Univerzita Pavla Jozefa Šafárika v Košiciach,
Prírodovedecká fakulta

VC dimenzia hypotézového priestoru H

Gabriela Andrejková, Ľubomír Antoni

EURÓPSKA ÚNIA
Európsky sociálny fond
Európsky fond regionálneho rozvoja
OPERAČNÝ PROGRAM
ĽUDSKÉ ZDROJE
MINISTERSTVO ŠKOLSTVA, VEDY, VÝSKUMU A ŠPORTU SLOVENSKEJ REPUBLIKY
UMB UNIVERZITA MATEJA BELA

Tento projekt sa realizuje vďaka podpore z Európskeho sociálneho fondu a Európskeho fondu regionálneho rozvoja v rámci Operačného programu Ľudské zdroje

www.minedu.sk www.employment.gov.sk/sk/esf/ www.itakademia.sk

- Konečné hypotézové priestory sú potenciálne naučiteľné. Vieme.
- Ako to je s nekonečnými hypotézovými priestormi?
- U nekonečných priestorov nás viac zaujíma, akú výpovednú silu majú jeho hypotézy.
- Kedy má hypotézový priestor H veľkú "expresívnu silu"?

Ak môže dosiahnuť všetky možné klasifikácie (pomocou svojich hypotéz) veľkej množiny príkladov (vždy konečnej). Uvažujeme pozitívne a negatívne príklady.

EURÓPSKA ÚNIA
Európsky sociálny fond
Európsky fond regionálneho rozvoja
OPERAČNÝ PROGRAM ĽUDSKÉ ZDROJE
MINISTERSTVO ŠKOLSTVA, VEDY, VÝSKUMU A ŠPORTU SLOVENSKEJ REPUBLIKY
UNIVERZITA MATEJA BELA
Tento projekt sa realizuje vďaka podpore z Európskeho sociálneho fondu a Európskeho fondu regionálneho rozvoja v rámci Operačného programu Ľudské zdroje
www.minedu.sk www.employment.gov.sk/sk/esf/ www.itakademia.sk

Úvod

Zaujíma nás:
Veľkosť najväčšej množiny príkladov $\{x_1, x_2, \ldots, x_m\}$, pre ktorú je možné nájsť v H konzistentnú hypotézu pre každé možné ohodnotenie týchto príkladov $\in \{+1, -1\}^m$.

Číslo označíme

$$VC(H)$$

EURÓPSKA ÚNIA
Európsky sociálny fond
Európsky fond regionálneho rozvoja
OPERAČNÝ PROGRAM ĽUDSKÉ ZDROJE
MINISTERSTVO ŠKOLSTVA, VEDY, VÝSKUMU A ŠPORTU SLOVENSKEJ REPUBLIKY
UMB UNIVERZITA MATEJA BELA
Tento projekt sa realizuje vďaka podpore z Európskeho sociálneho fondu a Európskeho fondu regionálneho rozvoja v rámci Operačného programu Ľudské zdroje
www.minedu.sk www.employment.gov.sk/sk/esf/ www.itakademia.sk

Úvod

Príklad 0.

$H = M_2$ – monočleny 2 literálov. Koľko máme monočlenov v M_2?
$3^2 = 9$

Príkladový priestor (2 premenné): 4 príklady, $\{00, 01, 10, 11\}$.
Počet možných ohodnotení je 16, lebo každý príklad môže byť kladný alebo záporný.

Vieme nájsť v H konzistentnú hypotézu pre každé ohodnotenie?
Napríklad pre ohodnotenie $\{0, 0, 1, 1\}$ máme monočlen $u_1 \in H$, pre $\{1, 1, 0, 0\}$ monočlen $\overline{u_1}$.

Aké sú monočleny pre ohodnotenia
$\{1, 0, 0, 0\}, \{0, 1, 0, 0\}, \{0, 0, 1, 0\}$, pre $\{0, 0, 0, 1\}$?

EURÓPSKA ÚNIA
Európsky sociálny fond
Európsky fond regionálneho rozvoja
OPERAČNÝ PROGRAM ĽUDSKÉ ZDROJE
MINISTERSTVO ŠKOLSTVA, VEDY, VÝSKUMU A ŠPORTU SLOVENSKEJ REPUBLIKY
UMB UNIVERZITA MATEJA BELA

Tento projekt sa realizuje vďaka podpore z Európskeho sociálneho fondu a Európskeho fondu regionálneho rozvoja v rámci Operačného programu Ľudské zdroje

Úvod

Možné ohodnotenia:

Pr	f_1	f_2	f_3	f_4	f_5	f_6	f_7	f_18	f_9	f_{10}	f_{11}	f_{12}	f_{13}	f_{14}	f_{15}	f_{16}
00	0	1	0	0	0	1	1	1	0	0	0	1	1	1	0	1
01	0	0	1	0	0	1	0	0	1	1	0	1	1	0	1	1
10	0	0	0	1	0	0	1	0	1	0	1	1	0	1	1	1
11	0	0	0	0	1	0	0	1	0	1	1	0	1	1	1	1

Pre každý príklad vieme nájsť hypotézy v H na jeho kladné aj záporné ohodnotenie. Teda $VC(H) \geq 1$.

Ak uvažujeme množiny po 2 príkladoch, vieme to tiež?
Máme 16 dvojíc príkladov a pre každú dvojicu treba overiť 4 možné ohodnotenia. Nájdeme monočleny pre všetky ohodnotenia dvojice príkladov $00, 11$?

