Výpočtové učenie

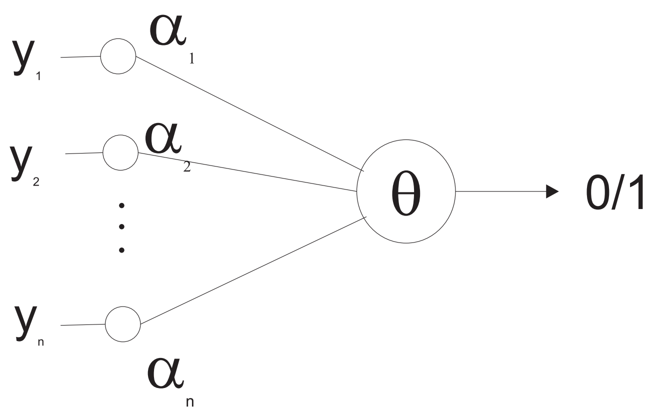
26. apríla 2006

Kapitola 1

VC Dimenzia

1.1 Rastová funkcia

Vapnik, Chervonenkis - 1971
Najvýznamnejší pojem pre teóriu výpočtového učenia.



Obrázok 1.1: Perceptrón

Perceptrón
Pre stav $\omega = (\alpha_1, \alpha_2, \ldots, \alpha_n \Theta)$ funkcia $h_\Omega \in H$ z $X = \mathbb{R}^n$ do $\{0, 1\}$ je daná

$$h_\omega(y) = \begin{cases} 1, & \text{ak } \sum_{i=1}^{n} \alpha_i y_i \geq \Theta \\ 0, & \text{inak} \end{cases}$$

$\omega \vdash h_\omega$ nie je injekcia;
pre ľubovoľné $\lambda \in \mathbb{R}$, stav $\lambda * \omega$ definuje tú istú funkciu.

$\Pi_H(x)$ =počet klasifikácií podľa H, t.j. počet rôznych vektorov tvaru

$$(h(\overline{x_1}), \ldots, h(\overline{x_m}))$$

kde h prebieha všetky hypotézy v H, $\overline{\mathbf{x}} = (\overline{x_1}, \ldots, \overline{x_m})$.

H môže byť nekonečný $\ldots H|E_{\overline{x}}$, $\quad E_x = \{\overline{x_1} \ldots \overline{x_m}\}$ je konečný a má kardinalitu $\Pi_H(\overline{\mathbf{x}})$.

$$\Pi_H(\overline{\mathbf{x}}) \leq 2^m$$

Definícia 1.1.1 ***Rastová funkcia*** *je*

$$\Pi_H(m) = \max\{\Pi_H(\overline{\mathbf{x}}) : \quad \overline{x} \in X^m\}$$

Definícia 1.1.2 *Hovoríme, že vzorka* $\overline{\mathbf{x}}$ *dĺžky* m ***je rozbitá podľa H*** *alebo* ***H rozbíja*** $\overline{\mathbf{x}}$*, ak počet možných klasifikácií podľa* H *je* 2^m*, t.j.* H *poskytuje všetky možné klasifikácie.*

- Ak príklady v $\overline{\mathbf{x}}$ nie sú rôzne, tak $\overline{\mathbf{x}}$ nemôže byť rozbitá.
- Ak príklady v $\overline{\mathbf{x}}$ sú rôzne, x je rozbitá podľa $H \Leftrightarrow$ ak pre ľubovoľné $S \subseteq E_x$ existuje nejaká hypotéza $h \in H$ taká, že pre ľubovoľné $1 \leq i \leq m$ platí

$$h(\overline{x_i}) = 1 \Leftrightarrow \overline{x_i} \in S.$$

  S je potom podmnožina sústreďujúca kladné príklady.

Definícia 1.1.3 ***VC dimenzia*** *H je maximálna dĺžka vzorky, ktorú H rozbíja. Ak neexistuje, hovoríme, že VC je* ∞*.*

$$VCdim(H) = \max\{m : \quad \Pi_H(m) = 2^m\}$$

Príklady:

1. $X = \mathbb{R}$, $H \subseteq$ lúče; $x = (x_1, \ldots, x_m)$ je tréningová vzorka, $x_1 < x_2 < \ldots < x_m$.
   Pre $\Theta \in \mathbb{R}$, $\quad r_\Theta = 1 \quad \Leftrightarrow \quad x_i \geq \Theta$.
   Množina klasifikovaných vektorov $\ldots m + 1$.
   $(0 \ldots 00), (0 \ldots 01), \ldots (1 \ldots 11)$
   ľubovoľná vzorka s rôznymi príkladmi má jeden z týchto klas. vektorov (spermutovaný). $\Rightarrow$ $\Pi_H(m) = m + 1$
   V prípade rovnakých príkladov počet klasifikácií je menší.

2. r_Θ, H je priestor lúčov. Nech je daná tréningová vzorka (x_1, x_2) dĺžky 2, $x_1 < x_2$.
   $\Rightarrow$ neexistuje lúč taký, že $h(x_1) = 1$ a $h(x_2) = 0$, pretože by muselo platiť $x2 < \Theta \leq x_1$.
   $\Rightarrow$ H nerozbíja žiadnu vzorku dĺžky 2. H rozbíja ľubovoľnú vzorku dĺžky 1 $\Rightarrow$ $VCdim(H) = 1$.

3. Nech $X = \mathbb{R}^2$. H je hypotézový priestor perceptrónu P_2, $x = (x_1, x_2, x_3)$ sú tri nekolineárne rôzne body. X je rozbitá pomocou $H \Leftrightarrow$ ak pre ľubovoľnú podmnožinu $S \subseteq E_x = \{x_1, x_2, x_3\}$ platí, že S a E_x sú separovateľné.
   $\Rightarrow$ $VCdim(H) \geq 3$ (P_2 rozbíja 3 nekolineárne body.)
   Rovnosť dokážeme tým, že nájdeme vzorku, ktorej príklady sa nedajú separovať $\Rightarrow$ existuje vzorka dĺžky 4, ktorú P_2 nerozbíja.

Tvrdenie 1 *Ak* H *je konečný hypotézový priestor, tak* $VCdim(H) \leq \ln |H|$.

Dôkaz:
Počet klasifikácií podľa konečného H vzorky ľubovoľnej dĺžky je najviac počet rôznych hypotéz v H
$\Rightarrow$ $\Pi_H(m) \leq |H|$
VCdim je najväčšie d, pre ktoré $\Pi_H(d) = 2^d$.

$$2^d = \Pi_H(d) \leq |H| \quad \Rightarrow \quad d = \Pi_H(d) \leq \ln |H|.$$

□

Príklad: Urobíme odhad pre $VCdim(M_n)$. Platí $|M_n| = 3^n$, M_n sú monočleny.
$VCdim(M_n) \leq (\ln 3) * n$ je horná hranica. Aká je dolná hranica? Pokúsime sa dokázať, že je rovná n.
Tvrdíme, že M_n rozbíja každú vzorku dĺžky n:
$(e_1, e_2, \ldots, e_n)$, kde $e_i = (0 \ldots 010 \ldots 0)$, kde 1 je na i-tom mieste, $1 \leq i \leq n$.
Predpokladajme, že $(q_1, \ldots, q_n) = q \in \{0, 1\}^n$. Ukážeme, že existuje $h \in M_n$ také, že $h(e_i) = q_i$ pre $1 \leq i \leq n$.
Ak $q = (1 \ldots 1)$ vezmeme za h prázdnu hypotézu.
Ak $q = (0 \ldots 00 \ldots 01)$ tak h vytvoríme ako konjunkciu negácií literálov, pre ktoré $q_j = 0$.
$\Rightarrow$ $VCdim(M_n) \geq n$ pre ľubovoľné n.

