Az együttes tanulást végző regressziós eljárás (COREG) pszeudókódja

A két szakértő közti különbséget a szakértőben használt távolság *p* paraméterével valósították meg. Az ún. Minkowski-távolság *N* elemű vektorok esetén: $\mathbf{x}^{(\ell)} = \left[x_1^{(\ell)}, x_2^{(\ell)}, ..., x_N^{(\ell)} \right]$; $M(\mathbf{x}^{(r)}, \mathbf{x}^{(s)}) = \sqrt[p]{\sum_{i=1}^{N} \left| x_i^{(r)} - x_i^{(s)} \right|^p}$

function COREG(L,U,k,T,p_1,p_2) **returns** L_1, L_2

input L: címkézett minták halmaza
U: címkézetlen minták halmaza
k: a k-NN regresszorban figyelembe vett szomszédok száma
T: az iterációk záma
p_1, p_2: a két k-NN regresszorban használt Minkowski-távolság paraméter

$L_1 \leftarrow L$

$L_2 \leftarrow L$

$U^{'} \leftarrow U$ véletlenszerűen kiválasztott elemeiből képzett részhalmaz //* számítási igény csökkentése a cél!

for $t = 1:T$ **do**

  for $\mathrm{j} = 1:2$ **do**

    for each $\mathbf{x}^{(u)} \in \mathrm{U}^{'}$ **do**

      $\hat{y}^{(u)} = kNN(L_j, k, p_j, \mathbf{x}^{(u)})$ //* a j-dik szakértő becslése a címkézetlen mintára

      $\Omega^{(u)} = Neighbours(L_j, k, \mathbf{x}^{(u)})$ //* a címkézetlen minta környezete – közelítés!

      $L_j^{(u)} = L_j \cup \left\{ \left(\mathbf{x}^{(u)}, \hat{y}^{(u)} \right) \right\}$ //* ideiglenes címkézett mintahalmaz

      $\Delta_{\mathbf{x}^{(u)}} = \sum_{\mathbf{x}^{(\ell)} \in \Omega^{(u)}} \left[\left(y^{(\ell)} - kNN\left(L_j, k, p_j, \mathbf{x}^{(\ell)} \right) \right)^2 - \left(y^{(\ell)} - kNN\left(L_j^{(u)}, k, p_j, \mathbf{x}^{(\ell)} \right) \right)^2 \right]$

    endfor

    if any $\Delta_{\mathbf{x}^{(u)}} > 0$

      $\tilde{\mathbf{x}}^{(r)} = \arg\max_{\mathbf{x}^{(u)} \in U^{'}} \Delta_{\mathbf{x}^{(u)}}$ //* amelyik nem címkézett pont a legnagyobb javulást hozza

      $\hat{y}^{(r)} = kNN(L_j, k, p_j, \tilde{\mathbf{x}}^{(r)})$

      $\pi_j \leftarrow \left\{ \left(\hat{\mathbf{x}}^{(r)}, \hat{y}^{(r)} \right) \right\}$

      $U \leftarrow U - \left\{ \left(\hat{\mathbf{x}}^{(r)}, \hat{y}^{(r)} \right) \right\}$ //* kivesszük a címkézetlen pontok halmazából

