Neuronové sítě

Doc. RNDr. Iveta Mrázová, CSc.

Katedra teoretické informatiky
Matematicko-fyzikální fakulta
Univerzity Karlovy v Praze

Neuronové sítě

– Vrstevnaté neuronové sítě – – analýza vlastností –

Doc. RNDr. Iveta Mrázová, CSc.

Katedra teoretické informatiky
Matematicko-fyzikální fakulta
Univerzity Karlovy v Praze

Kolmogorovova věta - 1957

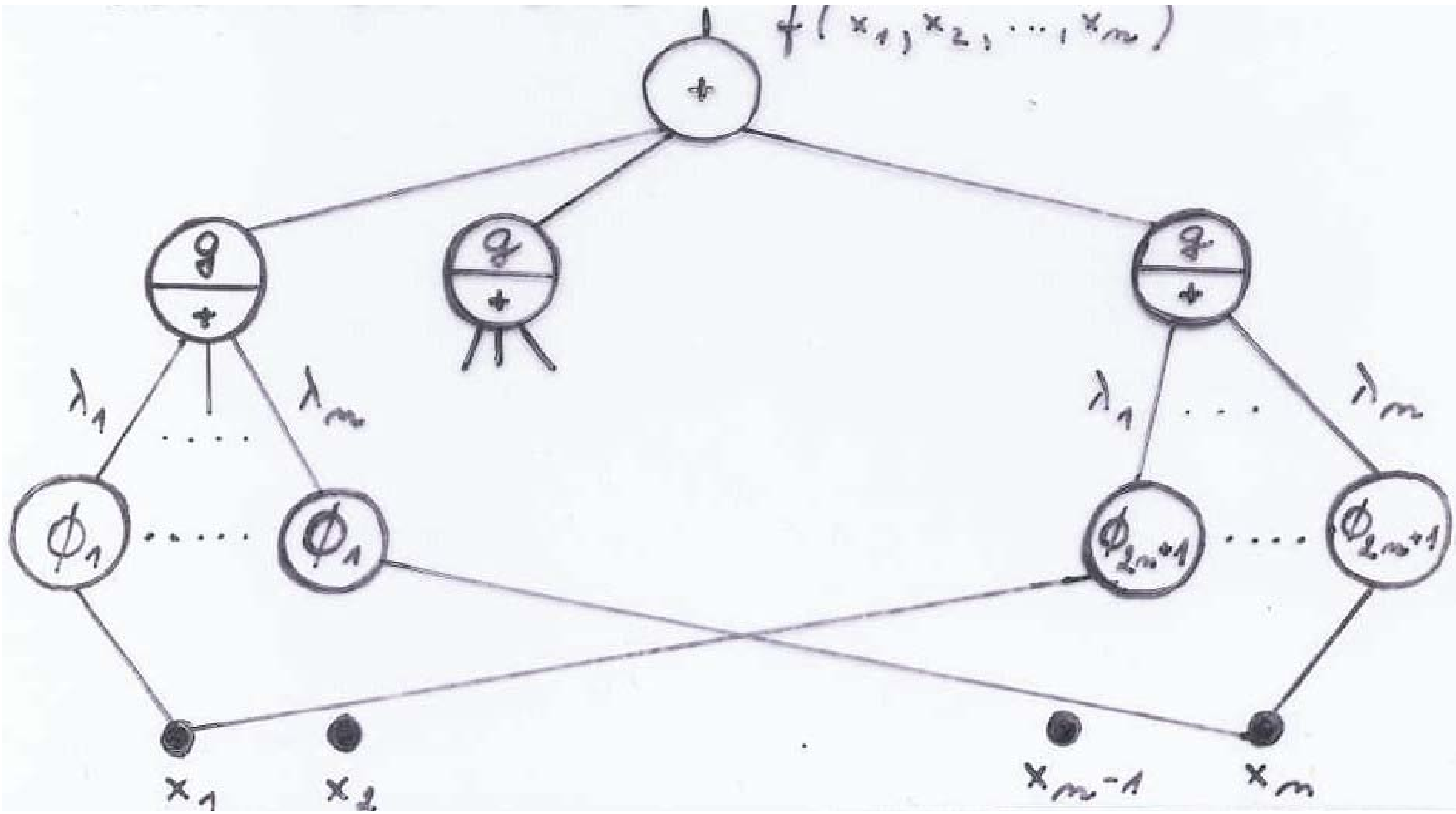
13. Hilbertův problém ~ spojité funkce ***n*** proměnných lze vyjádřit pomocí konečného počtu funkcí jediné proměnné a sčítání

- Příklad: $x \cdot y = \exp(\ln x + \ln y)$

V: Nechť $f : [0, 1]^n \rightarrow [0, 1]$ je spojitá funkce. Potom existují funkce jediné proměnné g a Φ_q, pro $q = 1, \ldots, 2n+1$ a konstanty λ_p, pro $p = 1, \ldots, n$ takové, že

$$f(x_1, \ldots, x_n) = \sum_{q=1}^{2n+1} g\left(\sum_{p=1}^{n} \lambda_p \Phi_q(x_p) \right)$$

Kolmogorovovy sítě



Aproximace funkcí (1)

- Libovolnou spojitou funkci lze vyjádřit pomocí sítě s odpovídajícím počtem výpočetních jednotek (× volba vhodné přenosové funkce)
- Nejlepší možná aproximace dané funkce (× volba vhodného počtu výpočetních jednotek s uvažovanou přenosovou funkcí)

Aproximace funkcí (2)