1.2 VC dimenzia reálnych perceptrónov

Veta 1 *Pre ľubovoľné n nech P_n je reálny perceptrón s n vstupmi. Potom*

$$VCdim(P_n) = n + 1.$$

Dôkaz: P_n je v stave $\omega = (\alpha_1 \alpha_2 \ldots \alpha_n \Theta)$
h_ω - funkcia, ktorú perceptrón počíta

$$h_\omega(y) = 1 \quad \Leftrightarrow \quad \alpha_1 y_1 + \ldots + \alpha_n y_n \geq \Theta.$$

Označme

$$l_\omega^+ = \{y \in \mathbb{R}^n \ : \ \sum_{i=1}^{n} \alpha_i y_i \geq \Theta\}$$

$$l_\omega^- = \{y \in \mathbb{R}^n \ : \ \sum_{i=1}^{n} \alpha_i y_i < \Theta\}$$

$$l_\omega = \{y \in \mathbb{R}^n \ : \ \sum_{i=1}^{n} \alpha_i y_i = \Theta\}$$

$C \subseteq \mathbb{R}^n$ je **konvexná**, ak pre ľubovoľné $x, y \in C$ a ľubovoľné reálne číslo λ, $0 \leq \lambda \leq 1$, bod $\lambda x + (1-\lambda) y \in C$.

Prienik ľubovoľných 2 konvexných množín v $\mathbb{R}^n$ je konvexná množina. Pre ľubovoľnú množinu bodov $S \subseteq \mathbb{R}^n$ existuje najmenšia konvexná množina obsahujúca S;

$conv(S)$ ...konvexný obal S je prienik všetkých konvexných množín obsahujúcich S;

Pripomeňme Radonovu vetu:

Veta 2 *Nech n je kladné celé číslo, E je ľubovoľná množina n+2 bodov z $\mathbb{R}^n$. Potom existuje $\emptyset \neq S \subseteq E$ taká, že*

$$conv(S) \cap conv(E \setminus S) \neq \emptyset.$$

Nech $x = \underbrace{(x_1 \ldots x_{n+2})}_{rozne}$ je ľubovoľná vzorka príkladov z $\mathbb{R}^n$ dĺžky n+2.

E_x je množina príkladov v x $\ldots |E_x| = n+2$. Podľa Radonovej vety existuje $S \neq \emptyset$, $S \subseteq E_x$

$$conv(S) \cap conv(E_x \setminus S) \neq \emptyset.$$

Predpokladajme, že existuje h_{omega} v P_n také, že S je množina pozitívnych príkladov h_ω v E_x.

$$\Rightarrow S \subseteq l_\omega^+,\ E_x \setminus S \subseteq l_\omega^-$$

Pretože uzavretý polpriestor l_ω^+ a otvorený l_ω^- sú disjunktné a sú konvexné v $\mathbb{R}^n$ $\Rightarrow$ $conv(S) \subseteq l_\omega^+$, $conv(E_x \setminus S) \subseteq l_\omega^-$
$conv(S) \cap conv(E \setminus S) \subseteq conv(S) \cap conv(E_x \setminus S) = \emptyset$

$\Rightarrow$ neexistuje taká h_ω, a preto x nie je rozbitá p_n. $\Rightarrow$ žiadna vzorka dĺžky n+2 nie je rozbitá pomocou p_n.

$\Rightarrow VCdim(P_n) \leq n+1$.

Opačná nerovnosť: $o \in \mathbb{R}^n$, $o = (0, 0 \ldots 0)$, $e_i = (0 \ldots 10 \ldots 0)$, $1 \leq i \leq n$. Ukážeme, že P_n rozbíja $x = (o, e_1, \ldots e_n)$ dĺžky $n+1$.

Nech $S \subseteq E_x = \{o, e_1 \ldots e_n\}$. Nech

$$\alpha i = \begin{cases} 1, & \text{ak } e_i \in S \\ -1, & \text{ak } e_i \notin S \end{cases}$$

$$\Theta = \begin{cases} -\frac{1}{2}, & \text{ak } o \in S \\ +\frac{1}{2}, & \text{ak } o \notin S \end{cases}$$

Priamou verifikáciou, ak $\omega = (\alpha_1 \ldots \alpha_n \Theta)$ je stav P_n, potom množina pozitívnych príkladov h_ω je práve $S \Rightarrow VCdim(P_n) \geq n+1$. $\square$

1.3 Sauerova lema

Rastová funkcia - miera počtu rôznych klasifikácií vzorky dĺžky m na pozitívne a negatívne príklady podľa H, pokiaľ $VCdim\ H$ je max. hodnota m, pre ktorú platí $\Pi_H(m) = 2^m$.

Veta 3 ***(Sauerova lema):*** *Nech $d \geq 0$ a $m \geq 1$ sú celé kladné čísla, nech H je hypotézový priestor s $VCdim(H) = d$. Potom*

$$\Pi_H(m) \leq \underbrace{1 + \binom{m}{1} + \binom{m}{2} + \ldots + \binom{m}{d}}_{\Phi(d,m)}.$$

Binomické koeficienty spĺňajú

$$\binom{a}{b} = \binom{a-1}{b} + \binom{a-1}{b-1}$$

Zaveďme funkciu:

$$\Phi(0,m) = 1 \quad (m \geq 1);$$

$$\Phi(d,1) = 2 \quad (d \geq 1)$$

$$\Phi(d,m) = \Phi(d,m-1) + \Phi(d-1,m-1), \quad (d \geq 1\, m \geq 2)$$

Dôkaz: Ak $VCdim(H) = d = 0$, potom príklad x-ľub, $h(x)$ je rovnaké (buď 0 alebo 1) pre ľub. hypotézu $h \in H$. $\Rightarrow$ $\Pi_H(x) = 1$ pre ľub. vzorku dĺžky m $\Rightarrow \Pi_H(m) = 1 = \Phi(0,m)$ $\Rightarrow$ veta platí pre $d = 0$.
Ak $m = 1$ a $d \geq 1 \Rightarrow$ $\Pi_H(1) \leq 2 = \Phi(d,1)$

Indukciou na $d + m$:

- Prípad $d + m = 2$ je dokázaný.
- Predpokladajme, že veta platí pre $d + m \leq k$, kde $k \geq 2$ a nech H je hypotézový priestor s $VCdim = d$, x je trénovacia vzorka dĺžky m, kde $d + m = k + 1$.
  Prípady $(d,m) = (0, k+1)$ a $(d,m) = (k,1)$ sú už dokázané.
  Nech $d \geq 1$, $m \geq 2$, x obsahuje rôzne príklady, E je množina príkladov v x, $H_E = H|E$ je obmedzenie hypotéz H na E.
  $\Rightarrow$ H_E je konečný a $\Pi_H(x) = |H_E|$.
  Potrebujeme ukázať, že $|H_E| \leq \Phi(d,m)$.
  Nech $F = E \setminus \{x_m\}$, $H_F = H|F$.
  Dve rôzne hypotézy $h, g \in H_E$ pri obmedzení na F dávajú tú istú hypotézu z H_F práve vtedy, keď sa zhodujú na F a nezhodujú na x_m.
  H_* je množina všetkých hypotéz, ktoré vzniknú takto:
  Ak $h^* \in H_*$, tak sú možné obe rozšírenia h^* na funkciu na E ... hypotézu z H_E. h^* je rozšírenie
  $\Rightarrow$ $|H_E| = |H_F| + |H_*|$.
  Nech $x' = (x_1 \ldots x_{m-1})$. Potom

  $$|H_F| = \Pi_H(x') \leq \Pi_H(m-1) \leq \Phi(d,m-1)$$

  pretože $d + (m-1) \leq k$.
  Tvrdíme, že $VCdim(H_*)$ je najviac $d-1$. Ak by $VCdim(H_*) = d \Rightarrow h^*$ rozbíja nejakú vzorku $z = (z_1 \ldots z_d)$ dĺžky d príkladov z F.
  Pre

  $$h^* \in H_* \begin{cases} h_1 \in H_E \ldots & h_1(x_m) = 0 \\ h_2 \in H_E \ldots & h_2(x_m) = 1 \end{cases}$$

  $\Rightarrow$ H_E a teda H rozbíja vzorku $(z_1 \ldots z_d x_m)$ dĺžky $d+1$, čo je v spore s $VCdim(H) \leq d$.
  $\Rightarrow$ $VCdim(H_*) \leq d - 1$.

