



Odhad ceny PC sestav (na základě 8 atributů)

Filip Pekárek
Michal Penkala



Vstupní data

- Náhodný výběr 100 pc sestav z www.alza.cz
- Sledování 8 vstupních parametrů:

- Procesor (11 druhů)
- Grafická karta (14 druhů)
- Operační paměť (15 druhů)
- Hard disk (14 druhů)
- Optická mechanika (7 druhů)
- Chipset (8 druhů)
- Operační systém (8 druhů)
- Značka sestavy (8 druhů)

- Jeden cílový atribut - cena

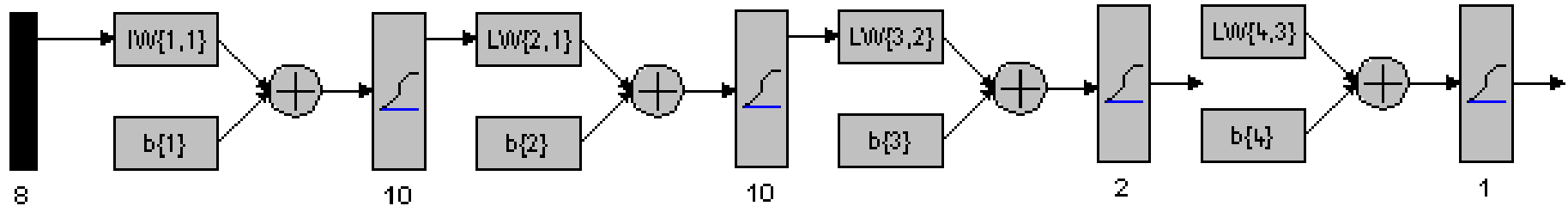
- Normalizace všech dat do intervalu (0,1)

```
ndata=[data(:,1)./16,data(:,2:4)./20,data(:,5)./10,data(:,6:8)./10,data(:,9)./40000];
```

- Data náhodně rozdělena v poměru
 - 70 % učení, 15 % validace, 15 % test

Použitá síť

- Topologie



```
newff([0 1;0 1;0 1;0 1;0 1;0 1;0 1;0 1],[10 10 2 1],{'logsig' 'logsig' 'logsig' 'logsig'},'trainlm');
```

