

Přednáška 4 – Model směsi kategorických komponent

Kategorické komponenty $f_j(y_t|\Theta)$

| | | | | |
|----------|------------|------------|-----|------------|
| y_t | 1 | 2 | ... | N |
| Θ | Θ_1 | Θ_2 | ... | Θ_N |

$j \in \{1, \dots, n_c\}$, n_c – počet komponent

Příklad:

Komponenta 1: $f_1(y_t|\Theta)$

| | | | | |
|----------|------|------|-----|------|
| y_t | 1 | 2 | 3 | 4 |
| Θ | 0.01 | 0.04 | 0.8 | 0.15 |

Komponenta 2: $f_2(y_t|\Theta)$

| | | | | |
|----------|-----|-----|------|------|
| y_t | 1 | 2 | 3 | 4 |
| Θ | 0.7 | 0.2 | 0.09 | 0.01 |

Komponenta 3: $f_3(y_t|\Theta)$

| | | | | |
|----------|------|------|------|-----|
| y_t | 1 | 2 | 3 | 4 |
| Θ | 0.01 | 0.04 | 0.15 | 0.8 |

Ukazovátka

ukazuje aktivní komponentu

$c_t \in \{1, 2, \dots, n_c\}$

- model ukazovátka – přepínání

| | | | | |
|-----------------|------------|------------|-----|----------------|
| c_t | 1 | 2 | ... | n_c |
| $f(c_t \alpha)$ | α_1 | α_2 | ... | α_{n_c} |

Ukazovátka: $c_t = 1$

Ukazovátka: $c_t = 2$

Ukazovátka: $c_t = 3$

Model ukazovátka:

| | | | |
|-----------------|-----|-----|-----|
| c_t | 1 | 2 | 3 |
| $f(c_t \alpha)$ | 0.3 | 0.4 | 0.3 |

Program – generování kategorických multimodálních dat

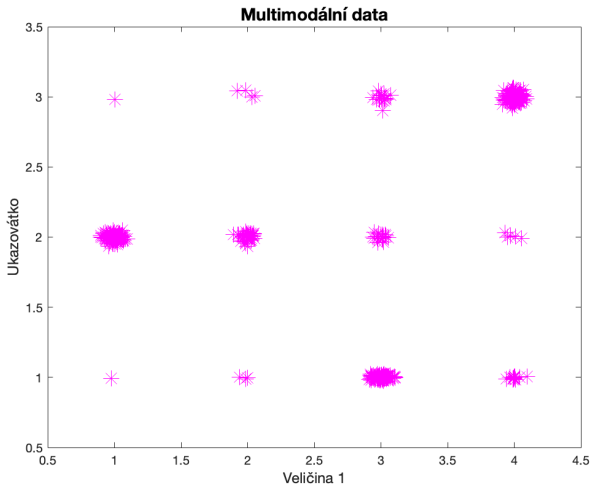
```
clear, clc, close
nd=500;      % počet dat

% parametry pro 3 kategorické komponenty a 4 hodnoty výstupu
th{1}=[0.01 0.04 0.8 0.15];
th{2}=[0.7 0.2 0.09 0.01];
th{3}=[0.01 0.04 0.15 0.8];

al=[.3 .4 0.3]; % parametry modelu ukazovátka

% Simulace
for t=1:nd
    c(t)=sum(rand(1,1)>cumsum(al))+1; % generování ukazovátka
    pp=cumsum(th{c(t)}); % kumulativní součet
    y(t)=sum(rand(1,1)>pp)+1; % generuje hodnoty 1,2,3,4
end
```

Výsledky - generování kategorických multimodálních dat



- Vzdálenost mezi shluky?
- Počet dat ve shlucích?

- Shlukování – odhad parametrů komponent Θ z průběžně měřených dat
(identifikujeme shluky)
- Klasifikace – třídíme data do odhadnutých komponent
(odhad ukazovátka c_t v čase t)

Základní princip – obecně:

- Inicializace komponent

V cyklu:

- Měříme data
- Vzdálenosti od jednotlivých komponent – proximity
- Normalizace proximit – váhy komponent
(pravděpodobnost, že je komponenta aktivní)
- Maximální váha – bodový odhad ukazovátka
(klasifikace)
- Update statistik s váženými daty
- Přepočítání bodových odhadů podle typu komponent
- Jdeme na krok 1

Pro čas $t = 0$

- 1 Nastavíme počet komponent + jejich počáteční statistiky
- 2 Vypočteme počáteční bodové odhady parametrů $(\hat{\Theta}_0)_j$