  Použitím indukčnej hypotézy

  $$|H_*| = \Pi_{H_*}(x'0) \leq \Pi_{H_*}(m-1) \leq \Phi(d-1,m-1)$$

  pretože $(d-1) + (m-1) \leq k$. Kombináciou oboch výsledkov dostaneme

  $$\Pi_H(x) = |H_E| = |H_F| + |H_*| \leq \Phi(d,m-1) + \Phi(d-1,m-1) = \Phi(d,m).$$

  □

Tvrdenie 2 *Pre všetky $m \geq d \geq 1$ platí $\Phi(d, m) < \left(\frac{e \cdot m}{d}\right)^d$.*

Tvrdenie 3 *Nech H je ľubovoľný hypotézový priestor obsahujúci aspoň 2 hypotézy a definovaný na konečnom príkladovom priestore X, potom $VCdim(H) > \frac{\ln |H|}{1 - \ln |X|}$.*

1.4 Úlohy k VC-dim

1. Ukážte, že ak $X = \mathrm{R}$ a H je množina všetkých uzavretých intervalov, tak $\Pi_H(m) = 1 + m + \frac{1}{2}m(m-1)$.

2. Popíšte expl. hypotézový priestor P_1 a ukážte, že $VCdim(P_1) = 2$.

3. Ukážte, že ak H je hypotézový priestor reálneho perceptrónu P_2, tak $\Pi_H(4) = 14$.

4. Nech H má konečnú $VCdim$. Pre $h \in H$ definujme $\overline{h}$

$$\overline{h} = 1 \quad \Leftrightarrow \quad h(x) = 0$$

   a nech komplement H je priestor $\{\overline{h} : \quad h \in H\}$. Dokážte, že majú oba priestory rovnakú VC dimenziu.

5. Dokážte
   (a) $\Phi(d, m) = \Phi(d, m-1) + \Phi(d-1, m-1), \quad d \geq 1, m \geq 2$
   (b) $\Phi(d, m) \leq m^d, \quad m \geq d > 1$.

6. Monočlen je monotónny, ak neobsahuje žiadne negované literály. Dokážte, že priestor monotónnych monočlenov definovaný na $\{0, 1\}^*$ má VC dimenziu práve n.

7. Hypotézový priestor H je lineárne usporiadaný, ak má aspoň 2 hypotézy a ak pre ľubovoľné dve $h, g \in H$ platí
   buď $h(x) = 1 \Rightarrow g(x) = 1$
   alebo $g(x) = 1 \Rightarrow h(x) = 1$
   Dokážte, že ak H je lineárne usporiadaný, $VCdim(H) = 1$.

8. Nech G_n je množina hypotéz z P_n, pre ktoré nulový vektor o je negatívny príklad. Predpokladajme, že vzorka $x = (x_1 \dots x_m)$ je rozbitá pomocou G_n. Prečo sa žiadne x_i nesmie rovnať 0? Dokážte, že vzorka $(x_1 \dots x_m, 0)$ je rozbitá pomocou P_n. Dokážte, že $VCdim(G_n) = n$.

Kapitola 2

Učenie a VC dimenzia

2.1 Úvod

Ukážeme, že pre ľub. hypotézový priestor H nutnou a postačujúcou podmienkou pre jeho potenciálnu naučiteľnosť je konečná VC dimenzia.

Máme potenciálne naučiteľné hypotézové priestory práve tie, ktoré majú konečnú VC dimenziu?

2.2 VC dimenzia a potenciálna naučiteľnosť

Označenie: $\overline{s} = (\overline{x}, \overline{b})$ pre $\overline{s} = ((x_1, b_1) \dots (x_m, b_m)) \in (X \times \{0,1\})^m$ Ak t je cieľový koncept a $\overline{s}$ je tréningová vzorka pre t, potom $\overline{s}$ označíme $(\overline{x}, t(\overline{x}))$.

Toto zdôrazňuje fakt, že keď $\overline{s} \in S(m,t)$ sú dané len hodnoty t na prvkoch tréningovej vzorky $\overline{x}$.

$H[\overline{x}, t]$ podmnožinu H, ktorá je v súlade s $\overline{s}$ $H[(\overline{x}, t(\overline{x}))]$

Pozorovaná chyba hypotézy $h \in H$ na $\overline{s}$

$$er_{\overline{s}}(h) = \frac{1}{m} |\{i : h(x_i) \neq b_i\}|$$

$H[\overline{s}] = \{h \in H | \quad h(x_i) = t(x_i), \quad 1 \leq i \leq m\}, H[\overline{s}]$ je množina hypotéz, ktoré majú nulovú chybu na $\overline{s}$.

Ak $\overline{s} = (\overline{x}, t(\overline{x}))$, potom $er_{\overline{s}}(h)$ sa nazýva pozorovaná chyba h na $\overline{x}$ vzhľadom na t $er_{\overline{s}}(h,t)$, alebo $er_{\overline{s}}(h)$, ak t je zrejmé.

Veta 4 *Ak hypotézový priestor má nekonečnú VC dimenziu, tak nie je potenciálne naučiteľný.*

Dôkaz:

Nech H má nekonečnú VC dimenziu, t.j. pre ľub. $m > 0$ existuje vzorka $\overline{z}$ dĺžky 2m, ktorá je rozbitá podľa H.

Nech $E = E_{\overline{z}}$ je množina príkladov v tejto vzorke a definujme rozdelenie pravdepodobnosti μ na X

$$\mu(x) = \begin{cases} \frac{1}{2m} & \text{pre } x \in E \\ 0 & \text{inak} \end{cases}$$

Vzhľadom na μ každý príklad v E má rovnakú pravdepodobnosť prezentácie a ostatné príklady majú pravdepodobnosť 0.

$$\mu^m \dots E^m \qquad \text{a 0 inde}$$

Teda s pravdepodobnosťou 1 náhodne vybratá vzorka $\overline{s}$ dĺžky m je vzorkou príkladov z E.

Nech $\overline{s} = (\overline{x}, t(\overline{x})) \in S(m,t)$. S pravdepodobosťou 1 (vzhľadom μ^m) máme $x_i \in E$ pre $1 \leq i \leq m$. Pretože $\overline{z}$ je rozbitá podľa H, ex. hypotéza $h \in H$ taká ,že

$$h(x_i) = t(x_i) \text{ pre } 1 \leq i \leq m \qquad \text{a}$$

$$h(x) \neq t(x) \text{ pre vsetky ostatne } x \in E$$

Z toho vyplýva, že $h \in H[\overline{s}]$: h má chybu aspoň $\frac{1}{2}$ vzhľadom na t.

Ukázali sme, že pre ľub. $m > 0$, ľub. t ex. pravdep. r oz. μ na X také, že udalosť $H[\overline{s}] \bigcap B_{\frac{1}{2}} = \emptyset$ má pravdepodobnosť 0.

Teda neex. žiadne $m_0 > 0$, $m_0 = m_0(\frac{1}{2}, \frac{1}{2})$ pre ktoré by sme mohli tvrdiť, že pre všetky $m > m_0 \mu^m \{\overline{s} \in S(m,t) | H[s] \cap B_{\frac{1}{2}} = \emptyset\} > \frac{1}{2}$.

$\Rightarrow$ H nie je potenciálne naučiteľné. $\square$

Príklad: Pre každú $A \subseteq R$ definujeme charakteristickú funkciu

$$\chi_A : \chi_A(y) = \begin{cases} 1 & \text{ak } y \in A \\ 0 & \text{inak} \end{cases}$$

U-systém všetkých podmnožín R takých, ktoré sú konečnými zjednoteniami uzavretých intervalov

$J = \{\chi_A : A \in U\}$ - priestor zjedn. intervalov

VCdim(J) je nekonečná!

$\overline{x} = (x_1 \dots x_m)$ je trén. vzorka príkladov z R, E_x je odp. množ. príkladov $S \subseteq E_x$; k S je možné skonštruovať $A \in U$ tak, že $S \subseteq A$ a $(E_{\overline{x}} \mathrm{S}) \bigcap A = \emptyset$ takto:

Pre každé $x_i \in S$ nech A_i je inerval, ktorý obsahuje x_i ale neobsahuje žiadny iný prvok $x \in E_{\overline{x}}$.

Nech $A = \bigcup A_i$, A je konečné zjednotenie

$$\chi_A = 1 \qquad na \qquad S$$

$$\chi_A = 0 \qquad na \qquad E_{\overline{x}} - S$$

Inak povedané, J rozbíja $\overline{x}$.

Pretože tento argument je použiteľný a platí len pre konečné vzorky $\Rightarrow$ VC dim(J) je nekonečná.

