

# Klasyfikacja wielokategorialna za pomocą sieci neuronowych

Rafał Grodzicki

---

# Klasyfikacja wielokategorialna (multilabel classification)

- Zbiór danych wejściowych:

$$X = R^d$$

- Zbiór klas (kategorii):

$$Y = \{0, 1, \dots, Q-1\}$$

- Zbiór uczący:

$$T = \{(x_1, Y_1), (x_2, Y_2), \dots, (x_m, Y_m)\}, \quad x_i \in X, \quad Y_i \subset Y, \quad Y_i \neq \emptyset$$

- Wielokategorialny system uczący się  
(multilabel learning system)

- Wygenerowanie wielokategorialnego klasyfikatora:

$$h: X \rightarrow 2^Y$$

# Klasyfikacja wielokategorialna (multilabel classification)

- Zamiast klasyfikatora ( $h$ ) system tworzy funkcję:

$$f : X \times Y \rightarrow R$$

- Dla pary uczącej  $(x_i, Y_i)$ ,  $x_i \in X$ ,  $Y_i \subset Y$ ,  $Y_i \neq \emptyset$  system dąży do wygenerowania funkcji spełniającej warunek  $(\forall y_1 \in Y_i \wedge y_2 \notin Y_i) f(x_i, y_1) > f(x_i, y_2)$
- Na podstawie utworzonej funkcji można wygenerować klasyfikator:

$$(\forall x \in X) \quad h(x) = \{y \in Y : f(x, y) > t(x)\}$$

gdzie  $t : X \rightarrow R$  jest funkcją progową

# Problem wielokategorialnej klasyfikacji

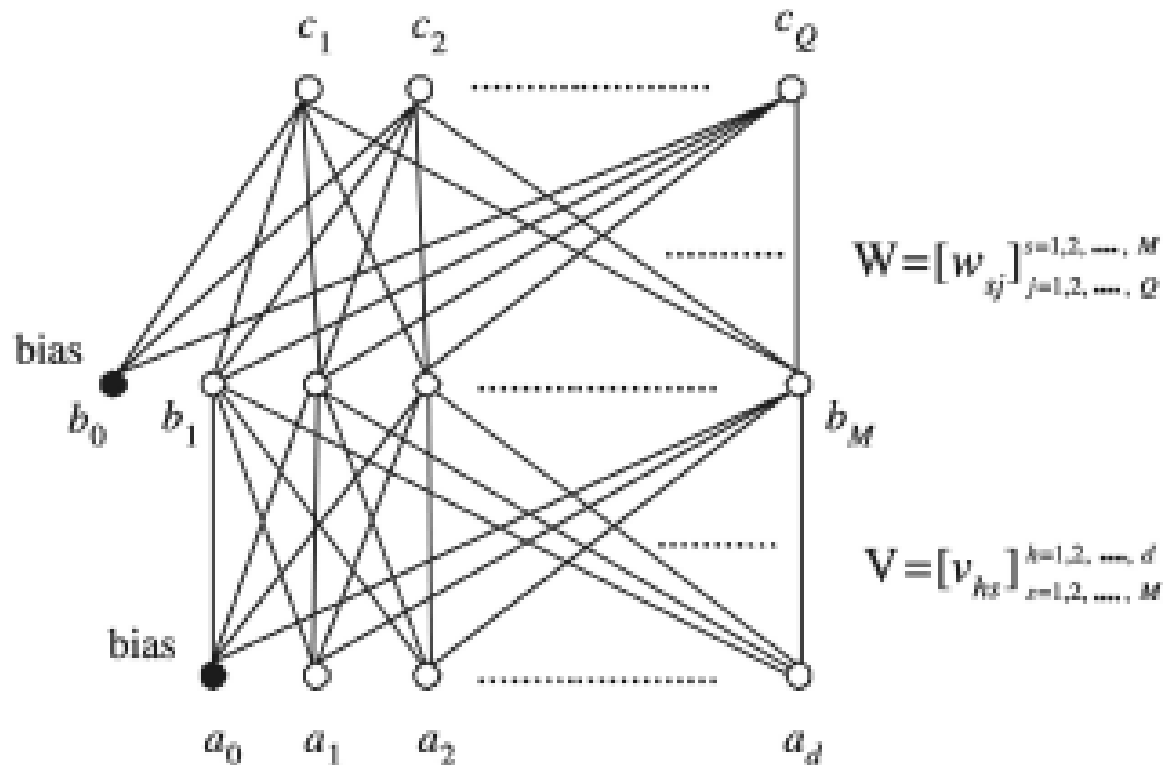
- Dekompozycja na niezależne problemy klasyfikacji binarnej
  - Nie uwzględnia korelacji pomiędzy różnymi klasami
- Dobry wielokategorialny system uczący się
  - Uwzględnianie korelacji pomiędzy różnymi klasami (kategoriami)
  - Zachowanie małej liczby klas