Pro čas $t = 1, 2, \dots$, pro každou komponentu

- 1 Měříme nová data y_t
- 2 Určíme proximity m_j – dosadíme do komponent $(\hat{\Theta}_{t-1})_j$ a y_t
- 3 Určíme váhy komponent

$$w_{j;t} = \frac{m_j}{\sum_{i=1}^{n_c} m_i}$$

- 4 Bodový odhad ukazovátka: $\hat{c}_t = \arg \max_j w_t$
- 5 Update statistik s váženými daty:

$$(\nu_{i;t})_j = (\nu_{i;t-1})_j + w_{j;t} \delta(i; y_t)$$

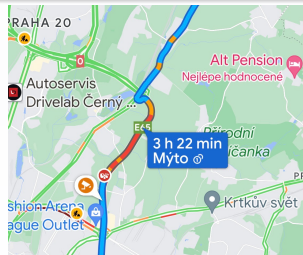
- 6 Přepočítání bodových odhadů $(\hat{\Theta}_{i;t})_j = \frac{(\nu_{i;t})_j}{\sum_{k=1}^n (\nu_{k;t})_j}$
- 7 Jdeme na krok 1

Příklad: klasifikace dopravní špičky online

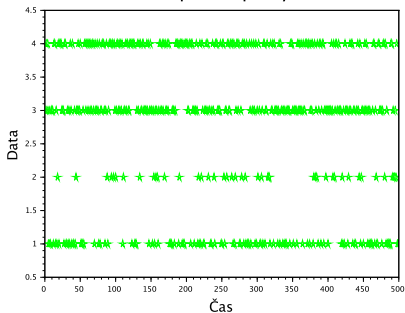
- $y_t \in \{1, 2, 3, 4\}$ – stupeň dopravy (volný provoz až dopravní kolaps)
- Čas t – 5 minut
- $c_t \in \{1, 2, 3\}$ – dopravní špička (ráno, poledne, odpoledne)

Inicializace:

- počet komponent – expertně
- ~~$x - y$ graf~~
- dimenze počáteční statistiky – podle maxima
- počáteční statistika – apriorní nebo expertní znalosti



Stupeň dopravy



Program – inicializace

```
% Inicializace - expertně
nc=3; % Počet komponent

% Počáteční statistika - expertně - apriorní znalosti
ni{1} = [1 1 15 1]; % dimenze podle maxima
ni{2} = [15 10 1 1];
ni{3} = [1 1 1 15];

% Počáteční odhad
for i = 1:nc
    theta_odhad{i} = ni{i} ./ sum(ni{i}, 2);
end
```

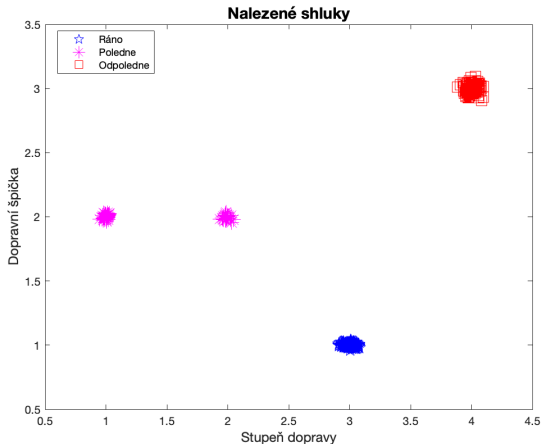
Program – shlukování a klasifikace

```
% Cyklus shlukování a klasifikace
for t = 1:nd
    m = zeros(nc,1);
    for i = 1:nc
        m(i) = theta_odhad{i}(y(t)); % Proximity
    end
    w = m / sum(m); % váhy

    for i = 1:nc
        ni{i}(y(t)) = ni{i}(y(t)) + w(i); % Update statistik

        % Bodový odhad parametrů
        theta_odhad{i} = ni{i} ./ sum(ni{i}, 2);
    end

    % Bodový odhad ukazovátka – klasifikace
    [~, cp(t)] = max(w);
```

Poznámky:

- online = ~~trénovací a testovací data~~
- přesnost klasifikace ?

Program – validace klasifikace z predikce z komponent

```
% Validace – predikce z komponent
[~,yp(t)] = max(theta_odhad{cp(t)}); % hodnoty predikce
end

% Výsledky
PE = sum(y == yp);
Acc = (PE * 100) / length(cp) % Přesnost predikce
```

Výsledky – predikce z komponent

