

情報意味論(10) PAC学習と VC次元

理工学部管理工学科
櫻井彰人

ポイント

- 仮説空間が無限集合のときのPAC学習
- VC 次元

PAC 学習

- 仮定: H 有限, 答えを含む; m 個の訓練事例は未知の分布 D からの独立サンプル; 誤差も分布 D で測る; アルゴリズムが決定的であっても事例が分布するため、確率的枠組みを考える
- このとき、少なくとも

$$m \geq \frac{1}{\varepsilon} (\ln |H| + \ln(1/\delta))$$

個の訓練例があれば、 $(1-\delta)$ の確率で、選ぶ仮説の“真の”誤差は高々 ε 。

PAC 学習 1 頁証明

- アルゴリズムは D と整合的な仮説を選ぶ
- 誤差 $\geq \varepsilon$ の仮説が整合的になる確率は高々 $(1-\varepsilon)^m$
- 誤差 $\geq \varepsilon$ の整合的仮説が選ばれる確率は高々 $|H|(1-\varepsilon)^m$
- $(1-\varepsilon) \leq e^{-\varepsilon}$ である故、この確率は高々 $|H|e^{-\varepsilon m}$

$$\delta \leq |H|e^{-\varepsilon m}$$

$$\ln \delta \leq \ln |H| - \varepsilon m$$

$$m \geq \frac{1}{\varepsilon} (\ln |H| + \ln(1/\delta))$$

PAC 学習可能性

- ある概念集合 C が PAC 学習可能であるとは、確率 $1-\delta$ と誤差 ε で学習可能、かつ $1/\delta$, $1/\varepsilon$, n , と $\text{size}(c)$ の多項式時間で学習可能なこと。
 - n は属性数
- これが意味するのは
 - 多項式サンプル複雑度 polynomial sample complexity
 - 多項式計算時間

例:

- ブーリアンリテラルの連言: $|H|=3^n$

$$m \geq \frac{1}{\varepsilon} (n \ln 3 + \ln(1/\delta))$$

整合的な仮説は多項式時間で見つけることができる

- バイアスのない学習者: $|H|=2^{2^n}$

$$m \geq \frac{1}{\varepsilon} (2^n \ln 2 + \ln(1/\delta))$$

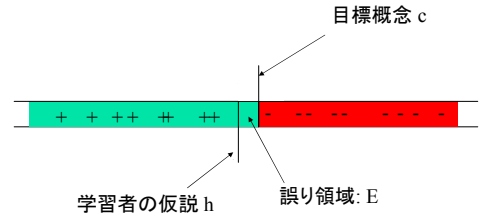
例(続)

- k-項の DNF:
 $T_1 \vee T_2 \vee \dots \vee T_k$
- この場合: $|H| \leq (3^n)^k$

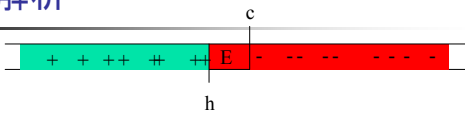
$$m \geq \frac{1}{\varepsilon} (kn \ln 3 + \ln(1/\delta))$$

- ゆえに k-項DNFs は PAC 学習可能?
- No! 整合的仮説を求める計算時間が NP-hard

仮説空間が無限集合だったら?



解析

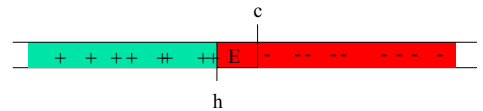


- 仮説 h の誤差: $P(h \neq c) = P(E)$
- $P(E) \geq \varepsilon$ となる確率は高々 $\delta \leq (1-\varepsilon)^m$ (どちらも)
- よって, 確率 $\delta \leq 2(1-\varepsilon)^m$ で誤差 $\geq \varepsilon$
- すなわち:

$$m \geq \frac{\ln(\delta/2)}{\ln(1-\varepsilon)} \geq \frac{\ln(\delta/2)}{\ln(e^{-\varepsilon})} = \frac{1}{\varepsilon} \ln(2/\delta)$$

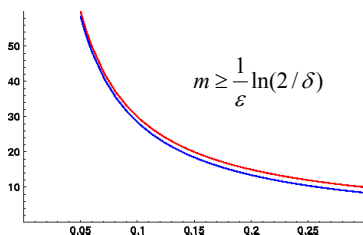
まとめ

- 仮説が無限集合の場合の基礎的結果!



$$m \geq \frac{1}{\varepsilon} \ln(2/\delta)$$

この限界評価の悪さは?



これは一般化できるか?

- 一次元の区間?
- 一般の概念空間?

Shattering:

仮説空間バイアス具合の表現

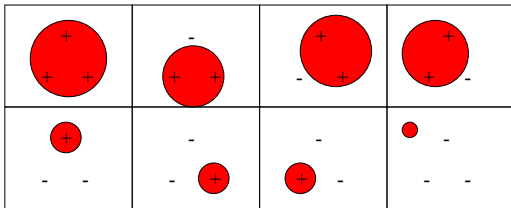
- 事例集合 S が仮説空間 H により shatter されるというのは、 S のどの2分割に対しても、それと整合的な仮説が H のなかにあること

例: Shattering

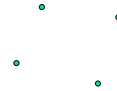


H を全部の円の集合だとする。この図の点集合は H で shatter できるか？

例: Shattering



これはどうだろう？



次にこれは？



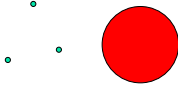
VC次元: $VC(H)$

仮説空間 H の VC次元 $VC(H)$ は、 H によって shatter される点の個数の最大値

(すなわち: H がバイアスをもたない点の最大個数)

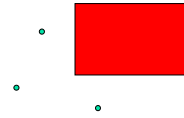
(補足: shatter できないとは、ある分割が H では表現できないということ、すなわち、バイアスがあるがためある事実(当該分割)が見えない)

例: 円の場合



- VC(H) = 3, というのも 3 点は shatter できるが、4 点はやできない

例: 軸並行な長方形



$$VC(H) = 4$$

言い忘れたが、一般位置にない4点は考えなくてよい

例: 1次元階段関数



VC(H) = 1 if “下がるだけ” のとき

VC(H) = 2 if “上がる” も “下がる” もあるとき

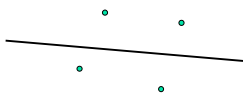
例: 区間



VC(H) = 2 if “間” だけのとき

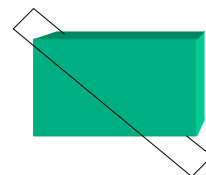
VC(H) = 3 if “間” も “間以外” もあるとき

例: 2D の線形決定境界



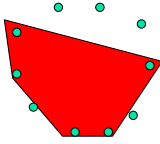
$$VC(H) = 3$$

例: n-次元線形決定境界



$$VC(H) = n+1$$

VC(H)=∞ はありうるか?



Yes! 凸多角形全部の集合

VC次元と PAC-学習

- 定理: 訓練例が m 個

$$m \geq \frac{1}{\varepsilon} (4 \log_2(2/\delta) + 8VC(H) \log_2(13/\varepsilon))$$

あれば、学習機械は、確率 $1-\delta$ で、誤差が高々 ε の仮説を生成することができる。

V=Vapnik, C=Chervonenkis

最初の例にもどると...



- VC(H)=2, であれば

$$m \geq \frac{1}{\varepsilon} (4 \log_2(2/\delta) + 16 \log_2(13/\varepsilon))$$

けれども上界としては、