# Sieć neuronowa jako wielokategoryjny klasyfikator

- BP-MLL (Backpropagation for Multilabel Learning)
- Autorzy:
  - Min-Ling Zhang
  - Zhi-Hua Zhou
- Pierwszy wielokategoryjny system uczący się oparty na sieciach neuronowych
- Perceptron ze zmodyfikowaną funkcją błędów
- Uczenie – wsteczna propagacja błędów

# Sieć neuronowa jako wielokategoryjny klasyfikator

- Architektura



# Sieć neuronowa jako wielokategoryalny klasyfikator

- Funkcja błędu

- Klasyczna – błąd średniokwadratowy

$$E = \sum_{i=1}^m \sum_{j=1}^Q (c_j^i - d_j^i)^2, \quad c_j^i = c_j(x_i), \quad d_j^i = \begin{cases} +1, & j \in Y_i \\ -1, & j \notin Y_i \end{cases}$$

- Uwzględnia poszczególne kategorie niezależnie
- Nie uwzględnia korelacji pomiędzy kategoriami (klasami)
- Modyfikacja: na wyjściu sieci powinny być większe wartości dla kategorii należących do  $Y_i$  niż dla kategorii spoza  $Y_i$

# Sieć neuronowa jako wielokategoryalny klasyfikator

- Funkcja błędu

- Zmodyfikowana

$$E_1 = \sum_{i=1}^m \frac{\sum_{(k,l) \in Y_i \times \bar{Y}_i} e^{-(c_k^i - c_l^i)}}{|Y_i| |\bar{Y}_i|}, \quad c_j(x_i) = c_j^i$$

- Koncentracja na różnicy pomiędzy wartościami wyjściowymi dla kategorii należących do  $Y_i$  a wartościami wyjściowymi dla kategorii spoza  $Y_i$
  - Silne karanie w przypadku wartości wyjściowych dla kategorii spoza  $Y_i$  większych niż dla kategorii z  $Y_i$
  - Uwzględnianie zależności pomiędzy różnymi klasami – większe wartości na wyjściu sieci dla kategorii należących do  $Y_i$  niż dla kategorii spoza  $Y_i$



# Sieć neuronowa jako wielokategoryalny klasyfikator

- Klasyfikacja

- Na podstawie wartości wyjściowych sieci ustalany jest zbiór kategorii (klas) odpowiadający danym wejściowym:

$$\{j \in Y : c_j > t(x)\}$$

gdzie  $t : X \rightarrow R$  jest funkcją progową

- Funkcja progowa:

