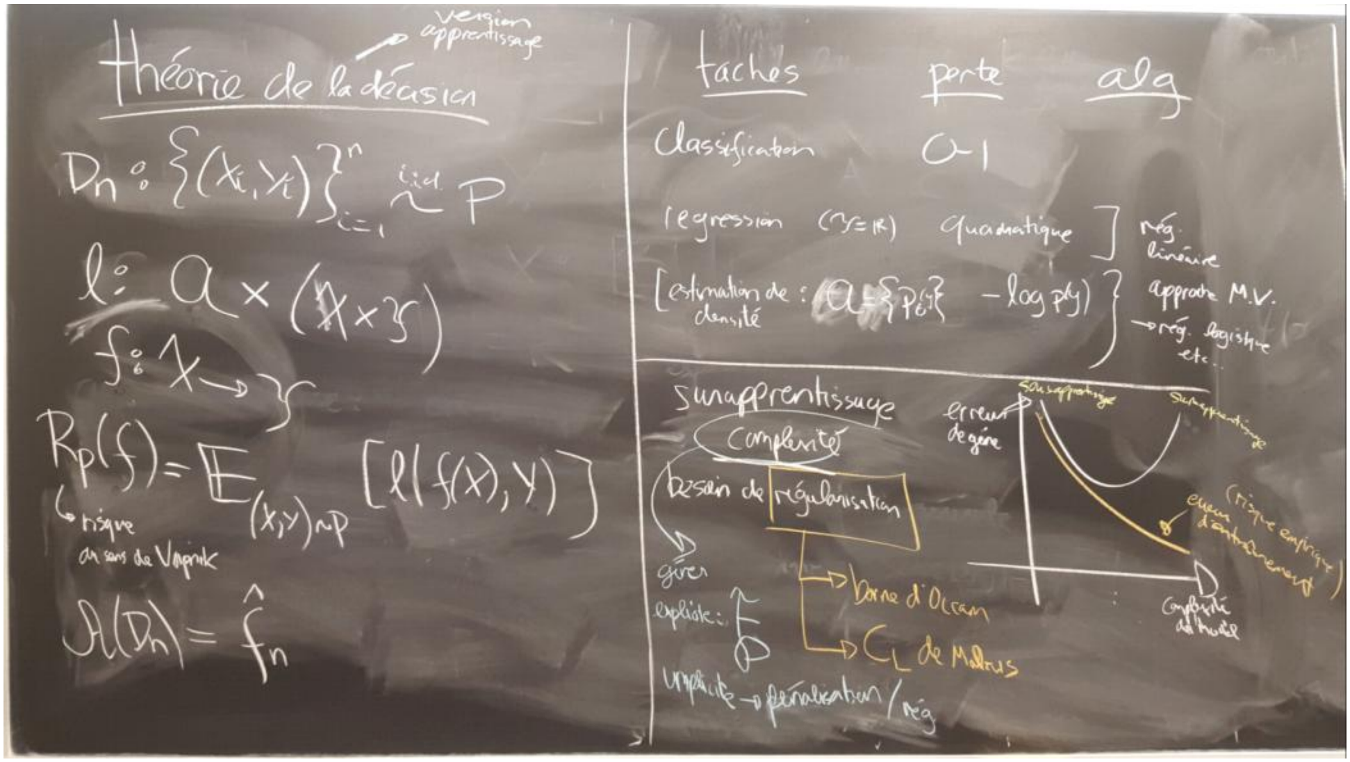
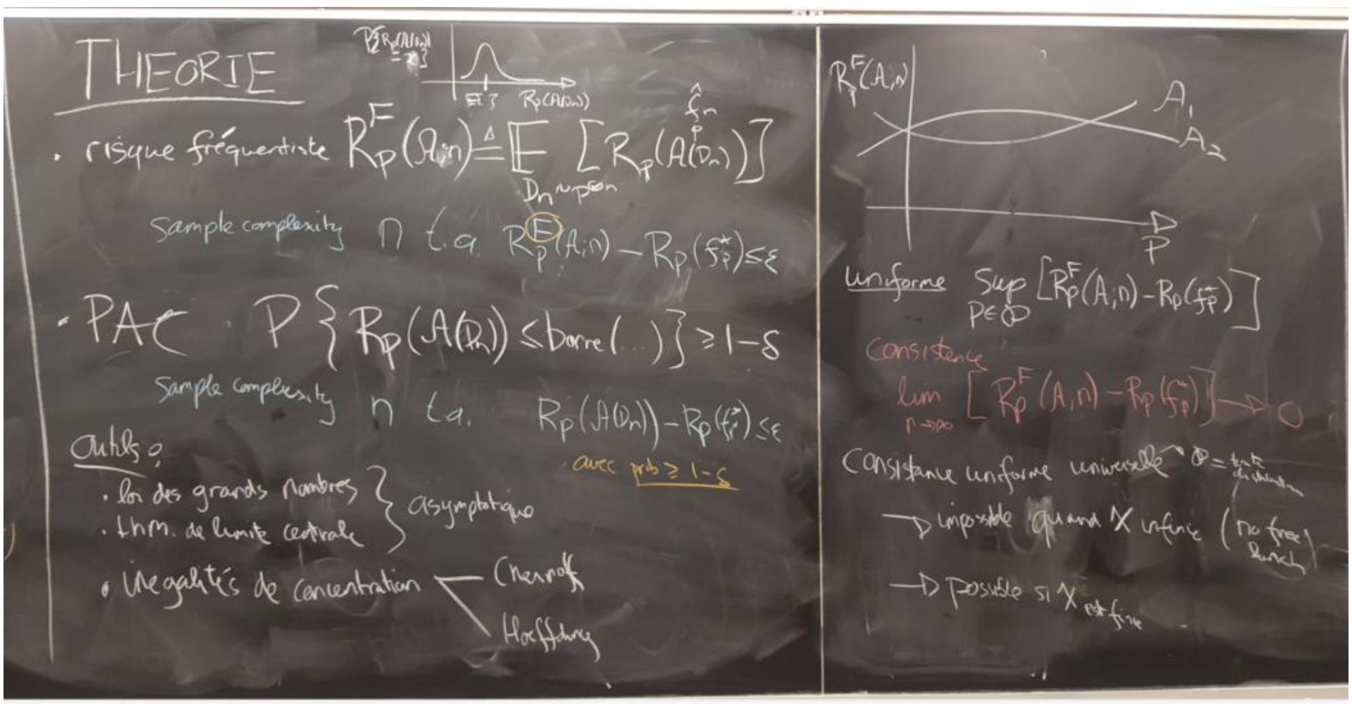
Cours 12 - révision

