

# **Klasyfikacja wielokategorialna za pomocą kombinacji sieci neuronowych**

Rafał Grodzicki

---

# Problem klasyfikacji wielokategorialnej (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:

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

# Problem klasyfikacji wielokategorialnej (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  $(\forall y_1 \in Y_i \wedge y_2 \notin Y_i) \quad f(x_i, y_1) > f(x_i, y_2)$
- 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

# Sieć neuronowa w klasyfikacji wielokategorialnej (metoda BP-MLL)

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

# Sieć neuronowa w klasyfikacji wielokategorialnej (metoda BP-MLL)

- 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 w klasyfikacji wielokategorialnej (metoda BP-MLL)

- 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

# Funkcja błędu – modyfikacje

- 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

# 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)

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

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

- Im mniejsza wartość tym lepiej
  - Najlepsza: 0
  - Najgorsza: 1
  - Losowa klasyfikacja: 0,5



# Miary oceny jakości klasyfikacji wielokategoryjnej

- One-error

- Określa jak często kategoria o najwyższej wartości wyjściowej nie należy do zbioru  $Y_i$

$$oneerror(f) = \frac{1}{p} \sum_{i=1}^p [(\arg \max_{y \in Y} f(x_i, y)) \notin Y_i]$$

- Im mniejsza wartość tym lepiej:

- Najlepsza: 0

- Najgorsza: 1

- Losowa klasyfikacja:  $\frac{1}{p} \sum_{i=1}^p \frac{|Y_i|}{Q}$

# Miary oceny jakości klasyfikacji wielokategoryjnej

- 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

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

- Im mniejsza wartość tym lepiej:
  - Najlepsza: 0
  - Najgorsza: 1
  - Losowa klasyfikacja: 0,5

# Miary oceny jakości klasyfikacji wielokategorialnej

## ● Coverage

- Określa uśrednioną liczbę kategorii o wyższej wartości wyjściowej niż kategoria o najniższej wartości wyjściowej należąca do zbioru

$$Y_{icoverage}(f) = \frac{1}{p} \sum_{i=1}^p \left( \max_{y \in Y_i} rank_f(x_i, y) - 1 \right)$$

$$rank_f : X \times Y \rightarrow \{1, 2, \dots, Q\}$$

$$f(x_i, y_1) > f(x_i, y_2) \Rightarrow rank_f(x_i, y_1) < rank_f(x_i, y_2)$$

- Im mniejsza wartość tym lepiej:

- Najlepszy:  $\frac{1}{p} \sum_{i=1}^p |Y_i| - 1$

- Najgorszy:  $Q - 1$

- Losowa klasyfikacja:  $\frac{1}{p} \sum_{i=1}^p \frac{|Y_i| + Q}{2} - 1$

# Miary oceny jakości klasyfikacji wielokategorialnej

- Average precision

- Określa jaka część kategorii o nie mniejszych wartościach wyjściowych niż poszczególne kategorie należące do zbioru  $Y_i$  należy do zbioru  $Y_i$

$$avgprec(f) = \frac{1}{p} \sum_{i=1}^p \frac{1}{|Y_i|} \sum_{y \in Y_i} \frac{|\{y' \in Y_i : rank_f(x_i, y') \leq rank_f(x_i, y)\}|}{rank_f(x_i, y)}$$

$$rank_f : X \times Y \rightarrow \{1, 2, \dots, Q\}$$

$$f(x_i, y_1) > f(x_i, y_2) \Rightarrow rank_f(x_i, y_1) < rank_f(x_i, y_2)$$

- Im większa wartość tym lepiej:
  - Najlepszy: 1
  - Najgorszy: 0
  - Losowa klasyfikacja: 0,5

# Kombinacja sieci neuronowych jako klasyfikator wielokategorialny

- Kombinacja sieci neuronowych (neural network ensemble)
  - N sieci neuronowych
    - Każda sieć uczona niezależnie algorytmem wstecznej propagacji błędu
    - Funkcja błędu