    else $\pi_j \leftarrow \{ \}$

    endif

  endfor

  $L_1 \leftarrow L_1 \cup \pi_2$

  $L_2 \leftarrow L_2 \cup \pi_1$

  if neither L_1, L_2 changes

    return L_1, L_2

  endif

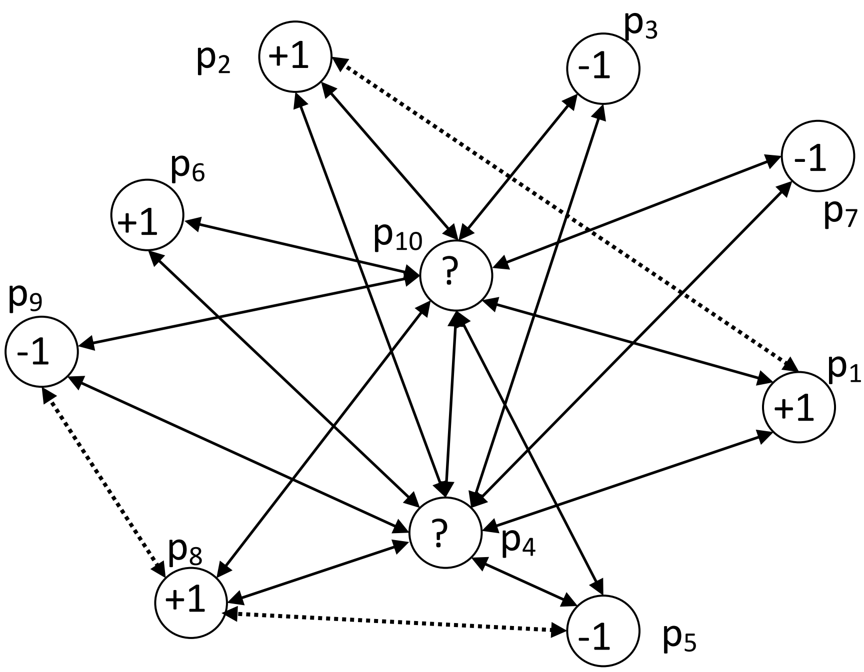
  $U^{'} \leftarrow U$ véletlenszerűen kiválasztott elemeiből képzett új részhalmaz

endfor

return L_1, L_2

endfunc

<u>MINCUT eljárás – mintapélda</u> $L=\{p_1,p_2,p_3,p_5,p_6,p_7,p_8,p_9\}$; $\mathrm{U}=\{p_4,p_{10}\}$



A szaggatott vonalakkal jelölt kapcsolatok (amelyek hasonlóságot jelölnek az elemek közt, tehát a nagyon hasonlókat egy csoportban szeretnénk tudni) nem függnek a U-beli két pont besorolásától. Természetesen jóval több ilyen van, mint ahányat az ábrán feltüntettünk. A folytonos vonallal jelölt kapcsolatok függnek a két ismeretlen besorolású pont besorolásától.

p_x	osztály	kapcsolat erőssége p_4-el	p_x	osztály	kapcsolat erőssége p_{10}-el
p_1	1	8,7	p_1	1	5,6
p_2	1	6,5	p_2	1	3,9
p_3	-1	11,3	p_3	-1	4,7
p_4	?	-	p_4	?	**8,3**
p_5	-1	1	p_5	-1	2,7
p_6	1	4,8	p_6	1	7,3
p_7	-1	7,2	p_7	-1	4,2
p_8	1	5,5	p_8	1	3,4
p_9	-1	6,3	p_9	-1	2,8
p_{10}	?	**8,3**	p_{10}	?	-

p_4 osztály	p_{10} osztály		veszteség-függvény	optimális besorolás
-1	-1	(8,7+6,5+4,8+5,5)+**0**+(5,6+3,9+7,3+3,4) = 25,5+0+20,2	45,5	
1	-1	(11,3+1+7,2+6,3)+**8,3**+(5,6+3,9+7,3+3,4) = 25,8+8,3+20,2	54,3	
-1	1	(8,7+6,5+4,8+5,5)+**8,3**+(4,7+2,7+4,2+2,8) = 25,5+8,3+14,4	48.2	
1	1	(11,3+1+7,2+6,3)+**0**+(4,7+2,7+4,2+2,8) = 25,8+0+14,4	40,2	!!!!!!

Természetesen, ha a kapcsolatok erőssége nem hasonlóságot (”vonzalmat”), hanem „összeférhetetlenséget” jelöl, akkor a legnagyobb veszteségfüggvény érték a legjobb nekünk.