- Přechodové funkce

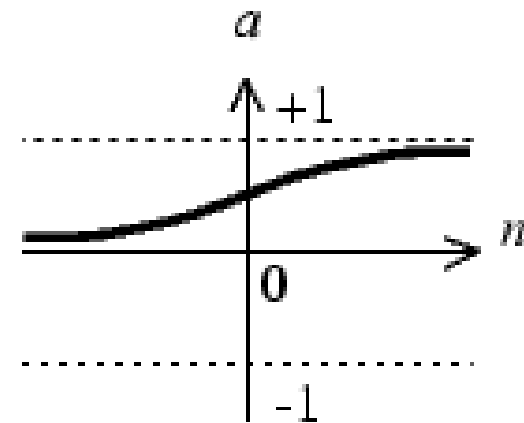
- Logsig ve všech vrstvách

- Učící funkce

- Trainlm ... Levenberg-Marquardt backpropagation

- Výkonostní funkce

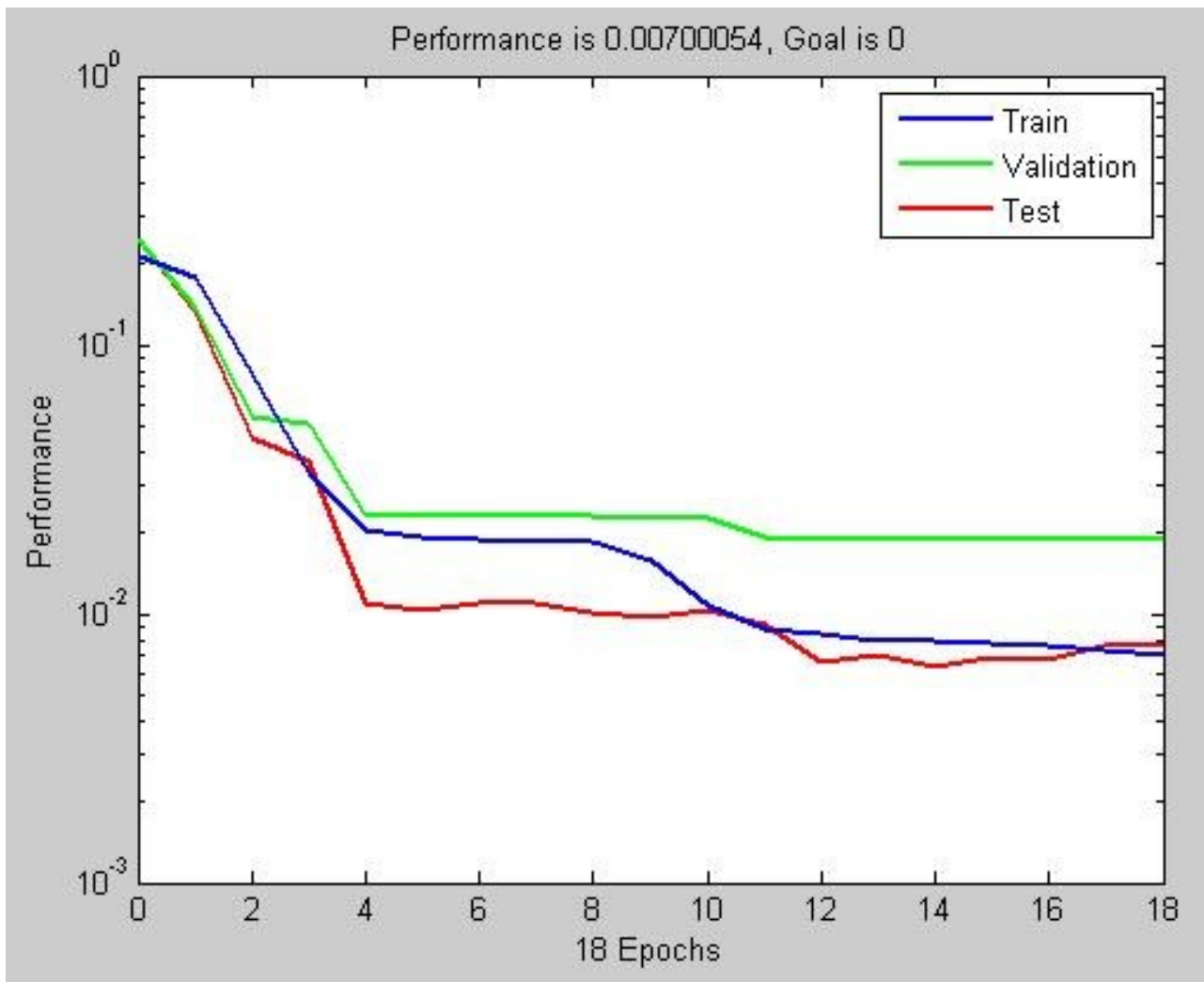
- MSE ... Mean squared error performance function