<u>*Poznamka*</u> :

J je obsiahnutý v priestore všetkých uzavretých podm. R (char. funkcií).

Čo by sa dalo povedať o VC dim tohto všeobec. priestoru?

Jeho VC dim je nekonečná.

H-hypotézový priestor def. na príkl. priestore X

t-cieľový koncept v H

μ - pravdepodobnostné rozdelenie na X

ϵ - ľub. r.č. $0 \leq \epsilon \leq 1$

Nech t, μ, ϵ - fixné, ale ľubovoľné

Definujme

$$Q_m^\epsilon = \{x \in X^m : H[x,t] \cap B_\epsilon \neq \emptyset\}$$

Pravdepodobnosť výberu trén. vzorky, pre ktorú existuje konzist., ale ϵ-zlá hypotéza je

$$\mu^m = \{\overline{s} \in S(m,t) : H[\overline{s}] \cap B_\epsilon \neq \emptyset\}$$

čo je vlastne $\mu^m(Q_m^\epsilon)$.

Teda ukázať, že H je potenciálne naučiteľný znamená nájsť hornú hranicu $f(m, \epsilon)$ pre $\mu^m(Q_m^\epsilon)$, ktorá je nezávislá od t a μ a ktorá ide k 0, keď $m \to \infty$

Potom

$$\forall (0 < \delta < 1) \exists (m_0) \forall (m \geq m_0) \qquad f(m, \epsilon) < \delta$$

$m_0 = m_0(\delta, \epsilon)$, m_0 je tiež hornou hranicou pre veľkosť trén. vzorky.

Tvrdenie: Nech H je hypotézový priestor definovaný na príkladovom priestore X, t, μ, ϵ sú ľub. ale pevne dané. Potom

$$\mu^m\{\overline{s} \in S(m,t) : H[\overline{s}] \cap B_\epsilon \neq \emptyset\} < \underbrace{2\Pi_H(2m)2^{\frac{-\epsilon m}{2}}}_{f(m,\epsilon)}$$

pre vš. $m \geq \frac{8}{\epsilon}$.

$$\Pi_H(m) \leq 1 + \binom{m}{1} + \binom{m}{2} + \ldots + \binom{m}{d} \qquad \text{polynóm}$$

$$lim_{n\to\infty} 2\Pi_H(2m)2^{\frac{-\epsilon m}{2}} \to 0 \qquad ?$$

Kľúčový výsledok: VC dim implikuje potenciálnu naučiteľnosť.

Veta 5 *Ak má hypotézový priestor konečnú VC dimenziu, potom je potenciálne naučiteľný.*

D: 4 fázy

- Ohraničiť $\mu^m(Q_m^\epsilon)$ pravdepodobnosťou určitej podmnožiny R_m^ϵ v X^{2m}
- Použitím skupiny akcií ohraničiť pravdepodobnosť R_m^ϵ pomocou konečných výrazov
- Vyjadriť túto hranicu pomocou Π_H - komb. argumentami
- Použiť argument posl. lemy na vyjadrenie výsledku, že $\mu^m(Q_m^\epsilon) \to 0$, keď $m \to \infty$

1.fáza:
$\overline{x}, \overline{y} \in X^m$, $\overline{xy} \in X^{2m}$ - je trén. vzorka dĺžky 2m
$R_m^\epsilon = \{\overline{xy} \in X^{2m} : \exists h \in B_\epsilon$ pre ktorú $er_{\overline{x}}(h) = 0$ a $er_{\overline{y}}(h) > \frac{\epsilon}{2}\}$
Lema 1: Pre vš. $m \leq \frac{8}{\epsilon}$ platí

$$\mu^m(Q_m^\epsilon) \geq 2\mu^{2m}(R_m^\epsilon).$$

D:
$\chi_Q \ldots$ charakteristická funkcia Q_m^ϵ:
$\chi_Q(\overline{x}) = 1$, ak $\overline{x} \in Q_m^\epsilon$
$\chi_Q(\overline{x}) = 0$, inak.
$\chi_R(\overline{xy}) = \chi_Q(\overline{x})\psi_{\overline{x}}(\overline{y})$, kde

$$\psi_{\overline{x}}(\overline{y}) = \begin{cases} 1 & \text{ak } \exists h \in H[\overline{x}] \cap B_\epsilon \text{ s } er_{\overline{y}}(h) > \frac{\epsilon}{2} \\ 0 & \text{inak} \end{cases}$$

$$\mu^{2m}(R_m^\epsilon) = \int \chi_R(\overline{xy} = \int (\chi_Q(\overline{x} \int \psi_{\overline{x}}(\overline{y}))$$

$\int$ sú cez celé relevantné priestory;
Vnútorný integrál je pravdepodobnosť, že pri danom $\overline{x}$ existuje $h \in B_\epsilon$, ktoré je konzistentné s $\overline{x}$ a splňuje $er_{\overline{y}}(h) > \frac{\epsilon}{2}$.
Tento integrál určite nie je menej než pravdep., že partikulárne $h \in B_\epsilon$ konzist. s $\overline{x}$ splňuje $er_{\overline{y}}(H) > \frac{\epsilon}{2}$.

$$?? \Rightarrow \mu^{2m}(R_m^\epsilon) \geq \int \frac{1}{2}\chi_Q(\overline{x}) = \frac{1}{2}\mu^m(Q_m^\epsilon)$$

Chernovova hranica (bin. rozdelenie)
$0 \leq p \leq 1$, LE(p,m,s)-pravdepodobnosť najviac s úspechov v m nezávislých pokusoch, z ktorých každý má pravdepodob. p.

$$\text{LE(p,m,(1-}\beta\text{)mp)} \leq e^{\frac{-\beta^2 mp}{2}} \qquad \text{pre ľub.} \qquad 0 \leq \beta \leq 1.$$

Nech $h \in B_\epsilon$, $er_\mu(h) = \epsilon_h > \epsilon$. Pre $\overline{y} \in X^m$, $mer_{overliney}(h)$ označuje počet komponentov, na ktorých t a h sa líšia = je to bin. rozdelená náh. prem.

$$\mu^m\{\overline{y} : er_{\overline{y}}(h) \leq \frac{\epsilon}{2}\} = \mu^m\{\overline{y} : mer_{\overline{y}}(h) \leq \frac{\epsilon}{2}m\}$$

$$\leq \mu^m\{\overline{y} : mer_{\overline{y}}(h) \leq \frac{\epsilon_h}{2}m\} = LE(\epsilon_h, m, (1-\frac{1}{2})m\epsilon_h)$$

$$\leq exp(\frac{-\epsilon_h m}{8}) < exp(\frac{-\epsilon m}{8})$$

??

2.fáza:

Nech $i \in \{1, \ldots m\}$

τ_i je permutácia, ktorá vymení i a m+i tý prvok;

Nech G_m je grupa generovaná permutáciami $\tau_i\{1 \leq i \leq m\}$ $|G_m| = 2^m$

Lema 2: Pre dané $\overline{z} \in X^{2m}$ nech $\Gamma(\overline{z})$ označuje počet $\sigma \in G_m$, pre ktoré $\sigma\overline{z} \in R^{\epsilon}_m$. Potom $|G_m|\mu^{2m}(R^{\epsilon}_m) \leq max\Gamma(\overline{z})$ kde max je cez vš. $\overline{z} \in X^{2m}$.

3.fáza:

Pre ľub. $h \in B_\epsilon$ nech

$$R^{\epsilon}_m(h) = \{\overline{xy} \in X^{2m} : er_{\overline{x}}(h) = 0 \qquad \text{a} \qquad er_{\overline{y}}(h) > \frac{\epsilon}{2}\}$$

Pre $\overline{z} \in X^{2m}$, $\Gamma(h, \overline{z})$ počet $\sigma \in G_m$, ktoré transformujú $\overline{z}$ na vektor v $R^{\epsilon}_m(h)$.

Lema 3: $m > 0$, $h \in B_\epsilon$. Potom $\Gamma(h, \overline{z}) < 2^{m(1-\frac{\epsilon}{2})}$ pre vš. $\overline{z} \in X^{2m}$.