EURÓPSKA ÚNIA
Európsky sociálny fond
Európsky fond regionálneho rozvoja
OPERAČNÝ PROGRAM ĽUDSKÉ ZDROJE
MINISTERSTVO ŠKOLSTVA, VEDY, VÝSKUMU A ŠPORTU SLOVENSKEJ REPUBLIKY
Tento projekt sa realizuje vďaka podpore z Európskeho sociálneho fondu a Európskeho fondu regionálneho rozvoja v rámci Operačného programu Ľudské zdroje
www.minedu.sk www.employment.gov.sk/sk/esf/ www.itakademia.sk

Úvod

Príklad 1: Vezmime nekonečnú množinu hypotéz:
$H = \{\langle a, b\rangle; a, b \in \mathbb{R}\}$.
Jednotlivé hypotézy sú uzavreté intervaly na reálnych číslach.
Príklady patriace do intervalu sú hypotézou ohodnotené ako $+1$, ostatné -1.

Analýza pre 1 príklad: Dosiahneme všetky možné klasifikácie?
Áno.

Obr.: $VC(H) \geq 1$

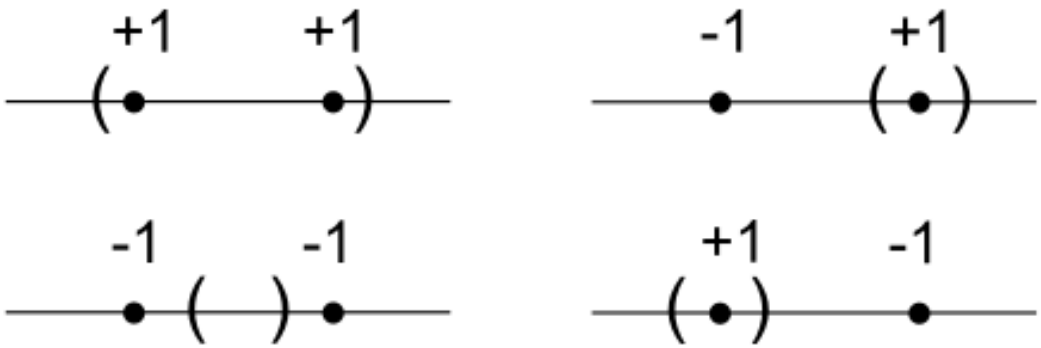
EURÓPSKA ÚNIA
Európsky sociálny fond
Európsky fond regionálneho rozvoja
OPERAČNÝ PROGRAM ĽUDSKÉ ZDROJE
MINISTERSTVO ŠKOLSTVA, VEDY, VÝSKUMU A ŠPORTU SLOVENSKEJ REPUBLIKY
UMB UNIVERZITA MATEJA BELA

Tento projekt sa realizuje vďaka podpore z Európskeho sociálneho fondu a Európskeho fondu regionálneho rozvoja v rámci Operačného programu Ľudské zdroje

www.minedu.sk www.employment.gov.sk/sk/esf/ www.itakademia.sk

Úvod

Analýza pre 2 príklady: Dosiahneme všetky možné klasifikácie?
Áno.
Pre každú dvojicu príkladov vieme nájsť vhodný interval.

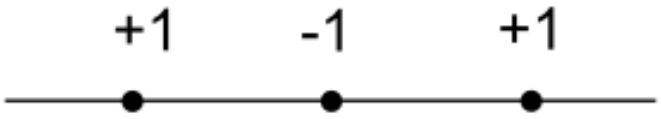


Obr.: $VC(H) \geq 2$

EURÓPSKA ÚNIA
Európsky sociálny fond
Európsky fond regionálneho rozvoja
OPERAČNÝ PROGRAM ĽUDSKÉ ZDROJE
MINISTERSTVO ŠKOLSTVA, VEDY, VÝSKUMU A ŠPORTU SLOVENSKEJ REPUBLIKY
Tento projekt sa realizuje vďaka podpore z Európskeho sociálneho fondu a Európskeho fondu regionálneho rozvoja v rámci Operačného programu Ľudské zdroje

Úvod

Analýza pre 3 príklady: Dosiahneme všetky možné klasifikácie?
Nie. Pre klasifikáciu



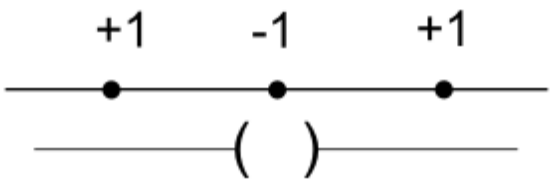
Obr.: $VC(H) < 3$ a zároveň $VC(H) \geq 2$

nemáme hypotézu.
Záver: Pre $H = \{\langle a, b\rangle; a, b \in \mathbb{R}\}$ platí $VC(H) = 2$

EURÓPSKA ÚNIA
Európsky sociálny fond
Európsky fond regionálneho rozvoja
OPERAČNÝ PROGRAM ĽUDSKÉ ZDROJE
MINISTERSTVO ŠKOLSTVA, VEDY, VÝSKUMU A ŠPORTU SLOVENSKEJ REPUBLIKY
UMB UNIVERZITA MATEJA BELA
Tento projekt sa realizuje vďaka podpore z Európskeho sociálneho fondu a Európskeho fondu regionálneho rozvoja v rámci Operačného programu Ľudské zdroje
www.minedu.sk www.employment.gov.sk/sk/esf/ www.itakademia.sk

