



Osztályozás – USPS

<http://www.cs.ubbcluj.ro/~csatol/mestint/USPS>

Mesterséges
Intelligencia

8

Csató Lehel

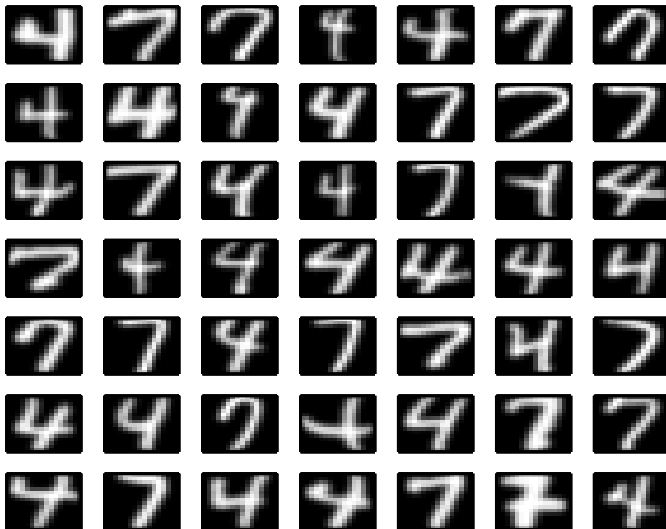
Grafikus modellek

GM feladatok

Írányítatlan

GM-ek

Matlab feladat





Mesterséges
Intelligencia

8

Csató Lehel

Grafikus modellek
GM feladatok

Írányítatlan
GM-ek

Matlab feladat

USPS adatok

- United States Postal Service;
- kézzel írott számjegyek 16×16 -os bit-térképe;
- ≈ 7200 tanulási- és ≈ 2000 teszt-adat;

Feladat:

- **Bináris osztályozás:** válasszuk el a négyeseket (4) a hetesektől (7).
- Írjunk **egy** osztályozási algoritmust a következők közül: Döntési fa; Neurális háló; SVM; Bayes osztályozó.
- Értékeljük az algoritmust a következők szerint:
 - osztályozási hiba tanulási adatokon;
 - **osztályozási hiba teszt-adatokon – 10% alatt!**;
 - tanulási idő;
 - vizsgáljuk meg a rosszul osztályozott „pontokat”;



Osztályozás

III

<http://www.cs.ubbcluj.ro/~csatol/mestint/USPS>

Mesterséges
Intelligencia

8

Csátó Lehel

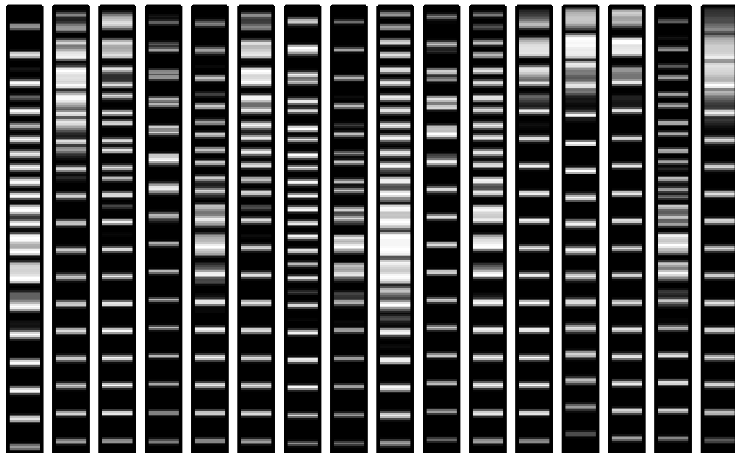
Grafikus modellek

GM feladatok

Írányítatlan

GM-ek

Matlab feladat



4 7 7 4 4 7 7 4 4 4 4 7 7 7 4 7
-1 +1 +1 -1 -1 +1 +1 -1 -1 -1 -1 +1 +1 +1 -1 +1

Mesterséges
Intelligencia

8

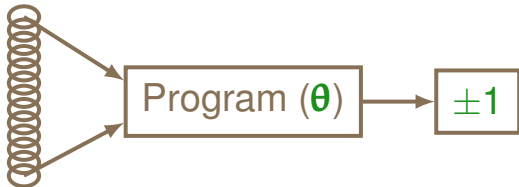
Csató Lehel

Grafikus modellek
GM feladatok

Írányítatlan
GM-ek

Matlab feladat

Osztályozó algoritmus:



Tanulás:

- θ_0 kezdőértékek
- tanulási adatok

$$\mathcal{D} = \{(\mathbf{x}_1, y_1), \dots, (\mathbf{x}_N, y_N)\}$$

- Becslés: $\theta_1 \rightarrow \dots \rightarrow$
- θ_{post}

Közben: tanulási hiba mérése.

Tesztelés:

- θ_{post} használata;
- új adatokra:

$$\mathcal{D}_{\text{test}} = \dots$$

- hiba mérése
tesztadatokon.