Lema 4: Pre ľub. $m > 0$ $\mu^{2m}(R^{\epsilon}_m < \Pi_H(2m)2^{\frac{-\epsilon m}{2}}$

4. fáza: - kombinácia predch. výsledkov

2.3 Zložitosť vzorky konzistentných algoritmov

Fakt: **Ak hypotézový priestor H má konečnú *VC* dimenziu, tak je potencionálne naučiteľný.**

Inak povedané:

K ľubovoľným parametrom δ - dôveryhodnosť, ϵ - presnosť ($0 < \delta, \epsilon < 1$), existuje vzorka dĺžky $m_0 = m_0(H, \delta, \epsilon)$ taká, že

$$m \geq m_0 \Longrightarrow \mu^m\{\mathbf{s} \in S(m,t); H[\mathbf{s}] \cap B_\epsilon = 0\} > 1 - \delta \tag{2.1}$$

pre ľubovoľné pravdepodobnostné rozdelenie μ na X ľubovoľný cieľový koncept $t \in H$.

Teda z toho vyplýva, že

- ľubovoľný konzistentný učiaci algoritmus L pre H je *pac* a ďalej
- ľubovoľné $\mu_0(H, \delta, \epsilon)$, pre ktoré platí vyššie uvedená podmienka, je horná hranica na zložitosť vzorky $m_L(H, \delta, \epsilon)$.

Ukážeme, že existuje presnejší výraz pre m_0, a teda pre hornú hranicu na zložitosť vzorky ľubovoľného konzistentného učiaceho algoritmu L pre H.

Pripomeňme, že bolo dokázané, že ak H je konečný hypotézový priestor a L je konzistentný učiaci algoritmus pre H, tak L je *pac* a platí

$$m_L(H, \delta, \epsilon) \leq \left\lceil \frac{1}{\epsilon} ln(\frac{|H|}{\delta}) \right\rceil \tag{2.2}$$

Horná hranica, ktorú odvodíme, závisí od VC dimenzie H a nie od mohutnosti H.

Veta: Predpokladajme, že H je hypotézový priestor s konečnou VC dimenziou $d \geq 1$ a platí $0 < \delta, \epsilon < 1$. Nech

$$m_0 = m_0(H, \delta, \epsilon) = \left\lceil \frac{4}{\epsilon}(d.lg(\frac{12}{\epsilon}) + lg(\frac{2}{\delta})) \right\rceil \tag{2.3}$$

Potom pre ľubovoľné $m \geq m_0$ platí

$$\mu^m\{\mathbf{s} \in S(m,t); H[\mathbf{s}] \cap B_\epsilon \neq \emptyset\} < \delta. \tag{2.4}$$

Dôsledok: Predpokladajme,že hypotézový priestor H má VC dimenziu $d \geq 1$. Potom ľubovoľný konzistentný učiaci algoritmus L pre H je *pac* so zložitosťou vzorky

$$m_L(H, \delta, \epsilon) \leq \left\lceil \frac{4}{\epsilon}(d.lg(\frac{12}{\epsilon}) + lg(\frac{2}{\delta})) \right\rceil \tag{2.5}$$

Toto už je sľúbený výsledok.

Príklad 1: Nech H je priestor lúčov. Jeho VC dimenzia je $VCdim(H) = 1$, teda ak L je ľubovoľný konzistentný učiaci algoritmus pre priestor lúčov, máme

$$m_L(H, \delta, \epsilon) \leq \left\lceil \frac{4}{\epsilon}(d.lg(\frac{12}{\epsilon}) + lg(\frac{2}{\delta})) \right\rceil \tag{2.6}$$

My sme priamo dokázali, že

$$m_0 = \left\lceil \frac{1}{\epsilon} ln(\frac{1}{\delta}) \right\rceil. \tag{2.7}$$

Vidíme, že hranica vypočítaná priamo je lepšia než vyplývajúca z VC dimenzie. Avšak priame argumenty sú mnohokrát ťažké a v tomto prípade je zrejmé, že ak δ a ϵ sú toho istého radu, tak tieto hranice sa líšia len o konštantu.

Príklad 2: Reálny perceptrón P_n má VC dimenziu $n + 1$. Predpokladajme, že pre ľubovoľnú tréningovú vzorku pre hypotézu v P_n máme nájsť stav ω perceptrónu taký, že h_ω je konzistentná so vzorkou. Potom keď použijeme tréningovú vzorku dĺžky

$$\left\lceil \frac{4}{\epsilon}((n+1).lg(\frac{12}{\epsilon}) + lg(\frac{2}{\delta})) \right\rceil \tag{2.8}$$

máme zaručenú približne správnu hypotézu bez ohľadu na cieľovú hypotézu a pravdepodobnostné rozdelenie príkladov.

2.4 Dolné hranice pre zložitosť vzorky

Pripomeňme, že sme dokázali, že ak priestor H má nekonečnú VC dimenziu, tak nie je potenciálne naučiteľný.

Veta: Predpokladajme, že hypotézový priestor H má VC dimenziu $d \geq 1$. Potom existuje konzistentný *pac* učiaci algoritmus L pre H taký, že pre ľubovoľné δ a ϵ zložitosť vzorky splňuje

$$m_L(H, \delta, \epsilon) \geq d(1 - \epsilon).$$

Uvedený výsledok je síce jednoduchý, ale je použiteľný len na konzistentné učiace algoritmy. Ďalej neposkytuje univerzálnu dolnú hranicu na zložitosť vzorky konzistentného učiaceho algoritmu; skôr sa zaoberá najhoršími možnými konzistentnými učiacimi algoritmami. Ukážeme silnejší výsledok Ehrenfeuchta et al. - 1989, ktorý poskytuje dolnú hranicu na zložitosť vzorky *ľubovoľného pac* učiaceho algoritmu pre hypotézový priestor konečnej VC dimenzie.

Veta: Pre ľubovoľný hypotézový priestor H s VC dimenziou $d \geq 1$ a pre ľubovoľný *pac* učiaci algoritmus L pre H platí

$$m_L(H, \delta, \epsilon) > \frac{d-1}{32\epsilon} \tag{2.9}$$

pre $\delta \leq \frac{1}{100}$ a $\epsilon \leq \frac{1}{8}$.

Ak by sme uvažovali, že zložitosť vzorky prekročí $(d_0 - 1)/32\epsilon$, kde d_0 je ľubovoľné kladné číslo spĺňajúce $VCdim(H) \geq d_0$, tak by sme mohli dokázať nasledujúce tvrdenie:

Tvrdenie: Ak hypotézový priestor H má konečnú VC dimenziu, tak neexistuje žiadny *pac* učiaci algoritmus pre H.

Tieto výsledky podporujú tvrdenie, že VC dimenzia je dobrá miera "expresívnej sily" hypotézového priestoru H: väčšia VC dimenzia H znamená väčšiu zložitosť vzorky pre *pac* učenie H. V skutočnosti výsledky môžu byť zovšeobecnené aj pre prípady, keď C je ľubovoľný konceptový priestor s VC dimenziou aspoň $d_0 \geq 1$ a H je ľubovoľný hypotézový priestor (nie nutne rovný C).

Ak L je učiaci algoritmus pre (C, H), vstupom do L musí byť tréningová vzorka dĺžky väčšej než $(d_0 - 1)/32\epsilon$, aby bola zaručená presnosť $\epsilon \leq \frac{1}{8}$ s pravdepodobnosťou $1 - \delta > 99/100$. Ak C má nekonečnú VC dimenziu, tak nemôže existovať žiadny učiaci algoritmus pre (C, H), ktorý je *pac* pre ľubovoľný hypotézový priestor H.

Príklad: Ak J je priestor všetkých zjednotení intervalov, tak pretože J má nekonečnú VC dimenziu, neexistuje žiadny *pac* učiaci algoritmus pre (J, H) pre ľubovoľný hypotézový priestor H. Vyššie uvedený výsledok je veľmi silný. Ukazuje nielen to, že neexistuje žiadny konzistentný alebo efektívny *pac* učiaci algoritmus, ale tiež, že pre dané neohraničené výpočtové zdroje žiadny algoritmus nemôže *pac* naučiť J nezávisle od toho ako by boli reprezentované výstupné hypotézy. Samozrejme, tieto závery platia pre ľubovoľný priestor nekonečnej VC dimenzie, takých ako priestor všetkých uzavretých množín alebo priestor charakteristických funkcií všetkých polygonálnych oblastí v $\mathbf{R}^2$.