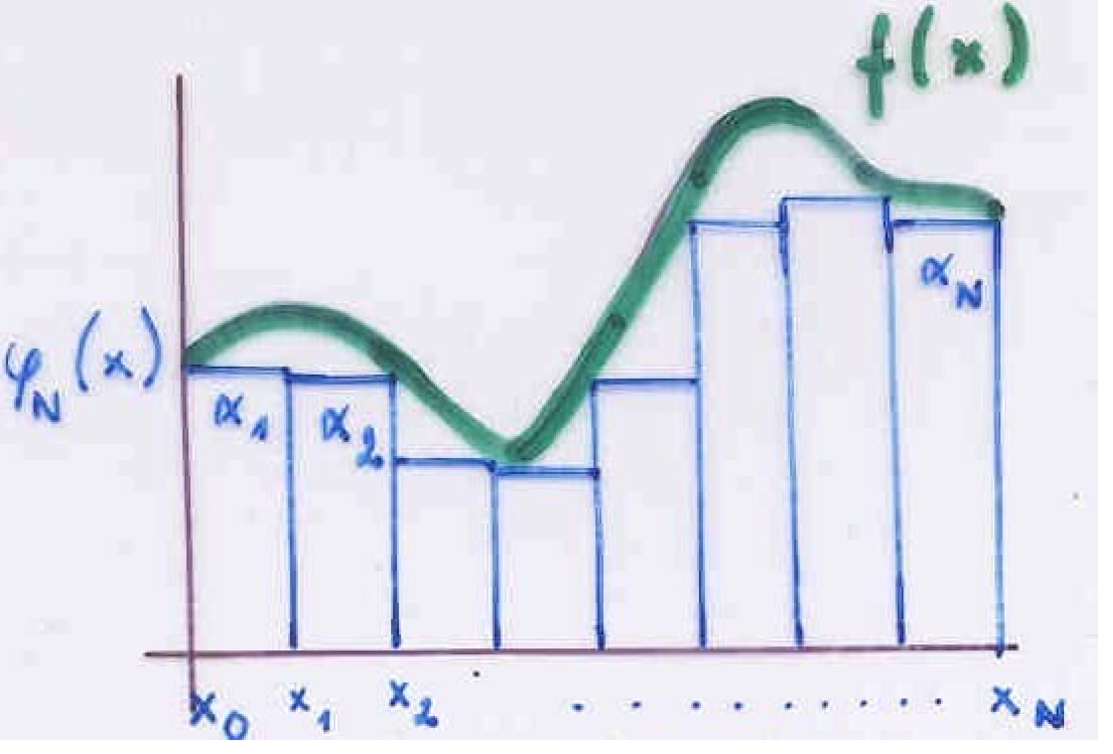
V: Spojitou reálnou funkci $f : [\,0\,,\,1\,] \rightarrow [\,0\,,\,1\,]$ lze aproximovat pomocí sítě prahových jednotek tak, že celková aproximační chyba E je menší než libovolné reálné číslo $\varepsilon > 0$:

$$E = \int_0^1 \left| f(x) - \tilde{f}(x) \right| dx < \varepsilon$$

kde $\tilde{f}$ označuje funkci realizovanou sítí prahových jednotek.

Aproximace funkcí (3)

Důkaz: Idea ~ aproximace f pomocí φ_N



Aproximace funkcí (4)

Důkaz (pokračování):

- Rozdělme interval $[\,0\,,1\,]$ do N stejně velkých podintervalů pomocí bodů $x_0, x_1, \ldots, x_N \in [\,0\,,1\,]\,;\, x_0 = 0,\, x_N = 1$
- Funkci φ_N definujme jako:

$$\varphi_N(x) = \min\{f(x'); x' \in [x_i, x_{i+1}] \text{ pro } x_i \leq x \leq x_{i+1}\}$$

- Dále nechť funkce φ_N aproximuje funkci f tak, že aproximační chyba E_N odpovídá:

$$E_N = \int_0^1 |\, f(x) - \varphi_N(x) \,| \; dx$$

Aproximace funkcí (5)

Důkaz (pokračování):

- Protože $f(x) \geq \varphi_N(x) \quad \forall x \in [0, 1]$, odpovídá

$$E_N = \int_0^1 f(x)\, dx - \int_0^1 \varphi_N(x)\, dx$$

~ dolní součet při výpočtu Riemannova integrálu f

- Spojité funkce lze integrovat $\rightarrow$ dolní součet konverguje pro $N \rightarrow \infty$ k integrálu f na intervalu $[0, 1]$
- Platí tedy, že $E_N \rightarrow 0$ pro $N \rightarrow \infty$, a proto pro libovolné reálné $\varepsilon > 0$ existuje M takové, že $E_N < \varepsilon \quad \forall N \geq M$
- Funkce φ_N je tedy požadovanou aproximací f.

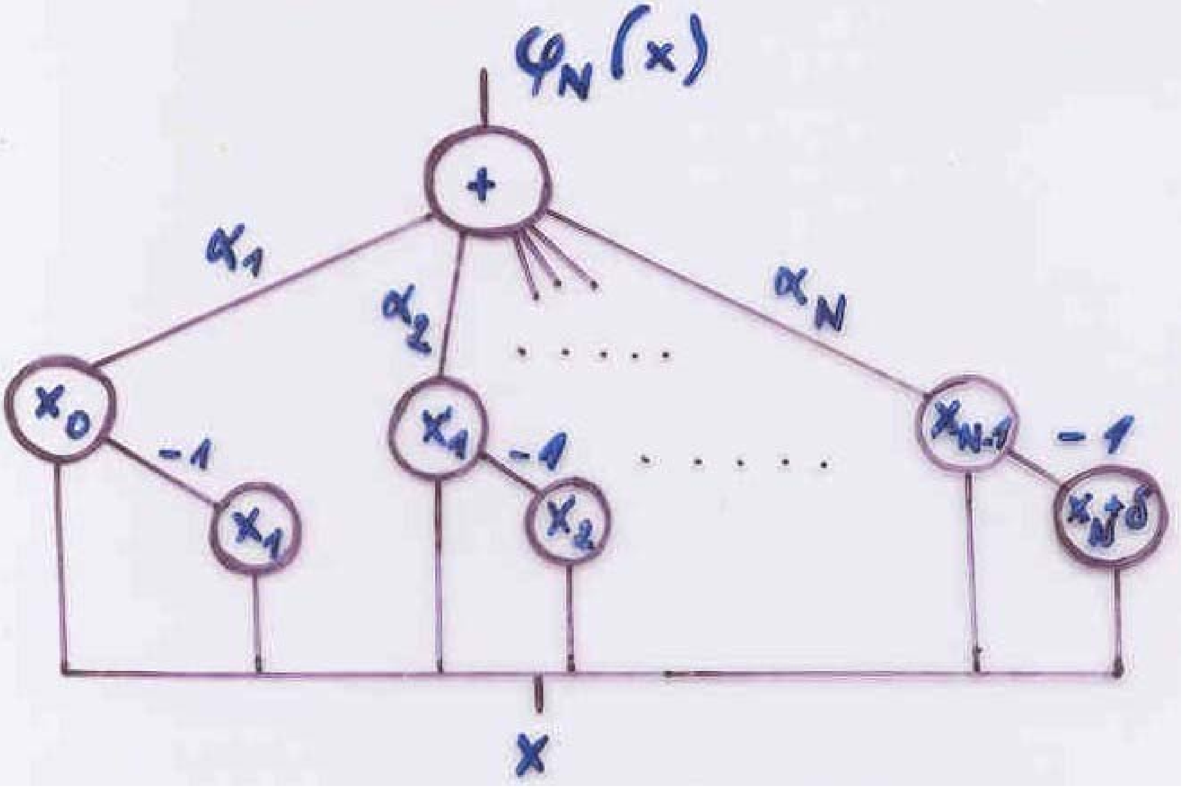
Aproximace funkcí (6)