$$a = \text{logsig}(n)$$

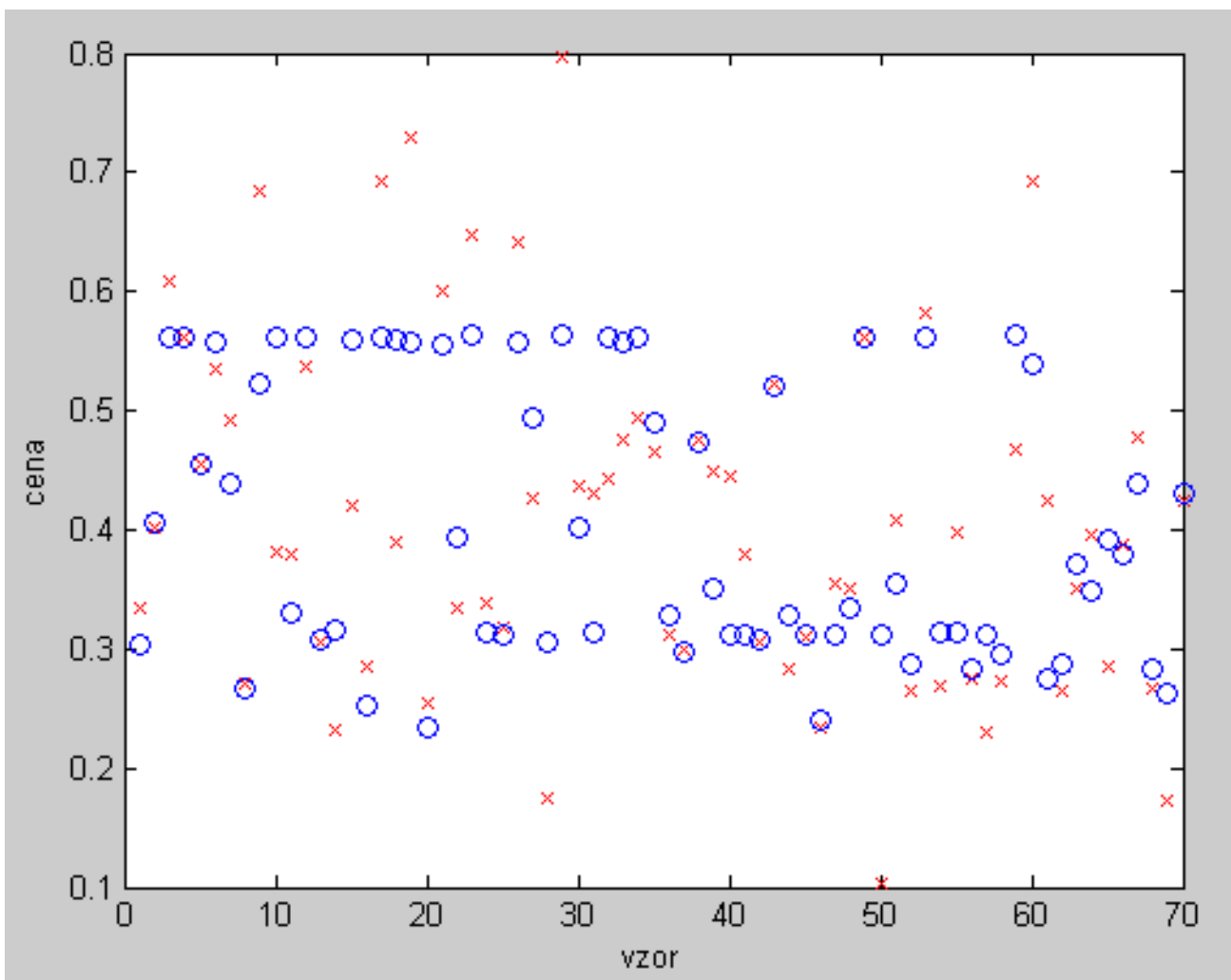
Výpočet

- Během hledání vhodné topologie se projevilo přeučení sítě
 - Použití validačních dat