Iný hypotézový priestor

Príklad 2: Vezmime nekonečnú množinu hypotéz, konečné intervaly a ich doplnky:
$H = \{\langle a, b\rangle; a, b \in \mathbb{R}\} \cup \{\mathbb{R} - \langle a, b\rangle; a, b \in \mathbb{R}\}$.
Pre každé ohodnotenie 3 príkladov už vieme nájsť konzistentnú hypotézu:



Obr.: $VC(H) \geq 3$

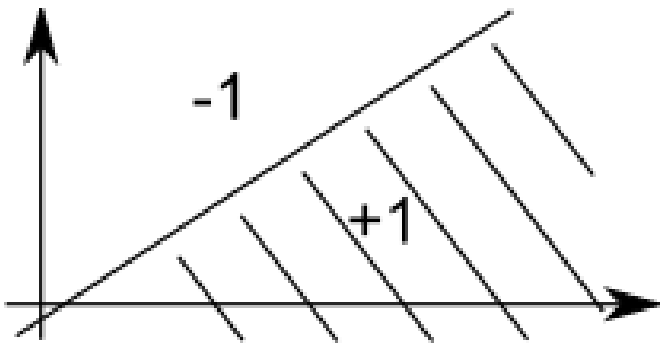
EURÓPSKA ÚNIA
Európsky sociálny fond
Európsky fond regionálneho rozvoja
OPERAČNÝ PROGRAM ĽUDSKÉ ZDROJE
MINISTERSTVO ŠKOLSTVA, VEDY, VÝSKUMU A ŠPORTU SLOVENSKEJ REPUBLIKY
Tento projekt sa realizuje vďaka podpore z Európskeho sociálneho fondu a Európskeho fondu regionálneho rozvoja v rámci Operačného programu Ľudské zdroje
www.minedu.sk www.employment.gov.sk/sk/esf/ www.itakademia.sk

Pre nasledujúce ohodnotenie 4 príkladov nevieme nájsť konzistentú hypotézu.

Obr.: $VC(H) < 4$, teda VC(H)=3

EURÓPSKA ÚNIA
Európsky sociálny fond
Európsky fond regionálneho rozvoja
OPERAČNÝ PROGRAM ĽUDSKÉ ZDROJE
MINISTERSTVO ŠKOLSTVA, VEDY, VÝSKUMU A ŠPORTU SLOVENSKEJ REPUBLIKY
Tento projekt sa realizuje vďaka podpore z Európskeho sociálneho fondu a Európskeho fondu regionálneho rozvoja v rámci Operačného programu Ľudské zdroje
www.minedu.sk www.employment.gov.sk/sk/esf/ www.itakademia.sk

Príklad 3: Nadroviny v $\mathbb{R}^2$. Nadrovina rozdelí priestor na 2 polpriestory. Hypotézy sú nadroviny.
Všetky 4 klasifikácie sú dosiahnuteľné pre 2 príklady.
Analýza pre 3 príklady:



Obr.: $VC(H) \geq 3$

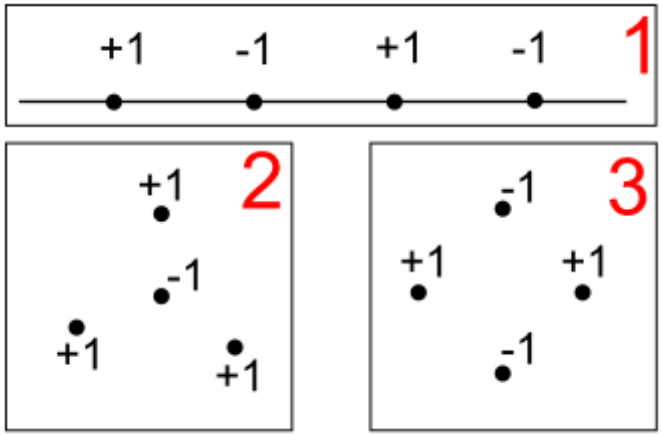
EURÓPSKA ÚNIA
Európsky sociálny fond
Európsky fond regionálneho rozvoja
OPERAČNÝ PROGRAM ĽUDSKÉ ZDROJE
MINISTERSTVO ŠKOLSTVA, VEDY, VÝSKUMU A ŠPORTU SLOVENSKEJ REPUBLIKY
Tento projekt sa realizuje vďaka podpore z Európskeho sociálneho fondu a Európskeho fondu regionálneho rozvoja v rámci Operačného programu Ľudské zdroje

Analýza pre 4 príklady:

Nevieme nájsť konzistentnú hypotézu pre štyri príklady. 3 prípady:

1. Príklady ležia na priamke
2. Jeden z príkladov leží v konvexnom obale ostatných
3. Žiaden z príkladov neleží v konvexnom obale ostatných

Pre nasledujúce rozloženia príkladov a ich ohodnotenia nevieme nájsť konzistentnú hypotézu v H (nadroviny):