Důkaz (pokračování):

- Realizace funkce $\varphi_N(x)$ pomocí sítě prahových jednotek (~ neuronová síť)
  - Funkce $\varphi_N(x)$ je skoková a v každém z N podintervalů $[0, 1]: [x_0, x_1), [x_1, x_2), \ldots, [x_{N-1}, x_N]$ nabývá odpovídající hodnoty $\alpha_0, \ldots, \alpha_N$

Aproximace funkcí (7)

Důkaz (pokračování):



Aproximace funkcí (8)

Důkaz (pokračování):

- Tato síť je schopna realizovat skokovou funkci $\varphi_N(x)$:
  - Jediným vstupem sítě je x
  - Každá dvojice jednotek s prahem x_i a x_{i+1} zaručuje, že jednotka s prahem x_i bude aktivní pouze pokud $x_i \leq x < x_{i+1}$.
  - Výstupní jednotka sčítá všechny výstupy předchozí vrstvy jednotek a jako výsledek vydá jejich (vážený) součet
  - Jednotka s prahem $x_N+\delta$, kde δ je malé kladné číslo, slouží k rozpoznání případů $x_{N-1} \leq x \leq x_N$.
- Tato síť realizuje funkci φ_N, která aproximuje funkci f nanejvýš s požadovanou chybou. ***QED***

Aproximace funkcí (9)

Poznámka:

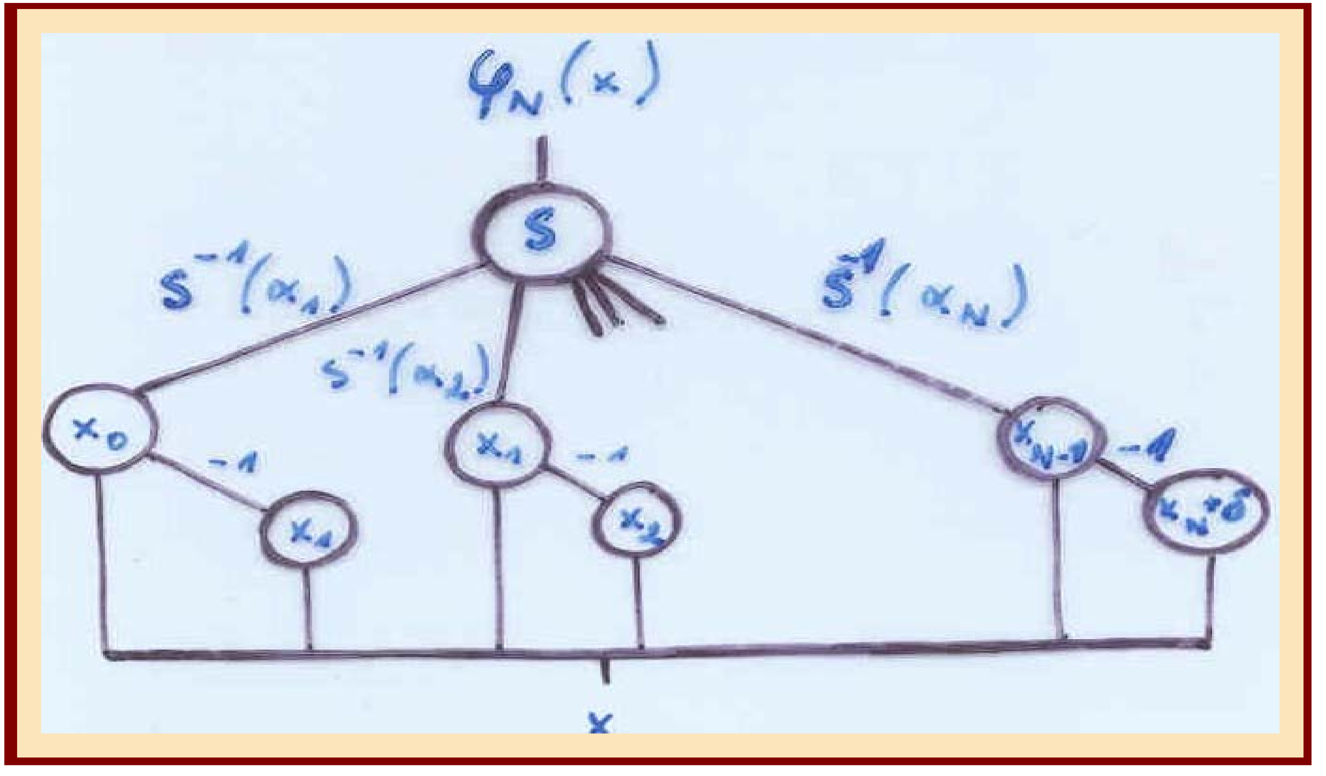
Věta platí i pro jednotky (~ neurony) se sigmoidální přenosovou funkcí, kde $f : [\,0\,,\,1\,] \rightarrow (\,0\,,\,1)$