Výsledky – trénovací množina

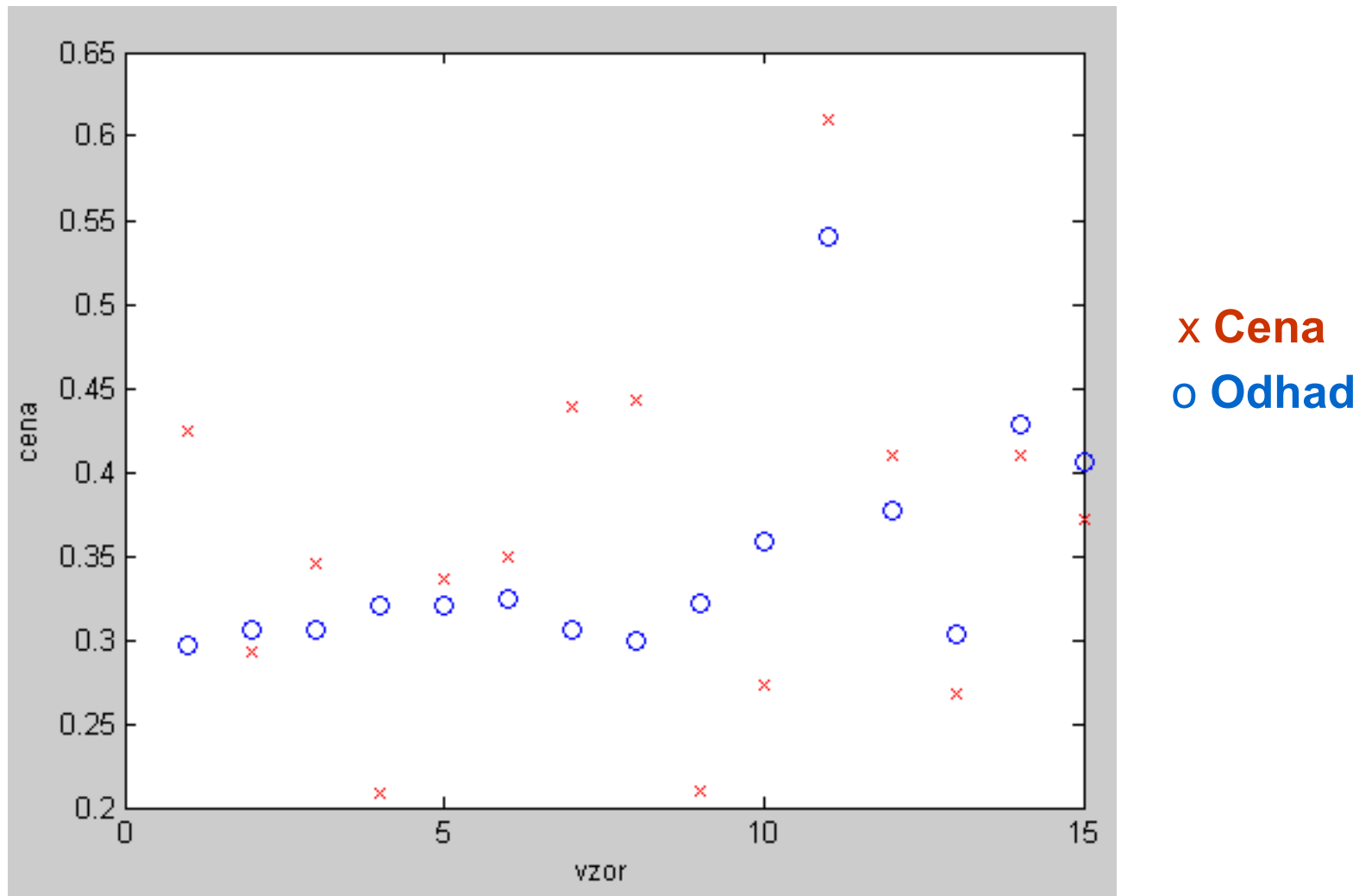
- Průměrná odchylka
 - 0,0607



x Cena
o Odhad

Výsledky – testovací množina

- Průměrná odchylka
 - 0,0669



Závěr

- Velká odchylka
 - Malý počet dat
 - Velká variabilita jednotlivých atributů
 - Náchylnost sítě k přeučení
- Zlepšení
 - Menší škály jednotlivých atributů
 - Větší množství dat
- Zkušenosti
 - Nejnáročnější bylo data dobýt
 - Krátkodobé použití sítě