- Stała funkcja ( $t(x) = 0$ )
- Wyznaczana na podstawie zbioru uczącego

# Modyfikacje funkcji błędu

- Włączenie wartości progowej do funkcji błędu

$$E_2 = \sum_{i=1}^m \frac{\sum_{(k,l) \in Y_i \times \bar{Y}_i} e^{-(c_k^i - c_l^i)} + \sum_{k \in Y_i} e^{-(c_k^i - c_Q^i)} + \sum_{l \in \bar{Y}_i} e^{-(c_Q^i - c_l^i)}}{|Y_i| |\bar{Y}_i| + |Y_i| + |\bar{Y}_i|}$$

$$h(x_i) = \{j \in Y : c_j(x_i) > c_Q(x_i)\}, \quad c_j(x_i) = c_j^i$$

- Wartość  $c_Q(x_i)$  jest interpretowana jako próg
- Funkcja progowa jest wyznaczana poprzez adaptację w procesie uczenia sieci neuronowej
- Koncentracja na różnicach pomiędzy wartościami wyjściowymi dla poszczególnych kategorii a wartością progową

## Modyfikacje funkcji błędu

- Włączenie do funkcji błędu wartości progowych niezależnych dla poszczególnych kategorii

$$E_3 = \sum_{i=1}^m \frac{\sum_{(k,l) \in Y_i \times \bar{Y}_i} e^{-(c_{2k}^i - c_{2l}^i)} + \sum_{k \in Y_i} e^{-(c_{2k}^i - c_{2k+1}^i)} + \sum_{l \in \bar{Y}_i} e^{-(c_{2l+1}^i - c_{2l}^i)}}{|Y_i| |\bar{Y}_i| + |Y_i| + |\bar{Y}_i|}$$

$$h(x_i) = \{j \in Y : c_{2j}(x_i) > c_{2j+1}(x_i)\}, \quad c_j(x_i) = c_j^i$$

- Wartość  $c_{2j}(x_i)$  jest interpretowana jako wartość wyjściowa dla j-tej kategorii
- Wartość  $c_{2j+1}(x_i)$  jest interpretowana jako próg dla j-tej kategorii
- Koncentracja na różnicach pomiędzy wartościami wyjściowymi dla poszczególnych kategorii a odpowiadającymi im wartościami progowymi

## Modyfikacje funkcji błędu

- Włączenie do funkcji błędu wartości progowych niezależnych dla poszczególnych kategorii

$$E_4 = \sum_{i=1}^m \frac{\sum_{(k,l) \in Y_i \times \bar{Y}_i} e^{-(c_{2k}^i - c_{2l}^i)} + \sum_{k \in Y_i} \sum_{t \in Y_i} e^{-(c_{2k}^i - c_{2t+1}^i)} + \sum_{l \in \bar{Y}_i} \sum_{t \in \bar{Y}_i} e^{-(c_{2t+1}^i - c_{2l}^i)}}{|Y_i| |\bar{Y}_i| + |Y_i|^2 + |\bar{Y}_i|^2}$$

$$h(x_i) = \{j \in Y : c_{2j}(x_i) > c_{2j+1}(x_i)\}, \quad c_j(x_i) = c_j^i$$

- Koncentracja na różnicach pomiędzy wartościami wyjściowymi dla poszczególnych kategorii a **wszystkimi** wartościami progowymi odpowiadającymi kategoriom **tego samego typu**
- Wartość  $c_{2j}(x_i)$  jest interpretowana jako wartość wyjściowa dla j-tej kategorii
- Wartość  $c_{2j+1}(x_i)$  jest interpretowana jako próg dla j-tej kategorii

## Modyfikacje funkcji błędu

- Włączenie do funkcji błędu wartości progowych niezależnych dla poszczególnych kategorii

$$E_5 = \sum_{i=1}^m \frac{\sum_{(k,l) \in Y_i \times \bar{Y}_i} \left( e^{-(c_{2k}^i - c_{2l}^i)} + e^{-(c_{2l+1}^i - c_{2k+1}^i)} \right) + \sum_{k \in Y_i} \sum_{t \in Y_i} e^{-(c_{2k}^i - c_{2t+1}^i)} + \sum_{l \in \bar{Y}_i} \sum_{t \in \bar{Y}_i} e^{-(c_{2t+1}^i - c_{2l}^i)}}{2|Y_i| |\bar{Y}_i| + |Y_i|^2 + |\bar{Y}_i|^2}$$