Obr.: $VC(H) < 4$

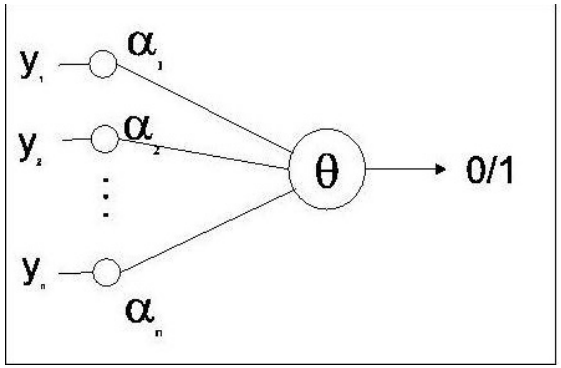
EURÓPSKA ÚNIA
Európsky sociálny fond
Európsky fond regionálneho rozvoja
Tento projekt sa realizuje vďaka podpore z Európskeho sociálneho fondu a Európskeho fondu regionálneho rozvoja v rámci Operačného programu Ľudské zdroje
www.minedu.sk www.employment.gov.sk/sk/esf/ www.itakademia.sk

Rastová funkcia

Vapnik, Chervonenkis - 1971

VC - dimenzia – najvýznamnejší pojem pre teóriu výp. učenia.

Pripomeňme si štruktúru perceptrónu:



EURÓPSKA ÚNIA
Európsky sociálny fond
Európsky fond regionálneho rozvoja
OPERAČNÝ PROGRAM ĽUDSKÉ ZDROJE
MINISTERSTVO ŠKOLSTVA, VEDY, VÝSKUMU A ŠPORTU SLOVENSKEJ REPUBLIKY
UMB UNIVERZITA MATEJA BELA

Tento projekt sa realizuje vďaka podpore z Európskeho sociálneho fondu a Európskeho fondu regionálneho rozvoja v rámci Operačného programu Ľudské zdroje

www.minedu.sk www.employment.gov.sk/sk/esf/ www.itakademia.sk

Perceptrón, hypotézový priestor perceptrónov H

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

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

Zobrazenie $\omega \vdash h_\omega$ nie je injekcia;
pre ľubovoľné $\lambda \in \mathbb{R}, \lambda \neq 0$, stav $\lambda * \omega$ definuje tú istú funkciu.

EURÓPSKA ÚNIA
Európsky sociálny fond
Európsky fond regionálneho rozvoja
OPERAČNÝ PROGRAM ĽUDSKÉ ZDROJE
MINISTERSTVO ŠKOLSTVA, VEDY, VÝSKUMU A ŠPORTU SLOVENSKEJ REPUBLIKY
UMB UNIVERZITA MATEJA BELA
Tento projekt sa realizuje vďaka podpore z Európskeho sociálneho fondu a Európskeho fondu regionálneho rozvoja v rámci Operačného programu Ľudské zdroje
www.minedu.sk www.employment.gov.sk/sk/esf/ www.itakademia.sk

$\Pi_H(\overline{\mathbf{x}})$ =**počet klasifikácií pomocou hypotéz z H, t.j. počet rôznych vektorov tvaru**

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

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

H môže byť nekonečný ... redukcia H na $H|E_{\overline{x}}$, $\quad E_x = \{\overline{x_1} \ldots \overline{x_m}\}$ je konečný a označme jeho kardinalitu $\pi_H(\overline{\mathbf{x}})$. Pre ľubovoľnú tréningovú vzorku dĺžky m platí $\pi_H(\overline{\mathbf{x}}) \leq 2^m$.
$\pi_H(\overline{\mathbf{x}})$ závisí od tréningovej vzorky.

EURÓPSKA ÚNIA
Európsky sociálny fond
Európsky fond regionálneho rozvoja
OPERAČNÝ PROGRAM ĽUDSKÉ ZDROJE
MINISTERSTVO ŠKOLSTVA, VEDY, VÝSKUMU A ŠPORTU SLOVENSKEJ REPUBLIKY
UMB UNIVERZITA MATEJA BELA
Tento projekt sa realizuje vďaka podpore z Európskeho sociálneho fondu a Európskeho fondu regionálneho rozvoja v rámci Operačného programu Ľudské zdroje
www.minedu.sk www.employment.gov.sk/sk/esf/ www.itakademia.sk

Definícia:

Rastová funkcia *je*

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

Definícia:

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 pre túto vzorku.*

EURÓPSKA ÚNIA
Európsky sociálny fond
Európsky fond regionálneho rozvoja
OPERAČNÝ PROGRAM ĽUDSKÉ ZDROJE
MINISTERSTVO ŠKOLSTVA, VEDY, VÝSKUMU A ŠPORTU SLOVENSKEJ REPUBLIKY
UMB UNIVERZITA MATEJA BELA

Tento projekt sa realizuje vďaka podpore z Európskeho sociálneho fondu a Európskeho fondu regionálneho rozvoja v rámci Operačného programu Ľudské zdroje

www.minedu.sk www.employment.gov.sk/sk/esf/ www.itakademia.sk

- Ak príklady v $\overline{\mathbf{x}}$ nie sú rôzne, tak $\overline{\mathbf{x}}$ nemôže byť rozbitá.
- Ak príklady v $\overline{\mathbf{x}}$ sú rôzne, $\overline{\mathbf{x}}$ 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

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

S je potom podmnožina sústreďujúca kladné príklady.