Iný výsledok týkajúci sa dolných hraníc zaviedol Blumer et al., 1989. Tento výsledok obsahuje δ a ϵ, ale nezávisí od VC dimenzie hypotézového priestoru. Toto sa dá použiť na netriviálne hypotézové priestory, t. j. také priestory, ktoré obsahujú viac než dve hypotézy.

Veta: Predpokladajme, že L je ľubovoľný *pac* učiaci algoritmus pre netriviálny hypotézový priestor H. Potom

$$m_L(H, \delta, \epsilon) > \frac{(1-\epsilon)}{\epsilon} ln(\frac{1}{\delta}), \tag{2.10}$$

pre ľubovoľné $0 < \delta, \epsilon < 1$.

2.5 Porovnanie hraníc zložitosti vzoriek

Ako už bolo uvedené, mnohé predchádzajúce výsledky môžu byť zovšeobecnené pre prípad, keď konceptový a hypotézový priestor sú rôzne. Ukážeme hranice pre zložitosť vzorky v týchto všeobecnejších prípadoch.

Veta: Nech C je konceptový a H je hypotézový priestor a predpokladajme, že H má konečnú VC dimenziu aspoň 1. Ak L je ľubovoľný konzistentný učiaci algoritmus pre (C, H), potom L je *pac* a pre zložitosť vzorky platí

$$m_L \leq \left\lceil \frac{4}{\epsilon}(VCdim(H) lg(\frac{12}{\epsilon}) + lg(\frac{2}{\delta})) \right\rceil \tag{2.11}$$

pre ľubovoľné δ a ϵ.

Veta: Nech C je konceptový a H je hypotézový priestor taký, že C má VC dimenziu aspoň 1. Predpokladajme, že L je ľubovoľný *pac* učiaci algoritmus pre (C, H). Potom pre zložitosť vzorky pre L platí

$$m_L > max(\frac{VCdim(C) - 1}{32\epsilon}, \frac{1}{\epsilon} ln(\frac{1}{\delta})) \tag{2.12}$$

pre všetky $\delta \leq 1/100$ a $\epsilon \leq 1/8$.

Označenie:

Píšeme $f = O(g)$, ak existuje nejaká konštanta $k > 0$ taká, že pre všetky relevantné x platí $f(x) \leq k.g(x)$.

Píšeme $f = \Omega(g)$, ak existuje nejaká konštanta $k > 0$ taká, že pre všetky relevantné x platí $f(x) \geq k.g(x)$.

Použitím toho označenia sformulujeme vzťahy pre zložitosť vzorky.

- Ak L je *pac*, potom C musí mať konečnú VC dimenziu a platí

$$m_L(C, \delta, \epsilon) = \Omega(\frac{VCdim(C)}{\epsilon} + \frac{1}{\epsilon} ln(\frac{1}{\delta})) \tag{2.13}$$

- Ak H má konečnú VC dimenziu a L je konzistentný, potom L je *pac* a platí

$$m_L(C,\delta,\epsilon) = O\Big(\frac{VCdim(C)}{\epsilon}ln(\frac{1}{\epsilon}) + \frac{1}{\epsilon}ln(\frac{1}{\delta})\Big) \tag{2.14}$$

- Ak H je konečný a L je konzistentný, potom L je *pac* a platí

$$m_L(C,\delta,\epsilon) = O\Big(\frac{1}{\epsilon}ln|H| + \frac{1}{\epsilon}ln(\frac{1}{\delta})\Big) \tag{2.15}$$

V prípade, že $C = H$, VC dimenzia d je konečná a L je konzistentný, máme dolné a horné hranice

$$m_L(H,\delta,\epsilon) = \Omega\Big(\frac{d}{\epsilon} + \frac{1}{\epsilon}ln(\frac{1}{\delta})\Big);$$

$$m_L(H,\delta,\epsilon) = \emptyset\Big(\frac{d}{\epsilon}ln(\frac{1}{\epsilon}) + \frac{1}{\epsilon}ln(\frac{1}{\delta})\Big);$$

Faktor $ln(1/\epsilon)$, ktorý odlíšuje hornú hranicu od dolnej, je nevyhnutný. Výsledky Hausslera, Littlestonea a Warmutha, 1988, ukazujú, že pre každé $d \geq 1$ existuje hypotézový priestor H_d a konzistentný učiaci algoritmus L pre H_d so zložitosťou vzorky rovnou hornej hranici. Na druhej strane je otvorený problém rozhodnúť, či pre každé d a pre každý konceptový priestor C s VC dimenziou d, existuje *nejaký* hypotézový priestor H a *nejaký* (C,H) učiaci algoritmus L, pre ktorý zložitosť vzorky má dolnú hranicu.

2.6 Cvičenia:

1. Nech H je hypotézový priestor s vlastnosťou, že pre ľubovoľné $t \in H$ a ľubovoľné $0 < \delta, \epsilon < 1$, existuje $m_0(t,\delta,\epsilon)$ také, že

   $$m \geq m_0(t,\delta,\epsilon) \Longrightarrow \mu^m\{\mathbf{s} \in S(m,t); H[\mathbf{s}] \cap B_\epsilon\} > 1-\delta$$

   pre ľubovoľné rozdelenie pravdepodobnosti μ na vstupnom priestore. Teda m_0 môže závisieť len na cieľovom koncepte. Ukážte, že H musí mať konečnú VC dimenziu a je preto potencionálne naučiteľný.

2. Dokážte, že pre ľubovoľné $c > 0$ platí

   $$lnx \leq (ln(\frac{1}{c}) - 1) + cx,$$

   pre všetky $x > 0$.

3. Ukážte,že priestor charakteristických funkcií uzavretých a ohrani4ených polygónových oblastí v rovine $\mathbf{R}^2$ nie *pac* naučiteľný.

4. Prepokladajme, že H je ľubovoľný hypotézový priestor konečnej VC dimenzie $d \geq 1$ a že L je ľubovoľný konzistentný učiaci algoritmus pre H. Je dané nejaké rozdelenie pravdepodobnosti na príkladovom priestore. Akú veľkú tréningovú vzorku potrebujeme, aby s aspoň z 90% nou šancou sme získali hypotézu s chybou menšou než 5%?

5. Booleovská funkcia f sa nazýva *symetrická*, ak $f(x)$ závisí len od počtu vstupov x, ktoré sú rovné 1. Napríklad, pre ľubovoľné n koncept parity definovaný na $\{0,1\}^n$ je symetrick'a. Nech n je kladné celé číslo a nech S_n označuje množinu všetkých symetrických funkcií na $\{0,1\}^n$. Aká je VC dimenzia S_n? Uveďte dolnú a hornú hranicu zložitosti vzorky pre ľubovoľný konzistentný *pac* učiaci algoritmus pre S_n. Poznamenajme, že ľubovoľná hypotéza h v S_n môže byť reprezentovaná vektorom $(h_0, h_1, \ldots, h_n) \in \{0,1\}^{n+1}$, kde h_i je hodnota h na príkladoch majúcich presne i jedničiek. Navrhnite konzistentný učiaci algoritmus pre S_n, ktorý reprezentuje priestor týmto spôsobom.

6. Nech H, G sú hypotézové priestory definované na tom istom príkladovom priestore X. Pre hypotézy $h \in H, g \in G$, definujme $h \vee g$ takto

   $$h \vee g = 1, \text{ ak } h(x) = 1 \text{ alebo } g(x) = 1; 0 \text{ inak }.$$

Nech $H \vee G = \{h \vee g; h \in H, g \in G\}$.

Dokážte, že

$$\Pi_{H \vee G} \leq \Pi_H(m)\Pi_G(m)$$

pre všetky m. Definujte $H \wedge G$ obvyklým spôsobom, dokážte analogický výsledok pre tento priestor. Ak H a G sú potenciálne naučiteľné, čo je možné povedať o $H \vee G$ a $H \wedge G$?

7. Nech H je hypotézový priestor konečnej VC dimenzie $d \geq 1$ a pre $s \geq 1$ definujme $H(s)$ induktívne takto: $H(1) = H$ a $H(k) = H \vee H(k-1), k \geq 2$. Dokážte, že pre $m > d$ platí

$$\Pi_{H(s)} \leq (\frac{em}{d})^{sd}.$$

Potom ukážte, že VC dimenzia $H(s)$ je najviac $2sdlg(3s)$.