$$h(x_i) = \{j \in Y : c_{2j}(x_i) > c_{2j+1}(x_i)\}, \quad c_j(x_i) = c_j^i$$

- Uwzględnianie różnic pomiędzy wartościami progowymi odpowiadającymi kategoriom **różnego typu**
- Mniejsze wartości progowe dla kategorii należących do  $Y_i$  niż dla kategorii spoza  $Y_i$
- Wartość  $c_{2j}(x_i)$  jest interpretowana jako wartość wyjściowa dla j-tej kategorii
- Wartość  $c_{2j+1}(x_i)$  jest interpretowana jako próg dla j-tej kategorii

# Miary oceny jakości klasyfikacji wielokategorialnej

- Hamming loss

- Określa jak często występuje błędna klasyfikacja (obiekt zaklasyfikowany jako należący do danej kategorii podczas gdy do niej nie należy)
- Im mniejsza wartość tym lepiej

$$hloss_s(h) = \frac{1}{p} \sum_{i=1}^p \frac{1}{Q} |h(x_i) \Delta Y_i|,$$

$$h(x_i) \Delta Y_i = (h(x_i) \cup Y_i) \setminus (h(x_i) \cap Y_i)$$

# Miary oceny jakości klasyfikacji wielokategorialnej

- One-error
  - Określa jak często kategoria o najwyższej wartości wyjściowej nie należy do zbioru  $Y_i$
  - Im mniejsza wartość tym lepiej

$$oneerror_s(f) = \frac{1}{p} \sum_{i=1}^p \left( \left[ \left( \arg \max_{y \in Y} f(x_i, y) \right) \notin Y_i \right] ? 1 : 0 \right)$$

# Miary oceny jakości klasyfikacji wielokategorialnej

- Ranking loss

- Określa uśrednioną część par kategorii  $(y_1, y_2)$ ,  $(y_1 \in Y_i \wedge y_2 \notin Y_i \wedge f(x_i, y_1) \leq f(x_i, y_2))$
- Wskazuje jak często kategoria należąca do zbioru  $Y_i$  ma nie większą wartość wyjściową niż kategoria spoza tego zbioru
- Im mniejsza wartość tym lepiej

$$rloss_S(f) = \frac{1}{p} \sum_{i=1}^p \frac{1}{|Y_i| |\bar{Y}_i|} \left| \left\{ (y_1, y_2) \in Y_i \times \bar{Y}_i : f(x_i, y_1) \leq f(x_i, y_2) \right\} \right|$$



# Klasyfikacja wielokategorialna – zastosowanie w bioinformatyce

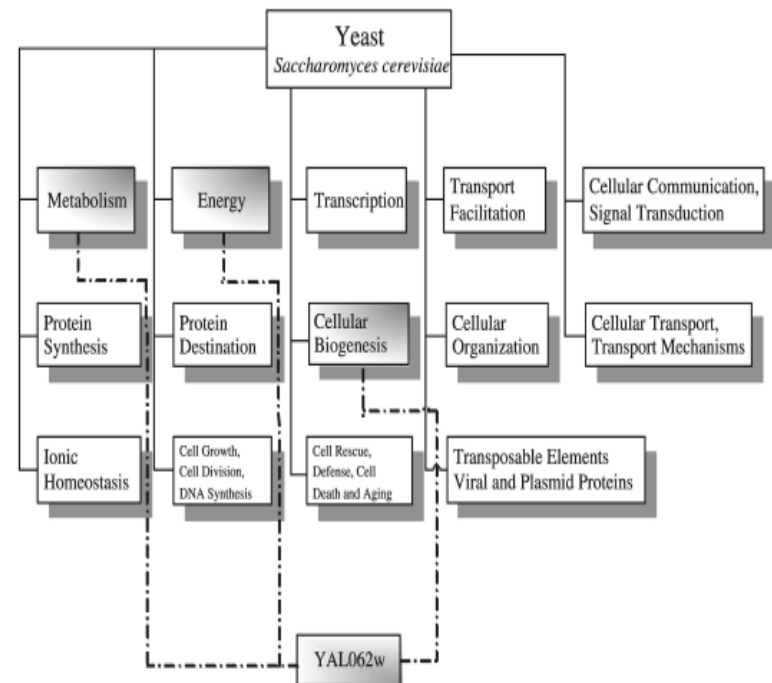
- Genomika funkcjonalna (functional genomics)
  - Określenie funkcji genów i kodowanych przez nie białek w oparciu o dane biologiczne:
    - Mikromacierze DNA
      - Poziomy ekspresji genów w różnych warunkach
    - Sekwencje
      - nukleotydów w danym genie
      - aminokwasów w białku kodowanym przez gen
    - Profile filogenetyczne
      - Ciąg bitów odpowiadających genomom różnych gatunków określających czy rozpatrywany gen występuje w danym gatunku (1) czy też nie (0)