EURÓPSKA ÚNIA
Európsky sociálny fond
Európsky fond regionálneho rozvoja
OPERAČNÝ PROGRAM ĽUDSKÉ ZDROJE
MINISTERSTVO ŠKOLSTVA, VEDY, VÝSKUMU A ŠPORTU SLOVENSKEJ REPUBLIKY
UNIVERZITA MATEJA BELA
Tento projekt sa realizuje vďaka podpore z Európskeho sociálneho fondu a Európskeho fondu regionálneho rozvoja v rámci Operačného programu Ľudské zdroje
www.minedu.sk www.employment.gov.sk/sk/esf/ www.itakademia.sk

Vapnik – Chervonenkisova dimenzia

Hypotézový priestor H má veľkú expresívnu silu, ak môže dosiahnuť všetky možné klasifikácie veľkej množiny príkladov. Miera tejto sily je **Vapnik – Chervonenkisova (VC) dimenzia**.

Definícia:

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

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

EURÓPSKA ÚNIA
Európsky sociálny fond
Európsky fond regionálneho rozvoja
OPERAČNÝ PROGRAM ĽUDSKÉ ZDROJE
MINISTERSTVO ŠKOLSTVA, VEDY, VÝSKUMU A ŠPORTU SLOVENSKEJ REPUBLIKY
UMB UNIVERZITA MATEJA BELA

Tento projekt sa realizuje vďaka podpore z Európskeho sociálneho fondu a Európskeho fondu regionálneho rozvoja v rámci Operačného programu Ľudské zdroje

www.minedu.sk www.employment.gov.sk/sk/esf/ www.itakademia.sk

Príklady

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

1. $X = \mathbb{R}$, $H \subseteq$ lúče; $x = (x_1, \ldots, x_m)$ je tréningová vzorka, $x_1 < x_2 < \cdots < x_m$.
   Pre $\Theta \in \mathbb{R}$, ohodnotenie príkladu je 1 $\quad \Leftrightarrow \quad x_i \geq \Theta$.
   Množina klasifikovaných príkladov $\ldots m+1$.
   Možné ohodnotenia: $(0\ldots00)$, $(0\ldots01)$, $\ldots(1\ldots11)$

   Ľubovoľná vzorka s rôznymi príkladmi má jedno z týchto ohodnotení (spermutované). $\Rightarrow$ $\Pi_H(m) = m+1$.

EURÓPSKA ÚNIA
Európsky sociálny fond
Európsky fond regionálneho rozvoja
OPERAČNÝ PROGRAM ĽUDSKÉ ZDROJE
MINISTERSTVO ŠKOLSTVA, VEDY, VÝSKUMU A ŠPORTU SLOVENSKEJ REPUBLIKY
UMB UNIVERZITA MATEJA BELA
Tento projekt sa realizuje vďaka podpore z Európskeho sociálneho fondu a Európskeho fondu regionálneho rozvoja v rámci Operačného programu Ľudské zdroje
www.minedu.sk www.employment.gov.sk/sk/esf/ www.itakademia.sk

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

   r_Θ, H je priestor lúčov. Zaujíma nás VC dimenzia.
   Platí $\Rightarrow$ $\Pi_H(m) = m + 1$

   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
   - $\Rightarrow$ $VCdim(H) = 1$.

EURÓPSKA ÚNIA
Európsky sociálny fond
Európsky fond regionálneho rozvoja
OPERAČNÝ PROGRAM ĽUDSKÉ ZDROJE
MINISTERSTVO ŠKOLSTVA, VEDY, VÝSKUMU A ŠPORTU SLOVENSKEJ REPUBLIKY
UMB UNIVERZITA MATEJA BELA
Tento projekt sa realizuje vďaka podpore z Európskeho sociálneho fondu a Európskeho fondu regionálneho rozvoja v rámci Operačného programu Ľudské zdroje
www.minedu.sk www.employment.gov.sk/sk/esf/ www.itakademia.sk

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

   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 $E_x - S$ 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.

EURÓPSKA ÚNIA
Európsky sociálny fond
Európsky fond regionálneho rozvoja
OPERAČNÝ PROGRAM ĽUDSKÉ ZDROJE
MINISTERSTVO ŠKOLSTVA, VEDY, VÝSKUMU A ŠPORTU SLOVENSKEJ REPUBLIKY
UMB UNIVERZITA MATEJA BELA
Tento projekt sa realizuje vďaka podpore z Európskeho sociálneho fondu a Európskeho fondu regionálneho rozvoja v rámci Operačného programu Ľudské zdroje
www.minedu.sk www.employment.gov.sk/sk/esf/ www.itakademia.sk

Veta:

Ak H je konečný hypotézový priestor, tak $VCdim(H) \leq \ln_2 |H|$.

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 $\Pi_H(d) = 2^d$.

$$2^d = \Pi_H(d) \leq |H| \quad \Rightarrow \quad d = \Pi_H(d) \leq \ln_2 |H|.$$

EURÓPSKA ÚNIA
Európsky sociálny fond
Európsky fond regionálneho rozvoja
OPERAČNÝ PROGRAM ĽUDSKÉ ZDROJE
MINISTERSTVO ŠKOLSTVA, VEDY, VÝSKUMU A ŠPORTU SLOVENSKEJ REPUBLIKY
UMB UNIVERZITA MATEJA BELA
Tento projekt sa realizuje vďaka podpore z Európskeho sociálneho fondu a Európskeho fondu regionálneho rozvoja v rámci Operačného programu Ľudské zdroje
www.minedu.sk www.employment.gov.sk/sk/esf/ www.itakademia.sk

VC dimenzia monočlenov - $VCdim(M_n)$

Platí $|M_n| = 3^n$, M_n sú monočleny.
$VCdim(M_n) \leq (\ln_2 3) * n$ je horná hranica.