Důkaz:

- Obor hodnot funkce f je búno omezen na interval $(\,0\,,\,1)$
- Funkci f lze aproximovat pomocí sítě

Aproximace funkcí (10)

Důkaz (pokračování):



Aproximace funkcí (11)

Důkaz (pokračování):

- Přenosová funkce jednotek s prahem x_i je dána pomocí $s_c(x - x_i)$, kde parametr c určuje strmost přenosové funkce

$$s_c(x - x_i) = \frac{1}{1 + e^{-c(x - x_i)}}$$

- Síť realizuje odhad funkce φ_N s takovou aproximační chybou, která je menší než libovolná požadovaná mez (> 0)

  (~ prahové funkce lze pomocí parametrizované sigmoidy aproximovat s libovolnou přesností)

Aproximace funkcí (12)

Důkaz (pokračování):

- Váhy (synapsí) mezi první vrstvou jednotek a výstupní jednotkou byly nastaveny tak, že sigmoida bude předávat jako výsledek požadované hodnoty α_i
- Dále je třeba zaručit, že pro každý vstup x bude první vrstva předávat výstupní vrstvě jen jedinou jedničku

  → první vrstva sítě určí, ke kterému z N segmentů x patří

QED

Aproximace funkcí (13)

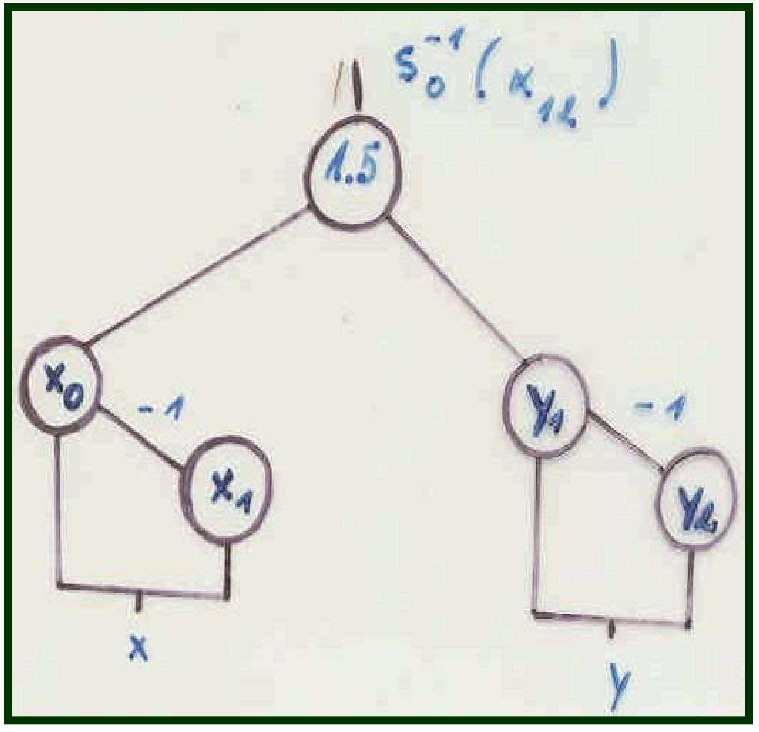
Vícerozměrný případ:

Síť pro aproximaci funkce $f: [0,1]^n \rightarrow (0,1)$ lze zkonstruovat na základě předchozích myšlenek:

- nutná rozšíření pro dvourozměrný případ

- Rozpoznání x -ových i y –ových „intervalů“
  - 2 jednotky vlevo pro $x_0 \leq x < x_1$
  - 2 jednotky vlevo pro $y_0 \leq y < y_1$
- Jednotka s prahem 1.5 rozpoznává konjunkci obou podmínek (pro x a y)

Aproximace funkcí (14)



- „Výstup“ má váhu $s_0^{-1}(\alpha_{12})$, takže výstupní jednotka se sigmoidou dává α_{12}

→ toto číslo odpovídá požadované aproximaci funkce f v intervalu: $[x_0, x_1) \times [y_1, y_2)$