# Klasyfikacja wielokategorialna – zastosowanie w bioinformatyce

- Genomika funkcjonalna (functional genomics)

- Problem klasyfikacji wielokategorialnej:

- Każdy gen powiązany ze zbiorem funkcji (klas)
- Przykład – genom drożdży:
  - Określone 14 klas funkcji genów
  - np. Gen *YAL062w* należy do klas:
    - Metabolism
    - Energy
    - Cellular Biogenesis



# Wyznaczanie klas funkcjonalnych genomu drożdży z pomocą sieci neuronowych

- Zbiór uczący
  - 2417 genów powiązanych z klasami funkcjonalnymi
  - Każdy gen
    - opisany przez 103-wymiarowy wektor (wejście sieci neuronowej)
      - Profile ekspresji genów (z mikromacierzy DNA)
      - Profile filogenetyczne
    - powiązany z podzbiorem zbioru 14 klas funkcjonalnych

# Wyznaczanie klas funkcjonalnych genomu drożdży z pomocą sieci neuronowych

- Parametry sieci neuronowej
  - Współczynnik uczenia: 0.05
  - Learning decay: 0.5
  - Liczba neuronów w warstwie ukrytej:
    - 20 dla funkcji błędu:
      - E1, E2
    - 40 dla funkcji błędu:
      - E3, E4, E5
  - 100 epok

# Wyznaczanie klas funkcjonalnych genomu drożdży z pomocą sieci neuronowych

- Przebieg uczenia
  - Walidacja krzyżowa
    - Zbiór danych losowo dzielony na 10 równych części
    - 10 cykli uczenia
      - 1 część – zbiór testowy
      - 9 części – zbiór uczący
      - Wyznaczenie miar oceny jakości klasyfikacji
  - 5 klasyfikatorów wielokategoryjnych (po jednym dla każdej funkcji błędu)

# Wyznaczanie klas funkcjonalnych genomu drożdży z pomocą sieci neuronowych

- Wyniki eksperymentów
  - Po 5 eksperymentów dla każdego klasyfikatora
  - Walidacja krzyżowa (10 cykli)
  - 50 ocen (hamming loss, one-error, ranking loss) dla każdego klasyfikatora
  - Porównanie klasyfikatorów
    - Test t na poziomie istotności 0.05

# Wyznaczanie klas funkcjonalnych genomu drożdży z pomocą sieci neuronowych

- Średnie i odchylenia standardowe ocen rozpatrywanych klasyfikatorów wielokategoryjnych

Funkcja błędu	Hamming loss		One-error		Ranking loss	
	Średnia	Odch.	Średnia	Odch.	Średnia	Odch.
$E_1$	0,2754	0,0188	0,2324	0,0292	0,1729	0,0150
$E_2$	0,2005	0,0071	0,2311	0,0255	0,1705	0,0102
$E_3$	0,2023	0,0094	0,2351	0,0215	0,1721	0,0107
$E_4$	0,1988	0,0094	0,2252	0,0230	0,1659	0,0120
$E_5$	0,1987	0,0089	0,2247	0,0242	0,1657	0,0117