Aká je dolná hranica? Pokúsime sa dokázať, že je rovná n.

Tvrdíme teda, že existuje vzorka dĺžky n, ktorú M_n rozbíja.

Vezmime vzorku n príkladov $(e_1, e_2, \ldots, e_n)$,
$e_i = (0 \ldots 010 \ldots 0)$, kde 1 je na i-tom mieste, $1 \leq i \leq n$.

EURÓPSKA ÚNIA
Európsky sociálny fond
Európsky fond regionálneho rozvoja
OPERAČNÝ PROGRAM ĽUDSKÉ ZDROJE
MINISTERSTVO ŠKOLSTVA, VEDY, VÝSKUMU A ŠPORTU SLOVENSKEJ REPUBLIKY
UMB UNIVERZITA MATEJA BELA
Tento projekt sa realizuje vďaka podpore z Európskeho sociálneho fondu a Európskeho fondu regionálneho rozvoja v rámci Operačného programu Ľudské zdroje
www.minedu.sk www.employment.gov.sk/sk/esf/ www.itakademia.sk

VC dimenzia monočlenov

Predpokladajme, že $(q_1, \ldots, q_n) = q \in \{0,1\}^n$ je dané ohodnotenie príkladov.

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, spĺňa všetko.
Ak $q = (0 \ldots 010 \ldots 10)$ tak h vytvoríme ako konjunkciu negácií literálov, pre ktoré $q_j = 0$.

$\Rightarrow$ $VCdim(M_n) \geq n$ pre ľubovoľné n.
Našli sme rozbitú tréningovú vzorku.

EURÓPSKA ÚNIA
Európsky sociálny fond
Európsky fond regionálneho rozvoja
OPERAČNÝ PROGRAM ĽUDSKÉ ZDROJE
MINISTERSTVO ŠKOLSTVA, VEDY, VÝSKUMU A ŠPORTU SLOVENSKEJ REPUBLIKY
UMB UNIVERZITA MATEJA BELA
Tento projekt sa realizuje vďaka podpore z Európskeho sociálneho fondu a Európskeho fondu regionálneho rozvoja v rámci Operačného programu Ľudské zdroje
www.minedu.sk www.employment.gov.sk/sk/esf/ www.itakademia.sk

VC dimenzia monočlenov

$n = 4$, príklady (e_1, e_2, e_3, e_4)

Tréningová vzorka: $(1, 0, 0, 0), (0, 1, 0, 0), (0, 0, 1, 0), (0, 0, 0, 1)$

Uvažujme ohodnotenie: $(1, 0, 1, 0)$

Hypotéza

$$\overline{u_2} \wedge \overline{u_4}$$

dáva ohodnotenie $(1, 0, 1, 0)$

EURÓPSKA ÚNIA
Európsky sociálny fond
Európsky fond regionálneho rozvoja
OPERAČNÝ PROGRAM ĽUDSKÉ ZDROJE
MINISTERSTVO ŠKOLSTVA, VEDY, VÝSKUMU A ŠPORTU SLOVENSKEJ REPUBLIKY
UMB UNIVERZITA MATEJA BELA
Tento projekt sa realizuje vďaka podpore z Európskeho sociálneho fondu a Európskeho fondu regionálneho rozvoja v rámci Operačného programu Ľudské zdroje
www.minedu.sk www.employment.gov.sk/sk/esf/ www.itakademia.sk

VC dimenzia reálnych perceptrónov

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

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

EURÓPSKA ÚNIA
Európsky sociálny fond
Európsky fond regionálneho rozvoja
OPERAČNÝ PROGRAM ĽUDSKÉ ZDROJE
MINISTERSTVO ŠKOLSTVA, VEDY, VÝSKUMU A ŠPORTU SLOVENSKEJ REPUBLIKY
UMB UNIVERZITA MATEJA BELA
Tento projekt sa realizuje vďaka podpore z Európskeho sociálneho fondu a Európskeho fondu regionálneho rozvoja v rámci Operačného programu Ľudské zdroje
www.minedu.sk www.employment.gov.sk/sk/esf/ www.itakademia.sk

VC dimenzia reálnych perceptrónov

Dôkaz:

Stav P_n je $\omega = (\alpha_1\alpha_2\ldots\alpha_n, \Theta)$; h_ω - funkcia, ktorú perceptrón počíta

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

Označme

$$I_\omega^+ = \{y \in \mathbb{R}^n : \quad \sum_{i=1}^n \alpha_i y_i \geq \Theta\}, \quad I_\omega^- = \{y \in \mathbb{R}^n : \quad \sum_{i=1}^n \alpha_i y_i < \Theta\}$$

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

EURÓPSKA ÚNIA
Európsky sociálny fond
Európsky fond regionálneho rozvoja
OPERAČNÝ PROGRAM ĽUDSKÉ ZDROJE
MINISTERSTVO ŠKOLSTVA, VEDY, VÝSKUMU A ŠPORTU SLOVENSKEJ REPUBLIKY
UMB UNIVERZITA MATEJA BELA

Tento projekt sa realizuje vďaka podpore z Európskeho sociálneho fondu a Európskeho fondu regionálneho rozvoja v rámci Operačného programu Ľudské zdroje

www.minedu.sk www.employment.gov.sk/sk/esf/ www.itakademia.sk

VC dimenzia reálnych perceptrónov

$C \subseteq \mathbb{R}^n$ 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;

