La dimension de Vapnik-Chervonenkis

Gilles Richard

Retour sur PAC !

Si C fini et il existe A... alors

- C PAC apprenable... et meme plus car...
- borne sur le nombre d'exemples en |C|

Si C infini...

- exemples: C = {les rectangles} ;-)
- pas de borne generale sur le nb d'exemples

Pb: trouver une autre borne;-)

Idee:

- ds le cas des rectangles, 4nbres suffisent!
- remplacer log(|C|) par similaire

LA VC-dimension

- C une classe de parties de X (individus)
- $A \subseteq X$, A fini (nos exemples!)

Def1: $tr(C,A) = tr_C(A) = \{c \cap A \mid c \text{ in } C\} \subseteq 2^A$

Def2: C pulverise A ssi $tr(C, A) = 2^A$

Def3: $VC\text{-}dim(C) = \text{argmax}\{|A| \mid C \text{ pulverise } A\}$

Ex.:

- Intervalles sur la droite reelle (2)
- Demi-plans dans le plan (3)
- Rectangles //axes: 4 (5: on choisit une classe differente pour le pt qui se trouve a l'int. du rect. defini par les 4 autres)
- Polygones a d cotes: 2d +1 (TBD)

Quelques petites proprietes

- **c_a = {x in R| sign(sin(ax))≥ 0}: concept de R**
- **C = {c_a | a in R} (C infini ici)**
- **VC-dim(C) = ∞**
- **|tr(C,A)| ≤ $2^{|m|}$ ou |A|=m**
- **VC-dim(C) mesure la « puissance d'expression » de C**

Prop. 1: Si C finie alors VC-dim(C) ≤ $\log_2(|C|)$

Preuve: le A max est tq : $2^{|A|} \leq |C|$;-)

Un beau resultat!

Prop. 2: VC-dim(C) infinie → NON PAC apprenable

Preuve: (1989)

- Idee de la preuve: par l'absurde et on suppose C PAC.
- m nbre d'ex. necessaires pour PAC apprendre avec ε=0.1 et δ=0.1
- On sait que moyenne erreur <=0.19 (vu avant) et
- On sait qu'il existe S pulverisable de cardinal 2m car VC-dim infinie ;-): on considere ce S comme notre TS
- On considere P uniforme sur TS et 0 ailleurs
- On construit ensuite un concept c tq $P(c(x_i) = 0) = ½$
- Alors l'erreur moyenne est ¼ > 0.19 : contradiction **CQFD**

Et la reciproque?? Si-dim finie ????

La fonction de croissance Π_C

Notons $\Pi_C(m) = \max\{ |tr(C,A)| \mid |A| = m\}$

1. Mesure la “complexite/puissance” de C
2. $\Pi_C(m) = 2^{|m|}$ ssi C pulverise un ensemble de card m
3. VC-dim(C) = max $\{d \mid \Pi_C(d) = 2^{|d|}\}$
4. VC-dim(C) infinie ssi $\forall$ m, $\Pi_C(m) = 2^{|m|}$
5. Croissance exponentielle attendue en fonction de m! et pourtant…

Prop. 3: Si VC-dim(C)=d alors $\Pi_C(m) \leq \Phi_d(m)$

Preuve: (lemme de Sauer – discontinuite ds la fonction de croissance)

- Notons $\Phi_d(m)$ le nbre de parties de moins de d elts ds un ens. a m elements.
- $\Phi_d(m) = \Phi_d(m-1) + \Phi_{d-1}(m-1)$ (explain !)
- D’ou $\Pi_C(m) \leq \Phi_d(m)$ par double induction sur m et d (c’est lourd !)

Prop. 4: m ≤ d alors $\Phi_d(m) = 2^{|m|}$ sinon $O(m^d)$

Preuve:

- (m ≤ d: oui) Si d < m, on multiplie par $(d/m)^d$ $\Phi_d(m)$ et on fait rentrer dans la somme car d/m ≤1: en allant jusqu’a m au lieu de d = dlpt en serie de $(1+x)^m$ ou x = d/m qui est inferieur a e^x
- On elimine $(d/m)^d$ de chaque cote … reste $(em/d)^d$ cqfd

A quoi ca sert?