NP-úplnost problému učení

Problém splnitelnosti

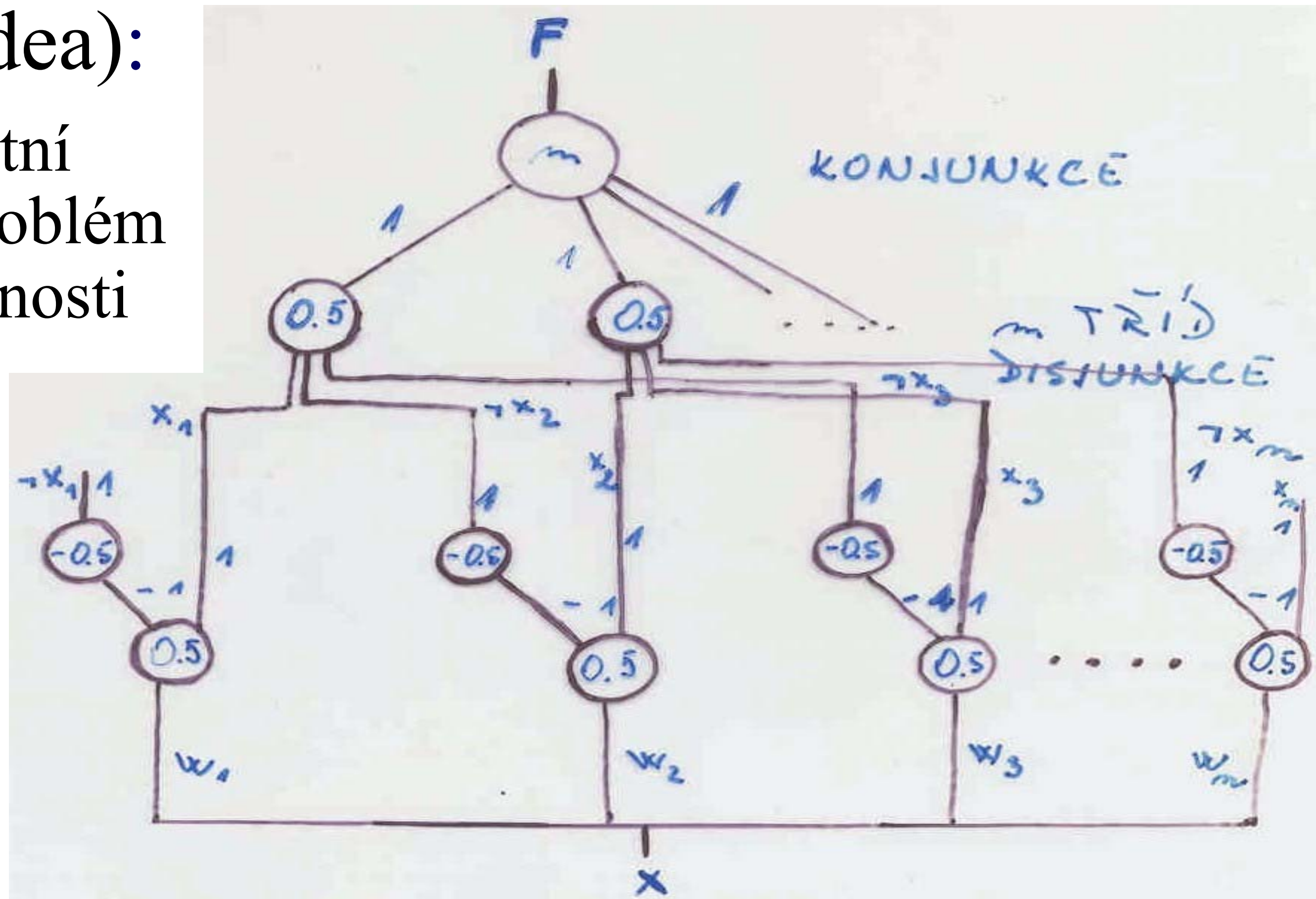
D: Nechť V je množina n logických proměnných a nechť F je logická formule v konjunktivní normální formě, která obsahuje jen proměnné z V. Problém splnitelnosti spočívá v přiřazení pravdivostních hodnot proměnným z V tak, aby měla formule F pravdivostní hodnotu *TRUE*.

V: **Obecný problém učení je pro sítě prahových jednotek NP-úplný.**

NP-úplnost problému učení (2)

Důkaz (idea):

ekvivalentní síť pro problém 3-splnitelnosti



NP-úplnost problému učení (3)

Důkaz (pokračování):

1. 3-splnitelnost logických formulí (3-SAT) lze redukovat (převést) na problém učení neuronových sítí

Logickou formuli $\boldsymbol{F}$ v konjunktivní normální formě, která obsahuje $\boldsymbol{n}$ proměnných, lze v polynomiálním čase převést na síť výše uvedeného typu:

- Každé proměnné $\boldsymbol{x_i}$ je přiřazena váha $\boldsymbol{w_i}$
- Spoje k výpočetním jednotkám ze třetí vrstvy jsou určeny příslušnou konjunktivní normální formou

NP-úplnost problému učení (4)

Důkaz (pokračování):

- Tyto operace lze provést (při vhodném kódování) v polynomiálním čase, protože pro $\boldsymbol{m}$ různých disjunkcí v 3-SAT-výrazu platí, že $\boldsymbol{m \leq (2n)^3}$

- Jestliže existuje instance $\boldsymbol{A}$ s pravdivostními hodnotami proměnných $\boldsymbol{x_i}$ taková, že $\boldsymbol{F}$ je splněna, potom existují váhy $\boldsymbol{w_1, w_2, \ldots, w_n}$, které řeší problém učení

NP-úplnost problému učení (5)

Důkaz (pokračování):

- Postačí zvolit váhy $w_i = 1$, jestliže $x_i = 1$;
  a $w_i = 0$, jestliže $x_i = 0$.
  (V obou případech tedy zvolíme $w_i = x_i$.)
- Podobně i opačným způsobem:
  jestliže existují váhy $w_1, w_2, \ldots, w_n$, které řeší problém učení, potom vede instance $x_i = 1$ pro $w_i \geq 0.5$ a $x_i = 0$ jinak ke splnění F

NP-úplnost problému učení (6)

Důkaz (pokračování):

2. Dále je třeba ukázat, že problém učení patří do třídy NP (řešení lze ověřit v polynomiálním čase)
   - Jsou-li dány váhy $w_1, w_2, \ldots, w_n$, potom lze po jediném „průchodu sítí“ ověřit, zda je její výstup F roven 1
   - Počet výpočetních kroků přímo závisí na počtu proměnných n a na počtu disjunkcí m (který je omezen polynomem $(2n)^3$)

NP-úplnost problému učení (7)

Důkaz (pokračování):

- Čas potřebný k ověření dané („uhodnuté“) instance je tedy omezen polynomem nad ***n***
- Daný problém učení tedy patří do třídy NP

QED

Poznámka:

Pro některé speciální typy jednoduchých neuronových sítí je problém učení řešitelný v polynomiálním čase (pomocí metod lineárního programování)

Počet oblastí v příznakovém prostoru (1)

- Kapacita neuronu závisí na dimenzi váhového prostoru a počtu „řezů dělicích nadrovin“

→ **Otázka:**
Kolik oblastí je určeno $\boldsymbol{m}$ dělicími nadrovinami dimenze $\boldsymbol{n - 1}$ v $\boldsymbol{n}$ - rozměrném prostoru?