# Wyznaczanie klas funkcjonalnych genomu drożdży z pomocą sieci neuronowych

- Wyniki testów statystycznych (test t na poziomie istotności 0.05)

Test	Hamming loss p-value	One-error p-value	Ranking loss p-value
$E_1$ vs. $E_2$	<b>0</b>	0,8202	0,3569
$E_1$ vs. $E_3$	<b>0</b>	0,5950	0,7786
$E_1$ vs. $E_4$	<b>0</b>	0,1778	<b>0,0116</b>
$E_1$ vs. $E_5$	<b>0</b>	0,1578	<b>0,0089</b>
$E_2$ vs. $E_3$	0,2787	0,4010	0,4335
$E_2$ vs. $E_4$	0,2981	0,2289	<b>0,0410</b>
$E_2$ vs. $E_5$	0,2594	0,2026	<b>0,0310</b>



# Wyznaczanie klas funkcjonalnych genomu drożdży z pomocą sieci neuronowych

- Porównanie klasyfikatorów wielokategorialnych

- Hamming loss

- $E_2$ ,  $E_3$ ,  $E_4$  i  $E_5$  istotnie lepsze od  $E_1$
- Nie ma statystycznie istotnych różnic pomiędzy  $E_2$  a  $E_3$ ,  $E_4$  i  $E_5$

$$E_2 = \sum_{i=1}^m \frac{\sum_{(k,l) \in Y_i \times \bar{Y}_i} e^{-(c_k^i - c_l^i)} + \sum_{k \in Y_i} e^{-(c_k^i - c_{\bar{0}}^i)} + \sum_{l \in \bar{Y}_i} e^{-(c_{\bar{0}}^i - c_l^i)}}{|Y_i| |\bar{Y}_i| + |Y_i| + |\bar{Y}_i|}$$

$$h(x_i) = \{j \in Y : c_j(x_i) > c_{\bar{0}}(x_i)\}, \quad c_j(x_i) = c_j^i$$

$$E_4 = \sum_{i=1}^m \frac{\sum_{(k,l) \in Y_i \times \bar{Y}_i} e^{-(c_{2k}^i - c_{2l}^i)} + \sum_{k \in Y_i} \sum_{t \in Y_i} e^{-(c_{2k}^i - c_{2t+1}^i)} + \sum_{l \in \bar{Y}_i} \sum_{t \in \bar{Y}_i} e^{-(c_{2t+1}^i - c_{2l}^i)}}{|Y_i| |\bar{Y}_i| + |Y_i|^2 + |\bar{Y}_i|^2}$$

$$h(x_i) = \{j \in Y : c_{2j}(x_i) > c_{2j+1}(x_i)\}, \quad c_j(x_i) = c_j^i$$

$$E_3 = \sum_{i=1}^m \frac{\sum_{(k,l) \in Y_i \times \bar{Y}_i} e^{-(c_{2k}^i - c_{2l}^i)} + \sum_{k \in Y_i} e^{-(c_{2k}^i - c_{2k+1}^i)} + \sum_{l \in \bar{Y}_i} e^{-(c_{2l+1}^i - c_{2l}^i)}}{|Y_i| |\bar{Y}_i| + |Y_i| + |\bar{Y}_i|}$$

$$h(x_i) = \{j \in Y : c_{2j}(x_i) > c_{2j+1}(x_i)\}, \quad c_j(x_i) = c_j^i$$

$$E_5 = \sum_{i=1}^m \frac{\sum_{(k,l) \in Y_i \times \bar{Y}_i} (e^{-(c_{2k}^i - c_{2l}^i)} + e^{-(c_{2l+1}^i - c_{2k+1}^i)}) + \sum_{k \in Y_i} \sum_{t \in Y_i} e^{-(c_{2k}^i - c_{2t+1}^i)} + \sum_{l \in \bar{Y}_i} \sum_{t \in \bar{Y}_i} e^{-(c_{2t+1}^i - c_{2l}^i)}}{2|Y_i| |\bar{Y}_i| + |Y_i|^2 + |\bar{Y}_i|^2}$$