Monday, January 18, 2016
21:32





R_P^F → pour perte quadratique $R_P^F(\hat f_n) - R_P(f_P^*) = \underbrace{\mathbb{E}_X\left[\left(\mathbb{E}_{D_n}[\hat f_n(X)|X] - f_P^*(X)\right)^2\right]}_{\text{biais}} + \underbrace{\mathbb{E}_X \mathbb{E}_{D_n}\left(\left(\hat f_n(X) - \mathbb{E}_{D_n}[\hat f_n(X)|X]\right)^2\right)}_{\text{variance}}$

$\boxed{\text{décomposition biais-variance}}$

- James-Stein
- C_L Mallows design X fixe

} régression

- Consistance de alg. partition [plug-in]

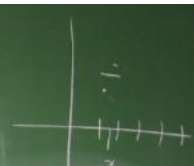
$$\begin{cases} h_n \to 0 & \} \text{ biais} \to 0 \\ n h_n^d \to +\infty & \} \text{ variance} \to 0 \end{cases}$$

$n \gg \frac{1}{h_n^{d}}$ → fléau de dimension

$\boxed{\text{borne d'Occam}}$ $\forall P$ $\left[\forall f \; R_P(f) \le \hat R_n(f) + \frac{1}{\sqrt{2n}} \underbrace{\sqrt{\ln 2\, |f|_{\pi} + \ln \frac{1}{\delta}}}_{\text{complexité}(f)}\right]$ avec prob. $\ge 1-\delta$

