

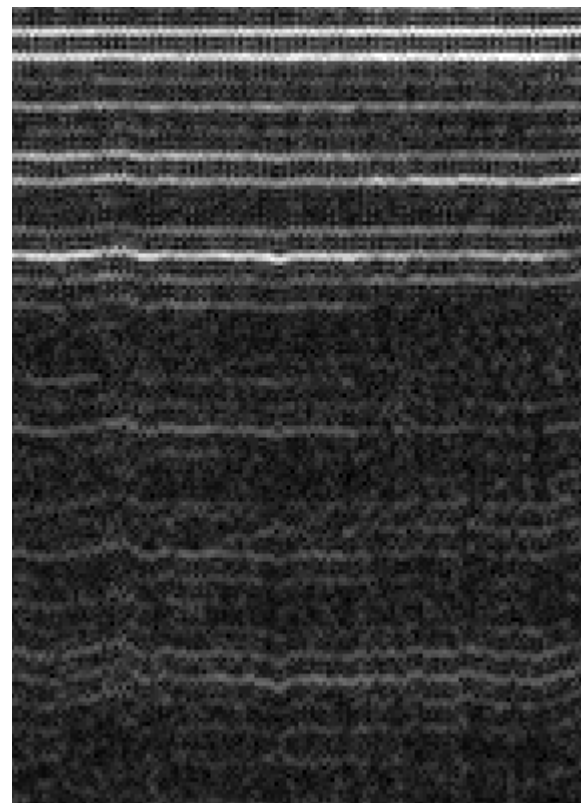
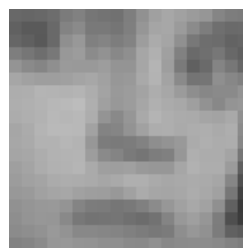
Samorganizace pro clusterování a rozpoznávání vzorů

Marek Kukačka
Res Informatica
Listopad 2009

Rozpoznávání vzorů

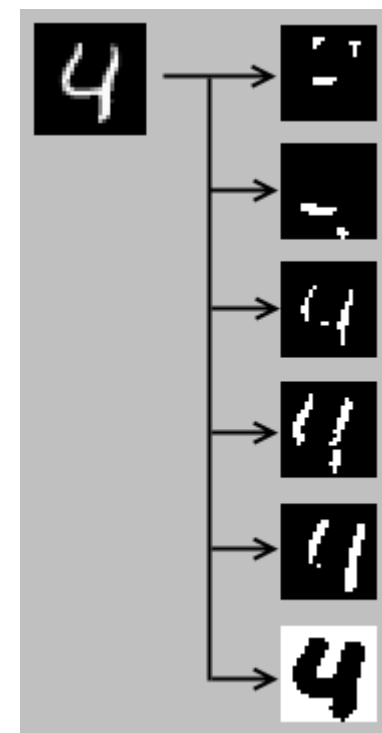
■ Problém – předzpracování dat pro rozpoznávání

- znaků
- obličejů
- zvuku
- a dalších ..



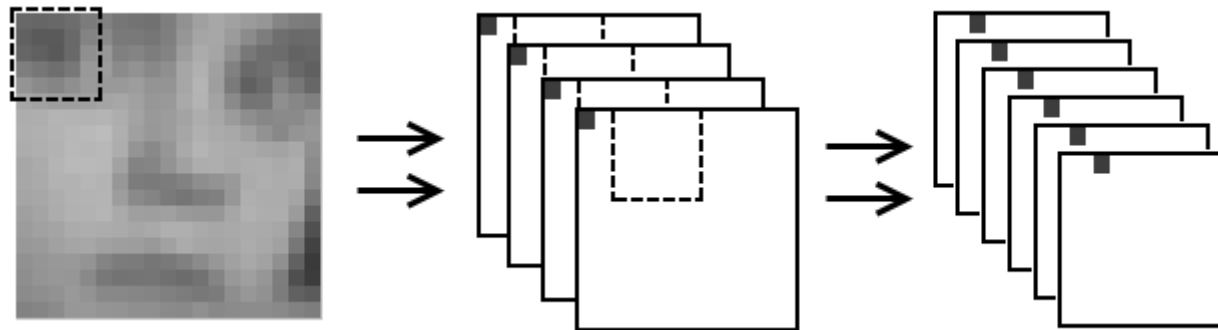
Využití lokálních příznaků

- Popis obrázku pomocí výskytu a relativní pozice lokálních příznaků



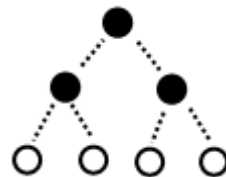
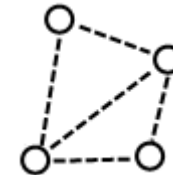
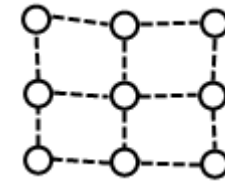
Automatické předzpracování

- Vícevrstvá samoorganizační síť
- Ve vyšších vrstvách jsou příznaky z nižších vrstev kombinovány do složitějších příznaků



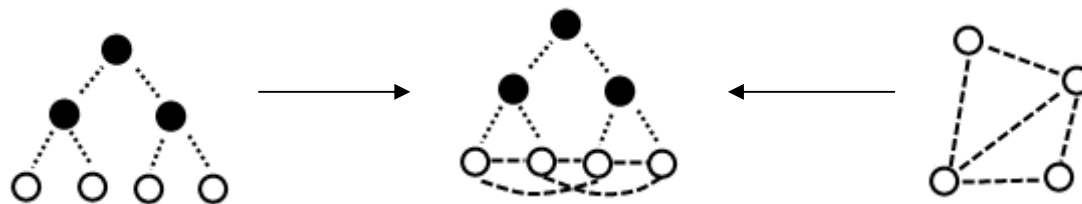
Samooorganizační sítě

- Nalezení příznaků v datech pomocí samoorganizačních sítí
 - Kohonen Maps
 - Growing Neural Gas
 - Evolving Tree



Hybridní samoorganizační síť

- Kombinuje modely Evolving Tree a Growing Neural Gas
- Výhody:
 - rychlejší adaptace a zpracování dat
 - menší chyba – lepší popis dat





Počet clusterů v datech

- Problém: kdy zastavit růst samoorganizační sítě? Jaký počet neuronů je nejvhodnější?
- Prořezávání sítě podle korelace výstupů neuronů
 - nefunguje moc dobře v prostorech s vysokou dimenzí



Informační kritéria

- např. Akaikeho kritérium (AIC),
Bayesovské kritérium (BIC)
- Používány pro porovnávání klasifikačních modelů

- Berou v potaz:
$$BIC = \frac{RSS}{\sigma_e^2} + k \cdot \ln(n)$$
 - chybu modelu
 - počet parametrů modelu



Kritérium pro on-line clusterování

- Minimalizace průměrné kvadratické vzdálenosti reprezentujícího neuronu od vzorku
 - průměrování přes počet vzorků reprezentovaných neuronem a přes počet neuronů v síti

$$NC = \frac{1}{|N|} \sum_{n \in N} \frac{1}{|C_n|} \sum_{x_n \in C_n} (x_n - w_n)^2$$



Další výzkum

- Lepší konvergenční kritérium pro rostoucí samoorganizační sítě
- Rozšíření modelu clusterování
- Testování na praktických úlohách
 - clusterování dat na základě detekovaných příznaků
 - rozpoznávání vzorů podle příznaků detekovaných automatickým předzpracováním



Děkuji za pozornost

mkukacka@email.cz