$$h(x_i) = \{j \in Y : c_{2j}(x_i) > c_{2j+1}(x_i)\}, \quad c_j(x_i) = c_j^i$$

# Wyznaczanie klas funkcjonalnych genomu drożdży z pomocą sieci neuronowych

- Porównanie klasyfikatorów wielokategorialnych
  - One-error
    - Wszystkie klasyfikatory porównywalne statystycznie

$$E_2 = \sum_{i=1}^m \frac{\sum_{(k,l) \in Y_i \times \bar{Y}_i} e^{-(c_k^i - c_l^i)} + \sum_{k \in Y_i} e^{-(c_k^i - c_{\bar{0}}^i)} + \sum_{l \in \bar{Y}_i} e^{-(c_{\bar{0}}^i - c_l^i)}}{|Y_i| |\bar{Y}_i| + |Y_i| + |\bar{Y}_i|}$$

$$h(x_i) = \{j \in Y : c_j(x_i) > c_{\bar{0}}(x_i)\}, \quad c_j(x_i) = c_j^i$$

$$E_4 = \sum_{i=1}^m \frac{\sum_{(k,l) \in Y_i \times \bar{Y}_i} e^{-(c_{2k}^i - c_{2l}^i)} + \sum_{k \in Y_i} \sum_{t \in Y_i} e^{-(c_{2k}^i - c_{2t+1}^i)} + \sum_{l \in \bar{Y}_i} \sum_{t \in \bar{Y}_i} e^{-(c_{2t+1}^i - c_{2l}^i)}}{|Y_i| |\bar{Y}_i| + |Y_i|^2 + |\bar{Y}_i|^2}$$

$$h(x_i) = \{j \in Y : c_{2j}(x_i) > c_{2j+1}(x_i)\}, \quad c_j(x_i) = c_j^i$$

$$E_3 = \sum_{i=1}^m \frac{\sum_{(k,l) \in Y_i \times \bar{Y}_i} e^{-(c_{2k}^i - c_{2l}^i)} + \sum_{k \in Y_i} e^{-(c_{2k}^i - c_{2k+1}^i)} + \sum_{l \in \bar{Y}_i} e^{-(c_{2l+1}^i - c_{2l}^i)}}{|Y_i| |\bar{Y}_i| + |Y_i| + |\bar{Y}_i|}$$

$$h(x_i) = \{j \in Y : c_{2j}(x_i) > c_{2j+1}(x_i)\}, \quad c_j(x_i) = c_j^i$$

$$E_5 = \sum_{i=1}^m \frac{\sum_{(k,l) \in Y_i \times \bar{Y}_i} (e^{-(c_{2k}^i - c_{2l}^i)} + e^{-(c_{2l+1}^i - c_{2k+1}^i)}) + \sum_{k \in Y_i} \sum_{t \in Y_i} e^{-(c_{2k}^i - c_{2t+1}^i)} + \sum_{l \in \bar{Y}_i} \sum_{t \in \bar{Y}_i} e^{-(c_{2t+1}^i - c_{2l}^i)}}{2|Y_i| |\bar{Y}_i| + |Y_i|^2 + |\bar{Y}_i|^2}$$

$$h(x_i) = \{j \in Y : c_{2j}(x_i) > c_{2j+1}(x_i)\}, \quad c_j(x_i) = c_j^i$$

# Wyznaczanie klas funkcjonalnych genomu drożdży z pomocą sieci neuronowych

- Porównanie klasyfikatorów wielokategorialnych
  - Ranking loss
    - $E_4$  i  $E_5$  są lepsze niż  $E_1$  i  $E_2$