π → à priori

Approches

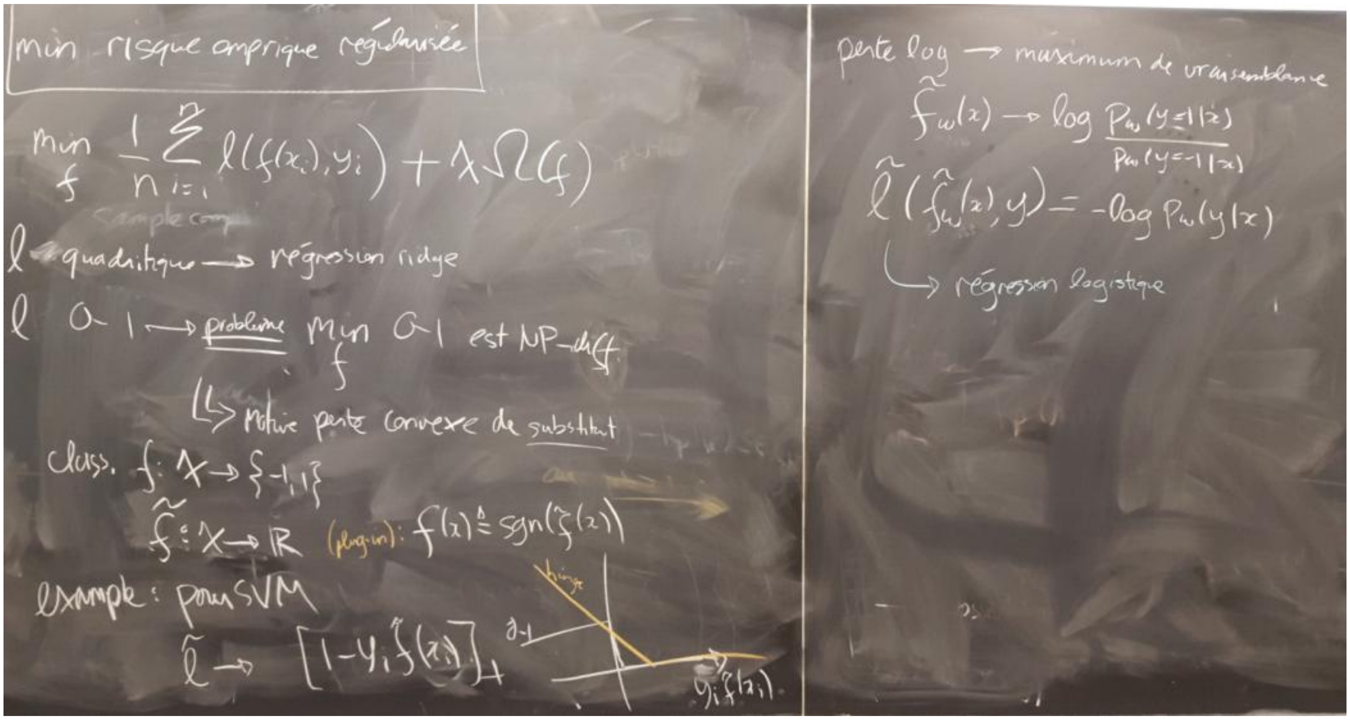
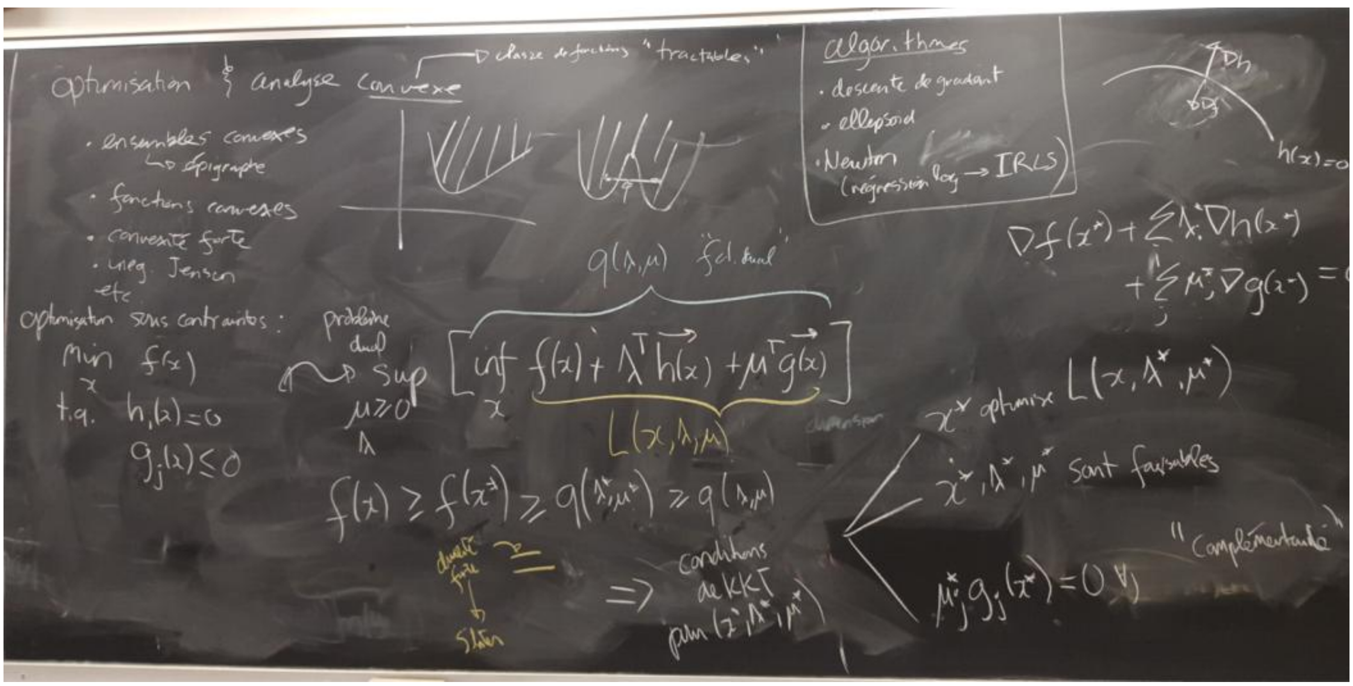