Notion de ε-reseau S pour c in C (c target concept):

- $\Delta_\varepsilon(c)$: ensemble des erreurs r=cΔc' tq P(r)≥ ε
- r de $\Delta_\varepsilon(c)$, il y a un elt de S dedans
  - un petit exemple C= les intervalles de [0,1]
  - S= kε avec 1 ≤ k≤ ⌊1/ ε ⌋(ie le plus petit entier ≥ 1/ ε)

Soit TS ε-reseau , soit A algo. polynomial capable de retourner h consistant avec c sur TS

alors erreur(h) ≤ ε (i.e. cΔh n'est pas dans $\Delta_\varepsilon(c)$)

Notons que pour TS de taille m

$p(TS \cap r = \emptyset) \leq (1-\varepsilon)^m$ donc p(il existe r…) ≤ $|\Delta_\varepsilon(c)|(1-\varepsilon)^m$

Petite idee de la preuve…

- Event A =“TS n’est pas un ε-reseau “
- But: borner mieux la proba de A :P(($\mu(e) \geq \varepsilon$))<P(A)
- Chercher autre event B tq Pr(A) $\leq$ k Pr(B)
- Intuitivement augmenter m = |TS| ;-)))
- On ajoute TS’ de taille m a TS
- B_r=“TS n’est pas un ε-reseau $\wedge$ TS’ touche r au moins εm/2 fois (ou r fixe non touche par TS)
- B = U B_r pour les r non touches par TS
- Of course P(B) $\leq$ P(A) car B -> A !
- Mais le plus beau est: P(A) $\leq$ 2P(B) pour m = $O(1/\varepsilon)$

On y est … presque

- La notion de trace reapparait ;-)))
- $P(B) \leq |\, tr_{\Delta\varepsilon(c)}(TS \cup TS') \,|\, P(B_r)$

  $\leq |\, tr_{C}(TS \cup TS') \,|\, 2^{-\varepsilon m/2} \leq \Phi_d(2m)\, 2^{-\varepsilon m/2}$

  $\leq (2em/d)^d\, 2^{-\varepsilon m/2}$ (via le lemme de Sauer)
- Donc $P(A) = P(\mu(e) \geq \varepsilon) \leq 2\,(2em/d)^d\, 2^{-\varepsilon m/2}$
- $P(\mu(e) \geq \varepsilon) \leq \delta$ … la meme chanson!

Et le resultat final

- **Prop.3 : C tq VC-dim(C))=d finie**

Si $m \geq c_0(1/\varepsilon \log 1/\delta + d/\varepsilon \log(1/\varepsilon))$ alors tt algo A rendant h consistant avec c est PAC.

- **Prop.4: au moins $\Omega(\text{VC-dim}(C)/\varepsilon)$ exemples pour PAC apprendre la classe C avec erreur $\leq \varepsilon$**

Preuve: on construit un TS de taille d pulverise par C, etc…

Conclusion: borne inferieure sur le nbre d'exemples (difficulte statistique!)

V. Vapnik & A. Chervonenkis?

- Education: Russie (Ouzbekistan)
- Now: Prof. London (Royal Holloway)
- Statistical learning theory – SVM
- Chervonenkis… bureau d'en face;-)

Conclusion

- VC-dim: tres importante
- Si finie, strategie de consistance OK
- Relation risque empirique/risque vrai
- Principe de minimisation du risque empirique (ou structurel) (ERM principle)

MAIS

- SVM VC-dim tres large ;-)
- Neural networks idem...
- ILP... infinie ;-)))

autre concept necessaire... probably ;-)

Induction versus Transduction

- **Cadre general (Faire un schema)**
  - Observations (x1,y1),...(xm,ym)
  - Nouveau cas a predire (x,?)
- **Induction**: 2 etapes (offline)

  particulier ->general ->particulier
  - ex.: SVM, neural networks, Bayesian network
- **Transduction:** 1 etape (online)

  particulier → particulier
  - ex.: k-nn, analogical transduction

Limite de PAC...

- **Plus d'hypothese h... donc**
  - Erreur ε: encore un sens
  - Confiance δ : pas si clair
  - PAC ... pas si interessant
- **Autre modele?**
  - Gamerman et Vogt (2000)
  - On garde ε – on jette δ
- **Cas de la classification**
  - Notion de conformite/non conformite
  - Notion de predicteur conforme