$$E_2 = \sum_{i=1}^m \frac{\sum_{(k,l) \in Y_i \times \bar{Y}_i} e^{-(c_k^i - c_l^i)} + \sum_{k \in Y_i} e^{-(c_k^i - c_{\bar{0}}^i)} + \sum_{l \in \bar{Y}_i} e^{-(c_{\bar{0}}^i - c_l^i)}}{|Y_i| |\bar{Y}_i| + |Y_i| + |\bar{Y}_i|}$$

$$h(x_i) = \{j \in Y : c_j(x_i) > c_{\bar{0}}(x_i)\}, \quad c_j(x_i) = c_j^i$$

$$E_4 = \sum_{i=1}^m \frac{\sum_{(k,l) \in Y_i \times \bar{Y}_i} e^{-(c_{2k}^i - c_{2l}^i)} + \sum_{k \in Y_i} \sum_{t \in Y_i} e^{-(c_{2k}^i - c_{2t+1}^i)} + \sum_{l \in \bar{Y}_i} \sum_{t \in \bar{Y}_i} e^{-(c_{2t+1}^i - c_{2l}^i)}}{|Y_i| |\bar{Y}_i| + |Y_i|^2 + |\bar{Y}_i|^2}$$

$$h(x_i) = \{j \in Y : c_{2j}(x_i) > c_{2j+1}(x_i)\}, \quad c_j(x_i) = c_j^i$$

$$E_3 = \sum_{i=1}^m \frac{\sum_{(k,l) \in Y_i \times \bar{Y}_i} e^{-(c_{2k}^i - c_{2l}^i)} + \sum_{k \in Y_i} e^{-(c_{2k}^i - c_{2k+1}^i)} + \sum_{l \in \bar{Y}_i} e^{-(c_{2l+1}^i - c_{2l}^i)}}{|Y_i| |\bar{Y}_i| + |Y_i| + |\bar{Y}_i|}$$

$$h(x_i) = \{j \in Y : c_{2j}(x_i) > c_{2j+1}(x_i)\}, \quad c_j(x_i) = c_j^i$$

$$E_5 = \sum_{i=1}^m \frac{\sum_{(k,l) \in Y_i \times \bar{Y}_i} (e^{-(c_{2k}^i - c_{2l}^i)} + e^{-(c_{2l+1}^i - c_{2k+1}^i)}) + \sum_{k \in Y_i} \sum_{t \in Y_i} e^{-(c_{2k}^i - c_{2t+1}^i)} + \sum_{l \in \bar{Y}_i} \sum_{t \in \bar{Y}_i} e^{-(c_{2t+1}^i - c_{2l}^i)}}{2|Y_i| |\bar{Y}_i| + |Y_i|^2 + |\bar{Y}_i|^2}$$

$$h(x_i) = \{j \in Y : c_{2j}(x_i) > c_{2j+1}(x_i)\}, \quad c_j(x_i) = c_j^i$$

# Bibliografia

- Min-Ling Zhang, Zhi-Hua Zhou - "Multilabel Neural Networks with Applications to Functional Genomics and Text Categorization", IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, VOL. 18, NO. 10, 2006
- A. Clare, "Machine Learning and Data Mining for Yeast Functional Genomics", PhD dissertation, Dept. of Computer Science, Univ. of Wales Aberystwyth, 2003
- A. Elisseeff, J. Weston - "A Kernel Method for Multi-Labelled Classification", Advances in Neural Information Processing Systems, vol. 14, pp. 681-687, 2002
- A. Clare, R.D. King, "Knowledge Discovery in Multi-Label Phenotype Data", Lecture Notes in Computer Science, vol. 2168, pp. 42-53, Berlin: Springer, 2001
- P. Pavlidis, J. Weston, J. Cai, and W.N. Grundy, "Combining Microarray Expression Data and Phylogenetic Profiles to Learn Functional Categories Using Support Vector Machines", Proc. Fifth Ann. Int'l Conf. Computational Molecular Biology (RECOMB '01), pp. 242-248, 2001