Mesterséges
Intelligencia

8

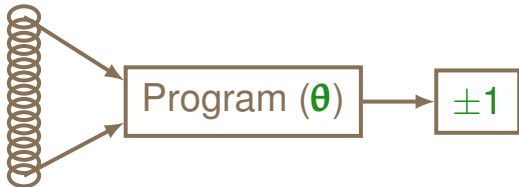
Csató Lehel

Grafikus modellek
GM feladatok

Irányítatlan
GM-ek

Matlab feladat

Osztályozó algoritmus:



Tanulás:

- θ_0 kezdőértékek
- tanulási adatok

$$\mathcal{D} = \{(\mathbf{x}_1, y_1), \dots, (\mathbf{x}_N, y_N)\}$$

- Becslés: $\theta_1 \rightarrow \dots \rightarrow$
- θ_{post}

Közben: tanulási hiba mérése.

Tesztelés:

- θ_{post} használata;
- új adatokra:

$$\mathcal{D}_{\text{test}} = \dots$$

- hiba mérése
tesztadatokon.



Mesterséges
Intelligencia

8

Csató Lehel

Grafikus modellek

GM feladatok

Irányítatlan

GM-ek

Matlab feladat

Program + dokumentáció:

(dolg...)

- Algoritmus kiválasztása;
- Tanulás – paraméterek becslése;
 - tanulási adatok beolvasása;
 - adatok transzformációja – pre–processzálas;
 - algoritmus paramétereinek inicializálása;
 - paraméterek tanulása.
- Tesztelés;
 - **teszt-adatok beolvasása** – függetlenek a tanulási mintáktól!;
 - algoritmus paramétereinek **beolvasása** – tanulási eredmény;
 - hiba mérése.
- Dokumentálás: algoritmus, módszer, **statisztika**



Mesterséges
Intelligencia

8

Csató Lehel

Grafikus modellek

GM feladatok

Irányítatlan
GM-ek

Matlab feladat

```
clear all;
% data generation
nTr = 200; nTe = 400;
[dX dY] = cl_data(nTr, [], 0);
ii_1 = find(dY==1);
d1 = dX(ii_1,:);
d2 = dX(setdiff([1:nTr],ii_1),:);
N_1 = size(d1,1);
N_2 = size(d2,1);

pi1 = N_1/nTr; pi2 = N_2/nTr;

mu1 = sum(d1,1)/N_1; % class means
mu2 = sum(d2,1)/N_2;

d = d1 - repmat(mu1,[N_1 1]);
sig1 = d'*d/N_1; sig1inv = sig1^(-1);

d = d2 - repmat(mu2,[N_2 1]);
sig2 = d'*d/N_2; sig2inv = sig2^(-1);

% generating nTe test data
[dX, dY] = cl_data(nTe, [], 0);
ii_1 = find(dY==1);
data1 = dX(ii_1,:);
data2 = dX(setdiff([1:nTe],ii_1),:);
N_1 = size(data1,1);
N_2 = size(data2,1);
```

```
% performing classification
log_sig1 = log(det(sig1))-log(pi1);
log_sig2 = log(det(sig2))-log(pi2);

% classification for data1
p_1=data1-repmat(mu1,[length(data1) 1]);
p_1=sum(((p_1*sig1inv).*p_1),2);

p_2=data1-repmat(mu2,[length(data1) 1]);
p_2=sum(((p_2*sig2inv).*p_2),2);

class(1,1) = length( ...
    find(-p_1+p_2 > log_sig1-log_sig2) );
class(1,2) = N_1 - class(1,1);

% SIMILARLY for the second dataset
p_1=data2-repmat(mu2,[length(data2) 1]);
p_1=sum(((p_1*sig2inv).*p_1),2);

p_2=data2-repmat(mu1,[length(data2) 1]);
p_2=sum(((p_2*sig1inv).*p_2),2);

class(2,2) = length( ...
    find( -p_1+p_2 > log_sig2-log_sig1));
class(2,1) = N_2 - class(2,2);

% the confusion matrix:
class
```

...grafikus ábrázolás...



Mesterséges
Intelligencia

8

Csató Lehel

Grafikus modellek

GM feladatok

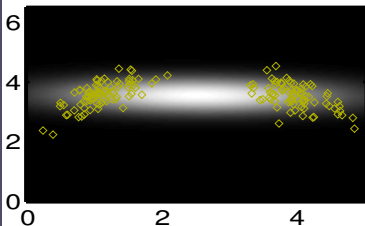
Irányítatlan

GM-ek

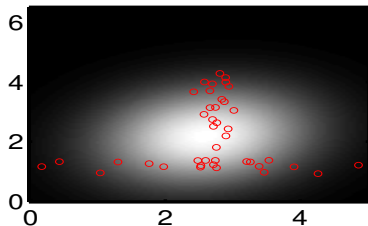
Matlab feladat

Eredmény:

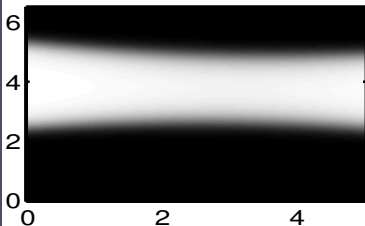
Class-cond. prob. C_1



Class-cond. prob. C_2



Posterior probability



Classification boundary