2.7 Zložitosť vzorky konzistentných algoritmov

Fakt: **Ak hypotézový priestor H má konečnú VC dimenziu, tak je potencionálne naučiteľný.**

Inak povedané:

K ľubovoľným parametrom δ - dôveryhodnosť, ϵ - presnosť ($0 < \delta, \epsilon < 1$), existuje vzorka dĺžky $m_0 = m_0(H, \delta, \epsilon)$ taká, že

$$m \geq m_0 \Longrightarrow \mu^m\{\mathbf{s} \in S(m,t); H[\mathbf{s}] \cap B_\epsilon = 0\} > 1 - \delta \tag{2.16}$$

pre ľubovoľné pravdepodobnostné rozdelenie μ na X ľubovoľný cieľový koncept $t \in H$.

Teda z toho vyplýva, že

- ľubovoľný konzistentný učiaci algoritmus L pre H je *pac* a ďalej
- ľubovoľné $\mu_0(H, \delta, \epsilon)$, pre ktoré platí vyššie uvedená podmienka, je horná hranica na zložitosť vzorky $m_L(H, \delta, \epsilon)$.

Ukážeme, že existuje presnejší výraz pre m_0, a teda pre hornú hranicu na zložitosť vzorky ľubovoľného konzistentného učiaceho algoritmu L pre H.

Pripomeňme, že bolo dokázané, že ak H je konečný hypotézový priestor a L je konzistentný učiaci algoritmus pre H, tak L je *pac* a platí

$$m_L(H, \delta, \epsilon) \leq \left\lceil \frac{1}{\epsilon} ln(\frac{|H|}{\delta}) \right\rceil \tag{2.17}$$

Horná hranica, ktorú odvodíme, závisí od VC dimenzie H a nie od mohutnosti H.

Veta: Predpokladajme, že H je hypotézový priestor s konečnou VC dimenziou $d \geq 1$ a platí $0 < \delta, \epsilon < 1$. Nech

$$m_0 = m_0(H, \delta, \epsilon) = \left\lceil \frac{4}{\epsilon}(d.lg(\frac{12}{\epsilon}) + lg(\frac{2}{\delta})) \right\rceil \tag{2.18}$$

Potom pre ľubovoľné $m \geq m_0$ platí

$$\mu^m\{\mathbf{s} \in S(m,t); H[\mathbf{s}] \cap B_\epsilon \neq \emptyset\} < \delta. \tag{2.19}$$

Dôsledok: Predpokladajme,že hypotézový priestor H má VC dimenziu $d \geq 1$. Potom ľubovoľný konzistentný učiaci algoritmus L pre H je *pac* so zložitosťou vzorky

$$m_L(H, \delta, \epsilon) \leq \left\lceil \frac{4}{\epsilon}(d.lg(\frac{12}{\epsilon}) + lg(\frac{2}{\delta})) \right\rceil \tag{2.20}$$

Toto už je sľúbený výsledok.

Príklad 1: Nech H je priestor lúčov. Jeho VC dimenzia je $VCdim(H) = 1$, teda ak L je ľubovoľný konzistentný učiaci algoritmus pre priestor lúčov, máme

$$m_L(H, \delta, \epsilon) \leq \left\lceil \frac{4}{\epsilon}(d.lg(\frac{12}{\epsilon}) + lg(\frac{2}{\delta})) \right\rceil \tag{2.21}$$

My sme priamo dokázali, že

$$m_0 = \left\lceil \frac{1}{\epsilon} ln(\frac{1}{\delta}) \right\rceil. \tag{2.22}$$

Vidíme, že hranica vypočítaná priamo je lepšia než vyplývajúca z VC dimenzie. Avšak priame argumenty sú mnohokrát ťažké a v tomto prípade je zrejmé, že ak δ a ϵ sú toho istého radu, tak tieto hranice sa líšia len o konštantu.

Príklad 2: Reálny perceptrón P_n má VC dimenziu $n+1$. Predpokladajme, že pre ľubovoľnú tréningovú vzorku pre hypotézu v P_n máme nájsť stav ω perceptrónu taký, že h_ω je konzistentná so vzorkou. Potom keď použijeme tréningovú vzorku dĺžky

$$\left\lceil \frac{4}{\epsilon}((n+1).lg(\frac{12}{\epsilon}) + lg(\frac{2}{\delta}))\right\rceil \tag{2.23}$$

máme zaručenú približne správnu hypotézu bez ohľadu na cieľovú hypotézu a pravdepodobnostné rozdelenie príkladov.

2.8 Dolné hranice pre zložitosť vzorky

Pripomeňme, že sme dokázali, že ak priestor H má nekonečnú VC dimenziu, tak nie je potenciálne naučiteľný.

Veta: Predpokladajme, že hypotézový priestor H má VC dimenziu $d \geq 1$. Potom existuje konzistentný *pac* učiaci algoritmus L pre H taký, že pre ľubovoľné δ a ϵ zložitosť vzorky splňuje

$$m_L(H,\delta,\epsilon) \geq d(1-\epsilon).$$

Uvedený výsledok je síce jednoduchý, ale je použiteľný len na konzistentné učiace algoritmy. Ďalej neposkytuje univerzálnu dolnú hranicu na zložitosť vzorky konzistentného učiaceho algoritmu; skôr sa zaoberá najhoršími možnými konzistentnými učiacimi algoritmami. Ukážeme silnejší výsledok Ehrenfeuchta et al. - 1989, ktorý poskytuje dolnú hranicu na zložitosť vzorky *ľubovoľného pac* učiaceho algoritmu pre hypotézový priestor konečnej VC dimenzie.

Veta: Pre ľubovoľný hypotézový priestor H s VC dimenziou $d \geq 1$ a pre ľubovoľný *pac* učiaci algoritmus L pre H platí

$$m_L(H,\delta,\epsilon) > \frac{d-1}{32\epsilon} \tag{2.24}$$

pre $\delta \leq \frac{1}{100}$ a $\epsilon \leq \frac{1}{8}$.

Ak by sme uvažovali, že zložitosť vzorky prekročí $(d_0-1)/32\epsilon$, kde d_0 je ľubovoľné kladné číslo spĺňajúce $VCdim(H) \geq d_0$, tak by sme mohli dokázať nasledujúce tvrdenie:

Tvrdenie: Ak hypotézový priestor H má konečnú VC dimenziu, tak neexistuje žiadny *pac* učiaci algoritmus pre H.

Tieto výsledky podporujú tvrdenie, že VC dimenzia je dobrá miera "expresívnej sily" hypotézového priestoru H: väčšia VC dimenzia H znamená väčšiu zložitosť vzorky pre *pac* učenie H. V skutočnosti výsledky môžu byť zovšeobecnené aj pre prípady, keď C je ľubovoľný konceptový priestor s VC dimenziou aspoň $d_0 \geq 1$ a H je ľubovoľný hypotézový priestor (nie nutne rovný C).

Ak L je učiaci algoritmus pre (C,H), vstupom do L musí byť tréningová vzorka dĺžky väčšej než $(d_0-1)/32\epsilon$, aby bola zaručená presnosť $\epsilon \leq \frac{1}{8}$ s pravdepodobnosťou $1-\delta > 99/100$. Ak C má nekonečnú VC dimenziu, tak nemôže existovať žiadny učiaci algoritmus pre (C,H), ktorý je *pac* pre ľubovoľný hypotézový priestor H.

Príklad: Ak J je priestor všetkých zjednotení intervalov, tak pretože J má nekonečnú VC dimenziu, neexistuje žiadny *pac* učiaci algoritmus pre (J,H) pre ľubovoľný hypotézový priestor H. Vyššie uvedený výsledok je veľmi silný. Ukazuje nielen to, že neexistuje žiadny konzistentný alebo efektívny *pac* učiaci algoritmus, ale tiež, že pre dané neohraničené výpočtové zdroje žiadny algoritmus nemôže *pac* naučiť J nezávisle od toho ako by boli reprezentované výstupné hypotézy. Samozrejme, tieto závery platia pre ľubovoľný priestor nekonečnej VC dimenzie, takých ako priestor všetkých uzavretých množín alebo priestor charakteristických funkcií všetkých polygonálnych oblastí v $\mathbf{R}^2$.