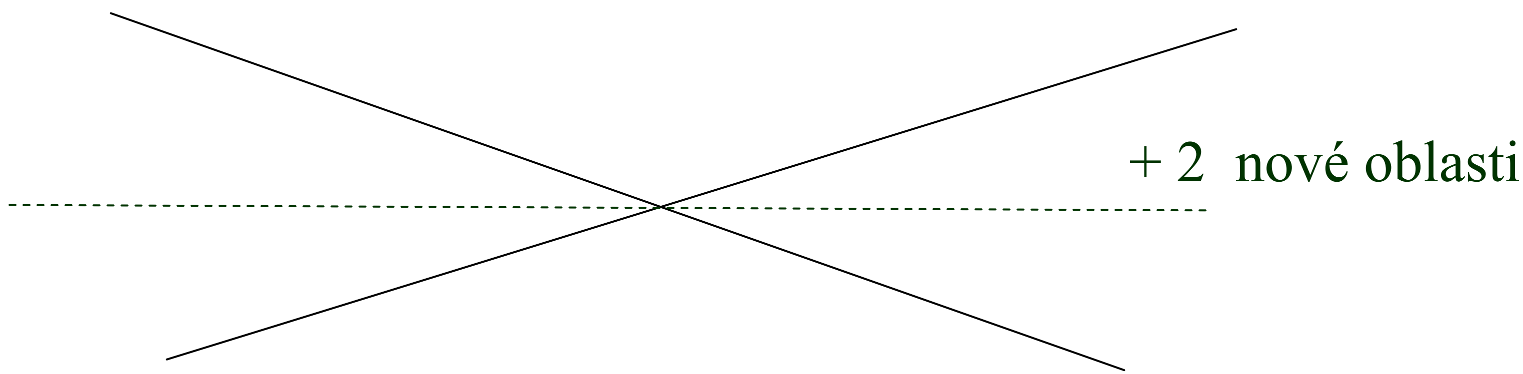
- přitom budeme uvažovat pouze takové nadroviny, které procházejí počátkem

→ Průnik $\boldsymbol{l}$ nadrovin ; $\boldsymbol{l} \leq \boldsymbol{n}$ bude dimenze $\boldsymbol{n - l}$

Počet oblastí v příznakovém prostoru (2)

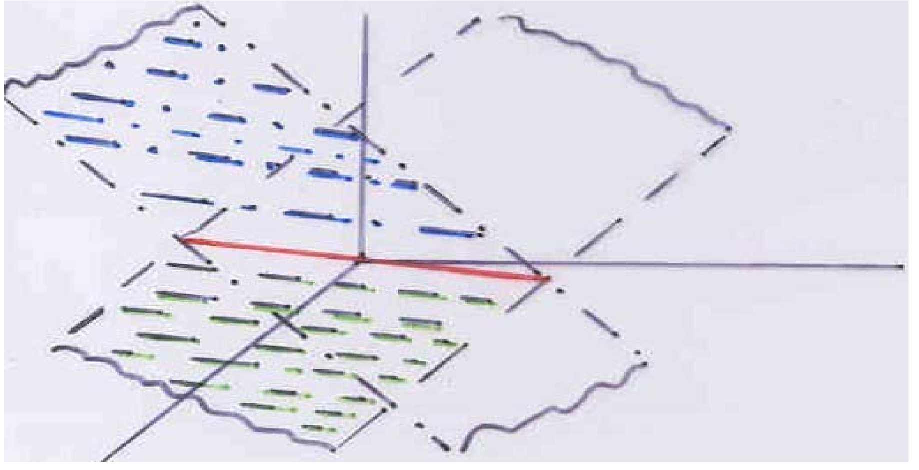
- <u>2 – rozměrný případ:</u>

m přímek pocházejících počátkem vytváří nanejvýš ***2 · m*** navzájem různých oblastí



Počet oblastí v příznakovém prostoru (3)

- <u>3 – rozměrný případ:</u>
  - každý nový řez zvýší počet oblastí až 2 ×



- <u>obecně:</u> ***n*** řezů ***(n – 1)*** – rozměrnými nadrovinami v ***n*** – rozměrném prostoru vytváří nanejvýš ***2^n*** různých oblastí

Počet oblastí v příznakovém prostoru (4)

Věta: Nechť $R(m, n)$ označuje počet oblastí určených m různými dělicími nadrovinami dimenze $n - 1$ v n – rozměrném prostoru. Dále nechť $R(1, n) = 2$ pro $n \geq 1$ a $R(m, 0) = 0 \quad \forall m \geq 1$.

Potom pro $n \geq 1$ a $m > 1$:

$$R(m, n) = R(m-1, n) + R(m-1, n-1)$$

Počet oblastí v příznakovém prostoru (5)

Důkaz (indukcí přes m):

1. $m = 2$ a $n = 1$: Platí, protože

   $R(2, 1) = R(1, 1) + R(1, 0) = 2 + 0 = 2$

2. $m = 2$ a $n \geq 2$: $R(2, n) = 4$ => platí, protože

   $R(2, n) = R(1, n) + R(1, n - 1) = 2 + 2 = 4$

3. $m + 1$ nadrovin dimenze $n - 1$ v n-rozměrném prostoru ($n \geq 2$) :

   - Prvních m nadrovin určuje $R(m, n)$ oblastí v n – rozměrném prostoru

Počet oblastí v příznakovém prostoru (6)

Důkaz (pokračování):

- $(m+1)$ – ní nadrovina protíná prvních m nadrovin v m nadrovinách dimenze $n-2$
- Těchto m nadrovin (dimenze $n-2$) rozděluje $(n-1)$ - rozměrný prostor do $R(m, n-1)$ oblastí
- Po řezu $(m+1)$-ní nadrovinou vzniklo $R(m, n-1)$ nových oblastí

→ Nový počet oblastí je tedy:

$$R(m+1, n) = R(m, n) + R(m, n-1)$$

QED

Počet oblastí v příznakovém prostoru (7)

- Možná alternativa výpočtu podle:

$$R(m,n) = 2\sum_{i=0}^{n-1} \binom{m-1}{i}$$

- S rostoucím ***n*** roste počet Boolovských funkcí výrazně rychleji než počet různých oblastí vytvořených nadrovinami v obecné poloze
  - tento počet může být obecně větší než počet prahových funkcí nad binárními vstupy

Počet oblastí v příznakovém prostoru (8)

Příklad:

n	POČET BOOLOVSKÝCH FUNKCÍ (2^{2^n})	POČET PRAHOVÝCH FUNKCÍ ($T(2^n, n)$)	POČET OBLASTÍ $R(2^n, n)$
1	4	2	2
2	16	14	14
3	256	104	128
4	65 536	1 882	3 882
5	4.3×10^9	94 572	412 736