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

# Kombinacja sieci neuronowych jako klasyfikator wielokategorialny

- Kombinacja sieci neuronowych (neural network ensemble)
  - Wartości wyjściowe stanowią średnie arytmetyczne odpowiadających sobie wyjść poszczególnych sieci neuronowych

$$C_j(x_i) = \frac{1}{N} \sum_{s=1}^N c_j^s(x_i), \quad j \in Y$$

$$C_\varrho(x_i) = \frac{1}{N} \sum_{s=1}^N c_\varrho^s(x_i)$$

$$h(x_i) = \{j \in Y : C_j(x_i) > C_\varrho(x_i)\}$$

# Wyznaczanie klas funkcjonalnych genomu drożdży – opis problemu

- 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
    - Profile filogenetyczne
      - Ciąg bitów odpowiadających genomom różnych gatunków określających czy rozpatrywany gen występuje w danym gatunku (1) czy też nie (0)
  - Każdy gen powiązany ze zbiorem funkcji (klas)

# Wyznaczanie klas funkcjonalnych genomu drożdży – opis problemu

- Zbiór danych dla genomu drożdży
  - 103-wymiarowy wektor wejściowy (dane z mikromacierzy DNA i profili filogenetycznych)
  - 14 klas funkcji genów (Metabolism, Protein Synthesis, Ionic Homeostasis, Energy, Protein Destination, Cell Growth and Division, Transcription, Cellular Biogenesis, Cell Rescue and Defense, Transport Facilitation, Cellular Organization, Transposable Elements Viral and Plasmid Proteins, Cellular Communication and Signal Transduction, Cellular Transport and Transport Mechanisms)
  - Zbiór uczący: 1500
  - Zbiór testowy: 917



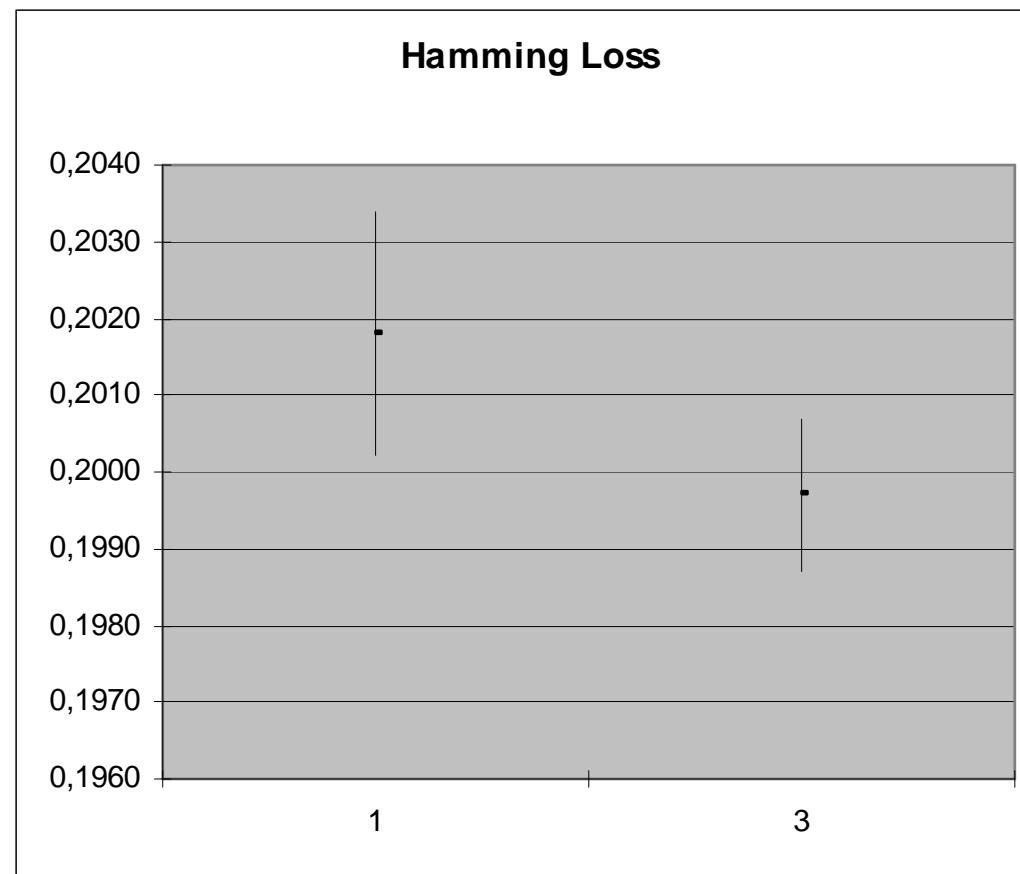
# Wyznaczanie klas funkcjonalnych genomu drożdży – eksperymenty