- moyennage local (régression) $\hat f(x) = \sum_{i=1}^{n} W_i(x) Y_i$
  - histogramme (partition) $W_i(x) \propto \mathbb{1}\{X_i \in A(x)\}$ $\sum_i W_i(x) = 1$
  - k p.p.v. $\mathbb{1}\{X_i \in V_k(x)\}$

↳ $\boxed{\text{plug-in}}$ $\hat f(x) = \mathbb{1}\{\hat\eta(x) \ge \frac{1}{2}\}$ $f^*(x) = \mathbb{E}[Y | X = x]$ → pour perte quadratique

$Y_i \in \{0,1\}$

$$\mathbb{E}_{D_n}[R^{0-1}(\hat f)] - R^{0-1}(f^*) \le 2\sqrt{\mathbb{E}[R^{reg}(\hat\eta)] - R^{reg}(\eta^*)}$$


min risque empirique régularisée
$\min_f \frac{1}{n}\sum_{i=1}^{n} \ell(f(x_i), y_i) + \lambda\Omega(f)$
ℓ quadratique → régression ridge
ℓ 0-1 → problème $\min_f$ 0-1 est NP-diff
↳ notre perte convexe de substitut
class. $f: X \to \{-1, 1\}$
$\tilde{f}: X \to \mathbb{R}$ (plug-in): $f(x) \triangleq \mathrm{sgn}(\tilde{f}(x))$
exemple: pour SVM
$\tilde{\ell} \to [1 - y_i \tilde{f}(x_i)]_+$
hinge
0-1
$y_i \tilde{f}(x_i)$
perte log → maximum de vraisemblance
$\tilde{f}_w(x) \to \log \frac{P_w(y=1|x)}{P_w(y=-1|x)}$
$\tilde{\ell}(\hat{f}_w(x), y) = -\log P_w(y|x)$
↳ régression logistique

optimisation & analyse convexe
→ classe de fonctions "tractables"
· ensembles convexes
↳ épigraphe
· fonctions convexes
· convexité forte
· inég. Jensen
etc
optimisation sous contraintes:
$\min_x f(x)$
t.q. $h_i(x) = 0$
$g_j(x) \leq 0$
problème dual
$g(\lambda, \mu)$ "fct. dual"
$\sup_{\mu \geq 0, \lambda} \left[\inf_x \underbrace{f(x) + \lambda^T \vec{h}(x) + \mu^T \vec{g}(x)}_{L(x,\lambda,\mu)}\right]$
$f(x) \geq f(x^*) \geq g(\lambda^*, \mu^*) \geq g(\lambda, \mu)$
dualité forte ≥ =
↓
Slater
⇒ conditions de KKT pour (x^*, λ^*, μ^*)
algorithmes
· descente de gradient
· ellipsoid
· Newton
(régression log → IRLS)
∇h
∇g
$h(x) = 0$
$\nabla f(x^*) + \sum_i \lambda_i^* \nabla h(x^*) + \sum_j \mu_j^* \nabla g(x^*) = 0$
x^* optimise $L(x, \lambda^*, \mu^*)$
x^*, λ^*, μ^* sont faisables
"complémentarité"
$\mu_j^* g_j(x^*) = 0 \ \forall j$

* sélection de modèle / hyperparamètre (λ)

  ∘ validation croisée

  ∘ C_L de Mallows

* générative vs. conditionelle

$p(x|y) = \sum_k \pi_k N(x \mid \mu_k, \Sigma_k)$

$p(x,y) \to$ LDA, QDA

$p(y|x) \to$ régression logistique $\to p(y|x) = \frac{1}{1+\exp(-\tilde{f}(x))}$