Počet oblastí v příznakovém prostoru (9)

Důsledek:

Problémy s učením ~ je-li počet vstupních vektorů příliš vysoký, nemusí být síť schopna vytvořit s předem daným pevným počtem skrytých neuronů potřebný počet oblastí

- **Zobecňování (generalizace)**
  - ~ očekávaný počet správně klasifikovaných vzorů
- **Přeučení**
  - ~ chybná interpolace vzorů mimo trénovací množinu
- **Vapnik – Chervonenkisova dimenze (VC-dimenze)**
  - ~ konečná VC-dimenze → „třídu konceptů“ lze naučit

Vapnik – Chervonenkisova dimenze (VC–dimenze) (1)

D: Nechť $C = \{ f_i \}$ je množina funkcí (concept class) **Množinu** m trénovacích vzorů $\{ t_k \}_{k=1,\ldots,m}$ **lze rozčlenit** pomocí C, jestliže pro každé ze 2^m možných označení těchto vzorů $1 / 0$, existuje alespoň jedna funkce f_i, která tomuto označení vyhovuje.

D: **VC-dimenze** V **množiny funkcí** C je definována jako největší m, pro které existuje množina m rozčlenitelných trénovacích vzorů.

Vapnik – Chervonenkisova dimenze (VC–dimenze) (2)

- Pokud existuje pro libovolné ***m*** množina ***m*** trénovacích vzorů, které lze rozčlenit pomocí ***C***, je VC-dimenze ***C*** rovna nekonečnu

  → Takový problém je „**N E N A U Č I T E L N Ý**“

- VC-dimenze množiny funkcí obecně nezávisí na počtu parametrů
- VC-dimenze je důležitá pro správné zobecňování sítí
  - Síť může mít mnoho parametrů, ale měla by mít malou VC-dimenzi → lepší generalizace
  - Velká VC-dimenze bývá spojena s horší generalizací

Vapnik – Chervonenkisova dimenze (VC–dimenze) (3)

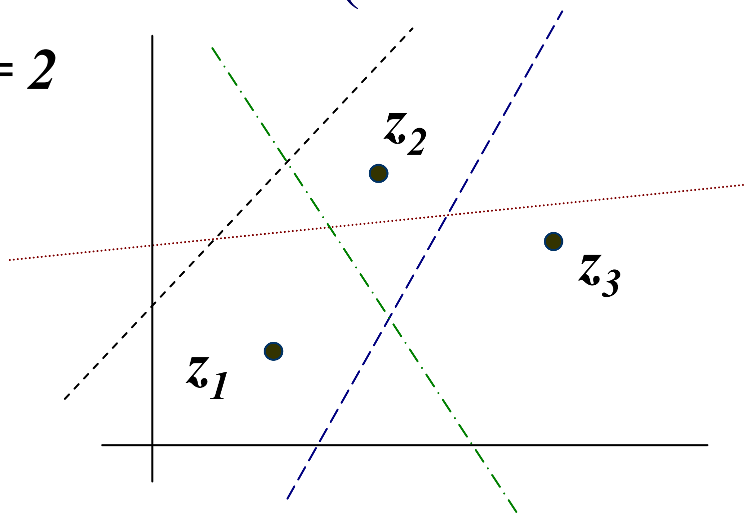
Příklad:

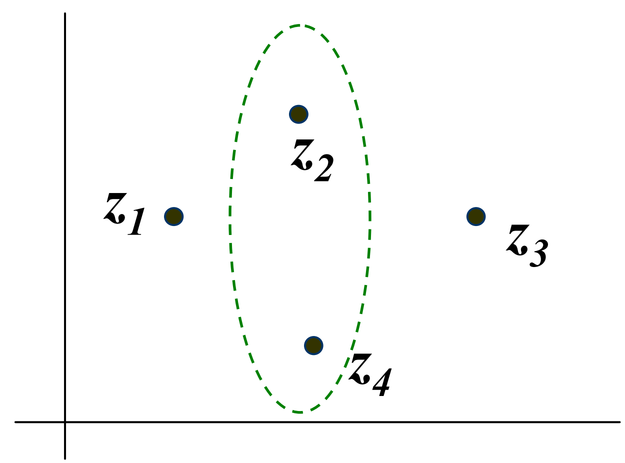
1. VC-dimenze množiny lineárních indikačních funkcí

$$Q(\vec{z},\alpha) = \Theta\left\{\sum_{p=1}^{n}\alpha_p z_p + \alpha_0\right\}$$ v $\boldsymbol{n}$ – rozměrném prostoru

je rovna $\boldsymbol{n+1}$ (tzn. lze rozčlenit nanejvýš $\boldsymbol{n+1}$ vzorů)

$\boldsymbol{n = 2}$





Vapnik – Chervonenkisova dimenze (VC–dimenze) (4)

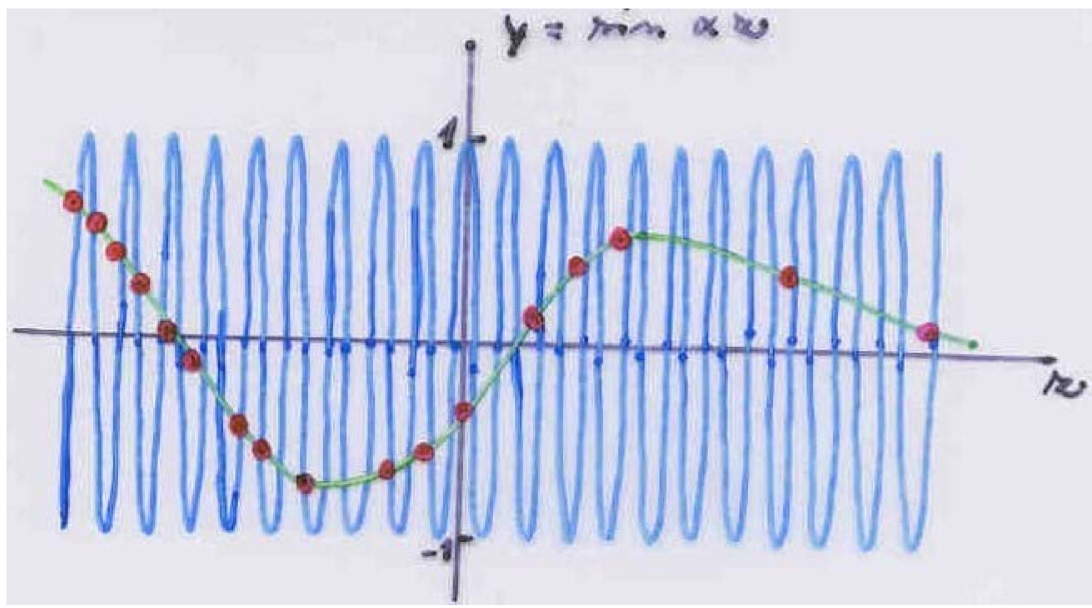
2. VC-dimenze množiny následujících funkcí
   $f(z, \alpha) = \theta(\sin \alpha z)\ ,\ \alpha \in R$ je nekonečná