EURÓPSKA ÚNIA
Európsky sociálny fond
Európsky fond regionálneho rozvoja
OPERAČNÝ PROGRAM ĽUDSKÉ ZDROJE
MINISTERSTVO ŠKOLSTVA, VEDY, VÝSKUMU A ŠPORTU SLOVENSKEJ REPUBLIKY
UMB UNIVERZITA MATEJA BELA
Tento projekt sa realizuje vďaka podpore z Európskeho sociálneho fondu a Európskeho fondu regionálneho rozvoja v rámci Operačného programu Ľudské zdroje
www.minedu.sk www.employment.gov.sk/sk/esf/ www.itakademia.sk

Pripomeňme Radonovu vetu:

Veta:

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

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

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

EURÓPSKA ÚNIA
Európsky sociálny fond
Európsky fond regionálneho rozvoja
OPERAČNÝ PROGRAM ĽUDSKÉ ZDROJE
MINISTERSTVO ŠKOLSTVA, VEDY, VÝSKUMU A ŠPORTU SLOVENSKEJ REPUBLIKY
UMB UNIVERZITA MATEJA BELA

Tento projekt sa realizuje vďaka podpore z Európskeho sociálneho fondu a Európskeho fondu regionálneho rozvoja v rámci Operačného programu Ľudské zdroje

www.minedu.sk www.employment.gov.sk/sk/esf/ www.itakademia.sk

Predpokladajme, že existuje h_ω v P_n také, že S je množina pozitívnych príkladov h_ω v E_x.

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

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

$\Rightarrow$ neexistuje taká h_ω, a preto x nie je rozbitá priestorom P_n.
$\Rightarrow$ žiadna vzorka dĺžky n+2 nie je rozbitá pomocou P_n.
$\Rightarrow$ $VCdim(P_n) \leq n + 1$.

EURÓPSKA ÚNIA
Európsky sociálny fond
Európsky fond regionálneho rozvoja
OPERAČNÝ PROGRAM ĽUDSKÉ ZDROJE
MINISTERSTVO ŠKOLSTVA, VEDY, VÝSKUMU A ŠPORTU SLOVENSKEJ REPUBLIKY
UMB UNIVERZITA MATEJA BELA
Tento projekt sa realizuje vďaka podpore z Európskeho sociálneho fondu a Európskeho fondu regionálneho rozvoja v rámci Operačného programu Ľudské zdroje
www.minedu.sk www.employment.gov.sk/sk/esf/ www.itakademia.sk

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

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

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

EURÓPSKA ÚNIA
Európsky sociálny fond
Európsky fond regionálneho rozvoja
OPERAČNÝ PROGRAM ĽUDSKÉ ZDROJE
MINISTERSTVO ŠKOLSTVA, VEDY, VÝSKUMU A ŠPORTU SLOVENSKEJ REPUBLIKY
UMB UNIVERZITA MATEJA BELA
Tento projekt sa realizuje vďaka podpore z Európskeho sociálneho fondu a Európskeho fondu regionálneho rozvoja v rámci Operačného programu Ľudské zdroje
www.minedu.sk www.employment.gov.sk/sk/esf/ www.itakademia.sk

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

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 $\Pi_H(m) = 2^m$.

EURÓPSKA ÚNIA
Európsky sociálny fond
Európsky fond regionálneho rozvoja
OPERAČNÝ PROGRAM ĽUDSKÉ ZDROJE
MINISTERSTVO ŠKOLSTVA, VEDY, VÝSKUMU A ŠPORTU SLOVENSKEJ REPUBLIKY
UMB UNIVERZITA MATEJA BELA
Tento projekt sa realizuje vďaka podpore z Európskeho sociálneho fondu a Európskeho fondu regionálneho rozvoja v rámci Operačného programu Ľudské zdroje
www.minedu.sk www.employment.gov.sk/sk/esf/ www.itakademia.sk

Veta:

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

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

Zaveďme funkciu:

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

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

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

EURÓPSKA ÚNIA
Európsky sociálny fond
Európsky fond regionálneho rozvoja
OPERAČNÝ PROGRAM ĽUDSKÉ ZDROJE
MINISTERSTVO ŠKOLSTVA, VEDY, VÝSKUMU A ŠPORTU SLOVENSKEJ REPUBLIKY
UMB UNIVERZITA MATEJA BELA

Tento projekt sa realizuje vďaka podpore z Európskeho sociálneho fondu a Európskeho fondu regionálneho rozvoja v rámci Operačného programu Ľudské zdroje

www.minedu.sk www.employment.gov.sk/sk/esf/ www.itakademia.sk

Idea dôkazu: Ak $VCdim(H) = d = 0$, potom pre ľubovoľný príklad x, je $h(x)$ rovnaké (buď 0 alebo 1) pre všetky hypotézy $h \in H$.
$\Rightarrow \quad \Pi_H(x) = 1$ pre ľubovoľnú vzorku $\mathbf{x}$ dĺžky m.

$\Rightarrow \Pi_H(m) = 1 = \Phi(0, m) \quad \Rightarrow$ veta platí pre $d = 0$.

Ak $m = 1$ a $d \geq 1$, tak pre ľubovoľné H máme
$\Rightarrow \quad \Pi_H(1) \leq 2 = \Phi(d, 1)$, takže veta je pravdivá aj v tomto prípade.