* astuce du noyau

$< \vec{\varphi}(x), \vec{\varphi}(x') > = K(x,x')$

$\to$ régression ridge design X: $\underbrace{X^T X}_{d\times d} \leadsto \underbrace{XX^T}_{n\times n} = K$

$\to$ SVM

∘ thm. du représentant

∘ dualité

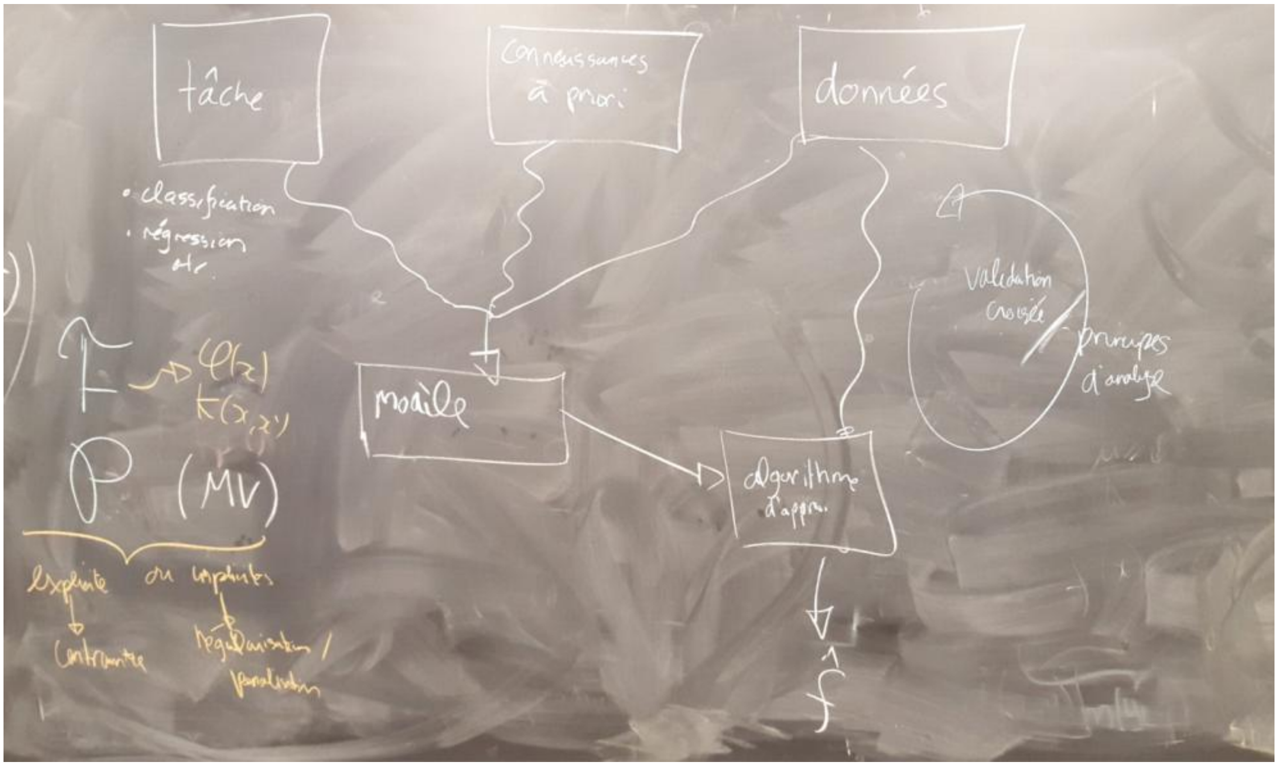
etc..

(RBF $K(x,x') = \exp\left(-\frac{\|x-x'\|^2}{2\sigma^2}\right)$

$\vec{\varphi}(x) = \exp\left(-\frac{\|x-\cdot\|^2}{2\sigma^2}\right)$

$\varphi(x) \in \mathbb{R}^{\mathbb{R}}$ ($\mathbb{R} \to \mathbb{R}$)

$< \varphi(x), \varphi(x') > = \int \exp\left(-\frac{\|x-u\|^2}{2\sigma^2} - \frac{\|x'-u\|^2}{2\sigma^2}\right) du$


tâche
connaissances à priori
données
• classification
• régression
etc
modèle
algorithme d'appr.
validation croisée
principes d'analyse
(MV)
explicite ou implicite
contrainte
régularisation / pénalisation