   - Body $z_1 = 10^{-1}, \ldots, z_m = 10^{-m}$ lze rozčlenit pomocí funkcí z této množiny
   - K rozčlenění těchto vzorů do dvou tříd $(+1 / -1)$ daných posloupností $\delta_1, \ldots, \delta_m;\ \delta_i \in \{0, 1\}$ stačí zvolit hodnotu parametru

$$\alpha = \pi \left(\sum_{i=1}^{m} (1 - \delta_i) 10^{i} + 1 \right)$$

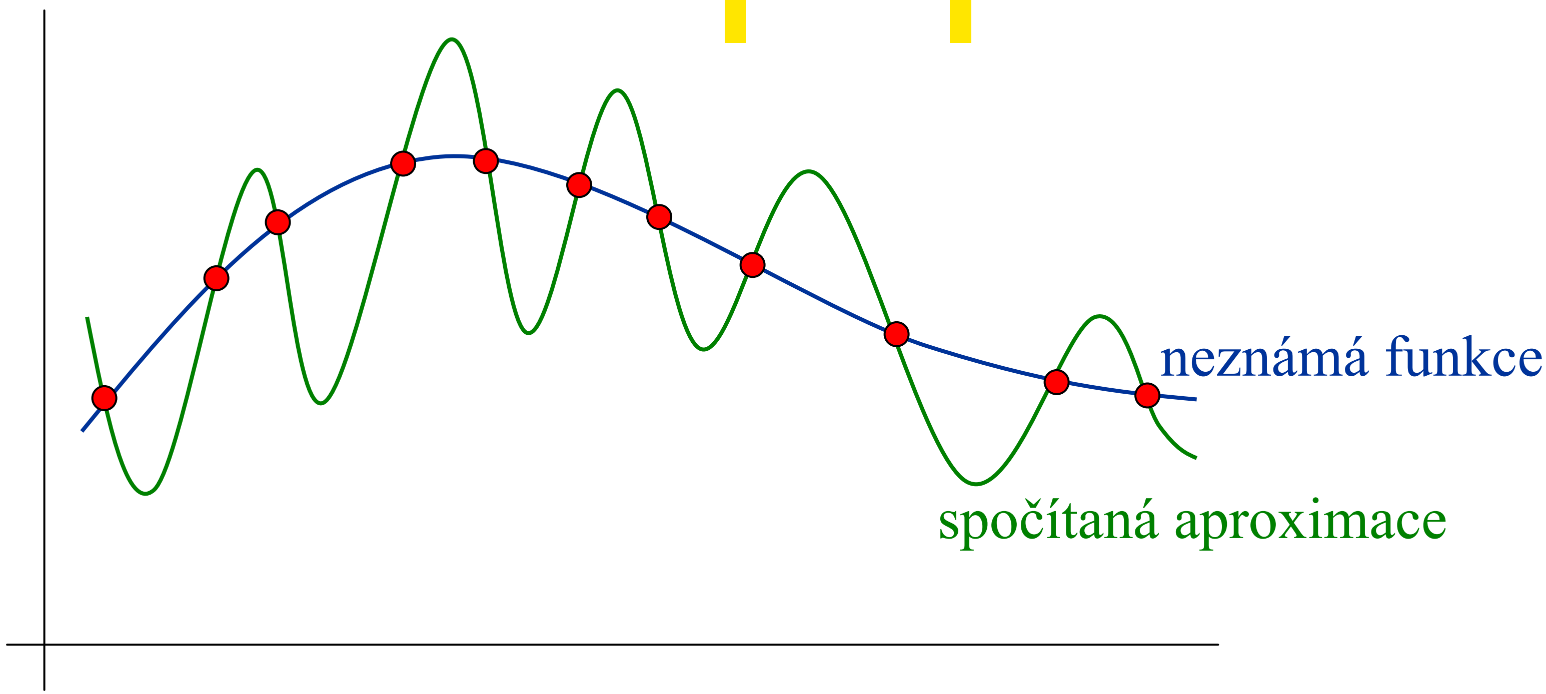
Vapnik – Chervonenkisova dimenze (VC–dimenze) (5)

- při volbě vhodného koeficientu α lze pro libovolný počet m zvolených bodů aproximovat libovolnou funkci omezenou v $< +1 / -1 >$ pomocí $\sin \alpha z$



Vapnik – Chervonenkisova dimenze (VC–dimenze) (6)

Problém „přeučení“ ~ síť se naučí i „šum“



Vapnik – Chervonenkisova dimenze (VC–dimenze) (7)

- Pro síť s počtem vah W a počtem neuronů N a s omezením pro generalizační chybu ε, je počet trénovacích vzorů P potřebných pro správné zobecňování: $P \geq (W/\varepsilon) \log_2 (N/\varepsilon)$
- Vrstevnatá síť s 1 skrytou vrstvou nemůže dobře zobecňovat, jestliže bylo méně než W/ε náhodně vybraných trénovacích vzorů, tj. $P \geq W/\varepsilon$
  - Pro požadovanou přesnost alespoň $90\ \%$ je třeba vybrat alespoň $10 \cdot W$ vzorů