EURÓPSKA ÚNIA
Európsky sociálny fond
Európsky fond regionálneho rozvoja
OPERAČNÝ PROGRAM ĽUDSKÉ ZDROJE
MINISTERSTVO ŠKOLSTVA, VEDY, VÝSKUMU A ŠPORTU SLOVENSKEJ REPUBLIKY
UMB UNIVERZITA MATEJA BELA
Tento projekt sa realizuje vďaka podpore z Európskeho sociálneho fondu a Európskeho fondu regionálneho rozvoja v rámci Operačného programu Ľudské zdroje
www.minedu.sk www.employment.gov.sk/sk/esf/ www.itakademia.sk

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 $VCdim(H) = d$ a $\mathbf{x}$ je tréningová vzorka dĺžky m, kde $d + m = k + 1$. Prípady $(d, m) = (0, k + 1)$ a $(d, m) = (k, 1)$ sú už dokázané.
  Nech $d \geq 1$, $m \geq 2$, $\mathbf{x}$ obsahuje aj opakované príklady, E je množina príkladov v $\mathbf{x}$, $H_E = H|E$ je obmedzenie hypotéz H na E.

  $\Rightarrow$ H_E je konečný a $\Pi_H(x) = |H_E|$.
  Potrebujeme ukázať, že $|H_E| \leq \Phi(d, m)$.
  Ponechané na vlastnú tvorivosť

EURÓPSKA ÚNIA
Európsky sociálny fond
Európsky fond regionálneho rozvoja
OPERAČNÝ PROGRAM ĽUDSKÉ ZDROJE
MINISTERSTVO ŠKOLSTVA, VEDY, VÝSKUMU A ŠPORTU SLOVENSKEJ REPUBLIKY
UMB UNIVERZITA MATEJA BELA
Tento projekt sa realizuje vďaka podpore z Európskeho sociálneho fondu a Európskeho fondu regionálneho rozvoja v rámci Operačného programu Ľudské zdroje
www.minedu.sk www.employment.gov.sk/sk/esf/ www.itakademia.sk

Príklad

Nech H je hypotézový priestor reálnych perceptrónov P_n. Potom $VCdim(H) = n + 1$, a preto pre ľubovoľné $m > 0$ platí $\Pi_H(m) \leq \Phi(n + 1, m)$.
Napríklad pre $n = 2$

$$\Pi_H(4) \leq \Phi(3, 4) = 1 + 4 + 6 + 4 = 15.$$

Skutočná analýza, ale ukáže, že $\Pi_H(4) = 14$.

Dá sa ukázať, že $\Phi(d, m)$ je polynomiálnou funkciou v m stupňa d:

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

EURÓPSKA ÚNIA
Európsky sociálny fond
Európsky fond regionálneho rozvoja
OPERAČNÝ PROGRAM ĽUDSKÉ ZDROJE
MINISTERSTVO ŠKOLSTVA, VEDY, VÝSKUMU A ŠPORTU SLOVENSKEJ REPUBLIKY
UMB UNIVERZITA MATEJA BELA

Tento projekt sa realizuje vďaka podpore z Európskeho sociálneho fondu a Európskeho fondu regionálneho rozvoja v rámci Operačného programu Ľudské zdroje

www.minedu.sk www.employment.gov.sk/sk/esf/ www.itakademia.sk

Úlohy

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

EURÓPSKA ÚNIA
Európsky sociálny fond
Európsky fond regionálneho rozvoja
OPERAČNÝ PROGRAM ĽUDSKÉ ZDROJE
MINISTERSTVO ŠKOLSTVA, VEDY, VÝSKUMU A ŠPORTU SLOVENSKEJ REPUBLIKY
UMB UNIVERZITA MATEJA BELA
Tento projekt sa realizuje vďaka podpore z Európskeho sociálneho fondu a Európskeho fondu regionálneho rozvoja v rámci Operačného programu Ľudské zdroje
www.minedu.sk www.employment.gov.sk/sk/esf/ www.itakademia.sk

Úlohy

4. Monočlen je monotónny, ak neobsahuje žiadne negované literály. Dokážte, že priestor monotónnych monočlenov definovaný na $\{0,1\}^*$ má VC dimenziu práve n.
5. 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, $VCdim(H) = 1$.

EURÓPSKA ÚNIA
Európsky sociálny fond
Európsky fond regionálneho rozvoja
OPERAČNÝ PROGRAM ĽUDSKÉ ZDROJE
MINISTERSTVO ŠKOLSTVA, VEDY, VÝSKUMU A ŠPORTU SLOVENSKEJ REPUBLIKY
Tento projekt sa realizuje vďaka podpore z Európskeho sociálneho fondu a Európskeho fondu regionálneho rozvoja v rámci Operačného programu Ľudské zdroje
www.minedu.sk www.employment.gov.sk/sk/esf/ www.itakademia.sk

Úlohy

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

EURÓPSKA ÚNIA
Európsky sociálny fond
Európsky fond regionálneho rozvoja
OPERAČNÝ PROGRAM ĽUDSKÉ ZDROJE
MINISTERSTVO ŠKOLSTVA, VEDY, VÝSKUMU A ŠPORTU SLOVENSKEJ REPUBLIKY
UMB UNIVERZITA MATEJA BELA
Tento projekt sa realizuje vďaka podpore z Európskeho sociálneho fondu a Európskeho fondu regionálneho rozvoja v rámci Operačného programu Ľudské zdroje
www.minedu.sk www.employment.gov.sk/sk/esf/ www.itakademia.sk