Iný výsledok týkajúci sa dolných hraníc zaviedol Blumer et al., 1989. Tento výsledok obsahuje δ a ϵ, ale nezávisí od VC dimenzie hypotézového priestoru. Toto sa dá použiť na netriviálne hypotézové priestory, t. j. také priestory, ktoré obsahujú viac než dve hypotézy.

Veta: Predpokladajme, že L je ľubovoľný *pac* učiaci algoritmus pre netriviálny hypotézový priestor H. Potom

$$m_L(H, \delta, \epsilon) > \frac{(1-\epsilon)}{\epsilon} ln(\frac{1}{\delta}), \tag{2.25}$$

pre ľubovoľné $0 < \delta, \epsilon < 1$.

2.9 Porovnanie hraníc zložitosti vzoriek

Ako už bolo uvedené, mnohé predchádzajúce výsledky môžu byť zovšeobecnené pre prípad, keď konceptový a hypotézový priestor sú rôzne. Ukážeme hranice pre zložitosť vzorky v týchto všeobecnejších prípadoch.

Veta: Nech C je konceptový a H je hypotézový priestor a predpokladajme, že H má konečnú VC dimenziu aspoň 1. Ak L je ľubovoľný konzistentný učiaci algoritmus pre (C, H), potom L je *pac* a pre zložitosť vzorky platí

$$m_L \leq \left\lceil \frac{4}{\epsilon}(VCdim(H) lg(\frac{12}{\epsilon}) + lg(\frac{2}{\delta})) \right\rceil \tag{2.26}$$

pre ľubovoľné δ a ϵ.

Veta: Nech C je konceptový a H je hypotézový priestor taký, že C má VC dimenziu aspoň 1. Predpokladajme, že L je ľubovoľný *pac* učiaci algoritmus pre (C, H). Potom pre zložitosť vzorky pre L platí

$$m_L > max(\frac{VCdim(C) - 1}{32\epsilon}, \frac{1}{\epsilon} ln(\frac{1}{\delta})) \tag{2.27}$$

pre všetky $\delta \leq 1/100$ a $\epsilon \leq 1/8$.

Označenie:
Píšeme $f = O(g)$, ak existuje nejaká konštanta $k > 0$ taká, že pre všetky relevantné x platí $f(x) \leq k.g(x)$.
Píšeme $f = \Omega(g)$, ak existuje nejaká konštanta $k > 0$ taká, že pre všetky relevantné x platí $f(x) \geq k.g(x)$.

Použitím toho označenia sformulujeme vzťahy pre zložitosť vzorky.

- Ak L je *pac*, potom C musí mať konečnú VC dimenziu a platí

$$m_L(C, \delta, \epsilon) = \Omega(\frac{VCdim(C)}{\epsilon} + \frac{1}{\epsilon} ln(\frac{1}{\delta})) \tag{2.28}$$

- Ak H má konečnú VC dimenziu a L je konzistentný, potom L je *pac* a platí

$$m_L(C, \delta, \epsilon) = O\Big(\frac{VCdim(C)}{\epsilon} ln(\frac{1}{\epsilon}) + \frac{1}{\epsilon} ln(\frac{1}{\delta})\Big) \tag{2.29}$$

- Ak H je konečný a L je konzistentný, potom L je *pac* a platí

$$m_L(C, \delta, \epsilon) = O\Big(\frac{1}{\epsilon} ln|H| + \frac{1}{\epsilon} ln(\frac{1}{\delta})\Big) \tag{2.30}$$

V prípade, že $C = H$, VC dimenzia d je konečná a L je konzistentný, máme dolné a horné hranice

$$m_L(H, \delta, \epsilon) = \Omega\Big(\frac{d}{\epsilon} + \frac{1}{\epsilon} ln(\frac{1}{\delta})\Big);$$

$$m_L(H,\delta,\epsilon) = \emptyset\Big(\frac{d}{\epsilon}ln(\frac{1}{\epsilon}) + \frac{1}{\epsilon}ln(\frac{1}{\delta})\Big);$$

Faktor $ln(1/\epsilon)$, ktorý odlíšuje hornú hranicu od dolnej, je nevyhnutný. Výsledky Hausslera, Littlestonea a Warmutha, 1988, ukazujú, že pre každé $d \geq 1$ existuje hypotézový priestor H_d a konzistentný učiaci algoritmus L pre H_d so zložitosťou vzorky rovnou hornej hranici. Na druhej strane je otvorený problém rozhodnúť, či pre každé d a pre každý konceptový priestor C s VC dimenziou d, existuje *nejaký* hypotézový priestor H a *nejaký* (C,H) učiaci algoritmus L, pre ktorý zložitosť vzorky má dolnú hranicu.

2.10 Cvičenia:

1. Nech H je hypotézový priestor s vlastnosťou, že pre ľubovoľné $t \in H$ a ľubovoľné $0 < \delta, \epsilon < 1$, existuje $m_0(t,\delta,\epsilon)$ také, že

$$m \geq m_0(t,\delta,\epsilon) \Longrightarrow \mu^m\{\mathbf{s} \in S(m,t); H[\mathbf{s}] \cap B_\epsilon\} > 1-\delta$$

pre ľubovoľné rozdelenie pravdepodobnosti μ na vstupnom priestore. Teda m_0 môže závisieť len na cieľovom koncepte. Ukážte, že H musí mať konečnú VC dimenziu a je preto potencionálne naučiteľný.

2. Dokážte, že pre ľubovoľné $c > 0$ platí

$$lnx \leq (ln(\frac{1}{c}) - 1) + cx,$$

pre všetky $x > 0$.

3. Ukážte,že priestor charakteristických funkcií uzavretých a ohrani4ených polygónových oblastí v rovine $\mathbf{R}^2$ nie *pac* naučiteľný.

4. Prepokladajme, že H je ľubovoľný hypotézový priestor konečnej VC dimenzie $d \geq 1$ a že L je ľubovoľný konzistentný učiaci algoritmus pre H. Je dané nejaké rozdelenie pravdepodobnosti na príkladovom priestore. Akú veľkú tréningovú vzorku potrebujeme, aby s aspoň z 90% nou šancou sme získali hypotézu s chybou menšou než 5%?

5. Booleovská funkcia f sa nazýva *symetrická*, ak $f(x)$ závisí len od počtu vstupov x, ktoré sú rovné 1. Napríklad, pre ľubovoľné n koncept parity definovaný na $\{0,1\}^n$ je symetrick'a. Nech n je kladné celé číslo a nech S_n označuje množinu všetkých symetrických funkcií na $\{0,1\}^n$. Aká je VC dimenzia S_n? Uveďte dolnú a hornú hranicu zložitosti vzorky pre ľubovoľný konzistentný *pac* učiaci algoritmus pre S_n. Poznamenajme, že ľubovoľná hypotéza h v S_n môže byť reprezentovaná vektorom $(h_0, h_1, \ldots, h_n) \in \{0,1\}^{n+1}$, kde h_i je hodnota h na príkladoch majúcich presne i jedničiek. Navrhnite konzistentný učiaci algoritmus pre S_n, ktorý reprezentuje priestor týmto spôsobom.

6. Nech H, G sú hypotézové priestory definované na tom istom príkladovom priestore X. Pre hypotézy $h \in H, g \in G$, definujme $h \vee g$ takto

$$h \vee g = 1, \text{ ak } h(x) = 1 \text{ alebo } g(x) = 1; 0 \text{ inak }.$$

Nech $H \vee G = \{h \vee g; h \in H, g \in G\}$.

Dokážte, že

$$\Pi_{H\vee G} \leq \Pi_H(m)\Pi_G(m)$$

pre všetky m. Definujte $H \wedge G$ obvyklým spôsobom, dokážte analogický výsledok pre tento priestor. Ak H a G sú potenciálne naučiteľné, čo je možné povedať o $H \vee G$ a $H \wedge G$?

7. Nech H je hypotézový priestor konečnej VC dimenzie $d \geq 1$ a pre $s \geq 1$ definujme $H(s)$ induktívne takto: $H(1) = H$ a $H(k) = H \vee H(k-1), k \geq 2$. Dokážte, že pre $m > d$ platí

$$\Pi_{H(s)} \leq (\frac{em}{d})^{sd}.$$

Potom ukážte, že VC dimenzia $H(s)$ je najviac $2sdlg(3s)$.