- Parametry
  - Liczba sieci neuronowych: 3
  - Dla każdej sieci neuronowej
    - Współczynnik uczenia: 0.05
    - Learning decay: 0.5
    - Liczba neuronów w warstwie ukrytej: 14
    - 100 epok
- 30 eksperymentów

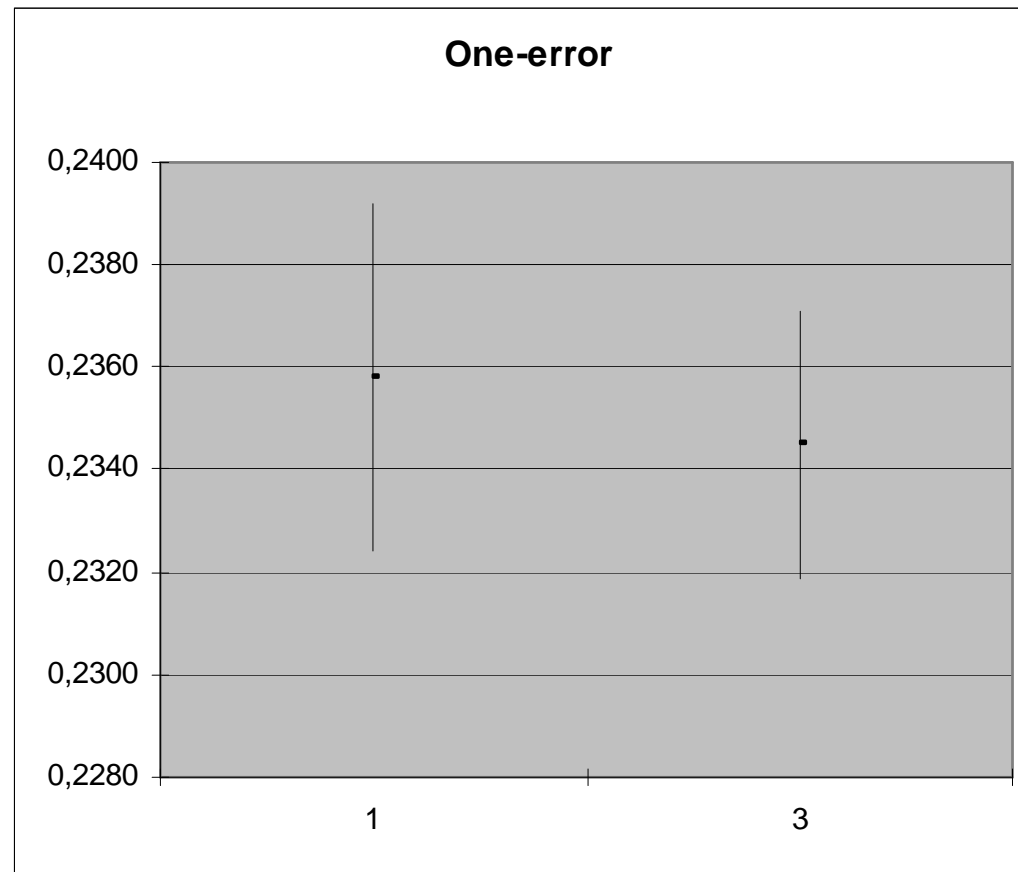
# Wyznaczanie klas funkcjonalnych genomu drożdży – wyniki

Miara	Najlepszy	Najgorszy	Losowa klasyfikacja	1 sieć neuronowa	3 sieci neuronowe	Najlepszy z literatury
Hamming Loss	0	1	0,5	0,2018 $\pm 0,0016$	0,1997 $\pm 0,001$	0,195
One-error	0	1	0,6971	0,2358 $\pm 0,0034$	0,2345 $\pm 0,0026$	0,217
Ranking Loss	0	1	0,5	0,1752 $\pm 0,0023$	0,1723 $\pm 0,001$	0,163
Coverage	3,24	13	8,12	6,5058 $\pm 0,0661$	6,4828 $\pm 0,0321$	-
Average Precision	1	0	0,5	0,7545 $\pm 0,0022$	0,7571 $\pm 0,0012$	0,773

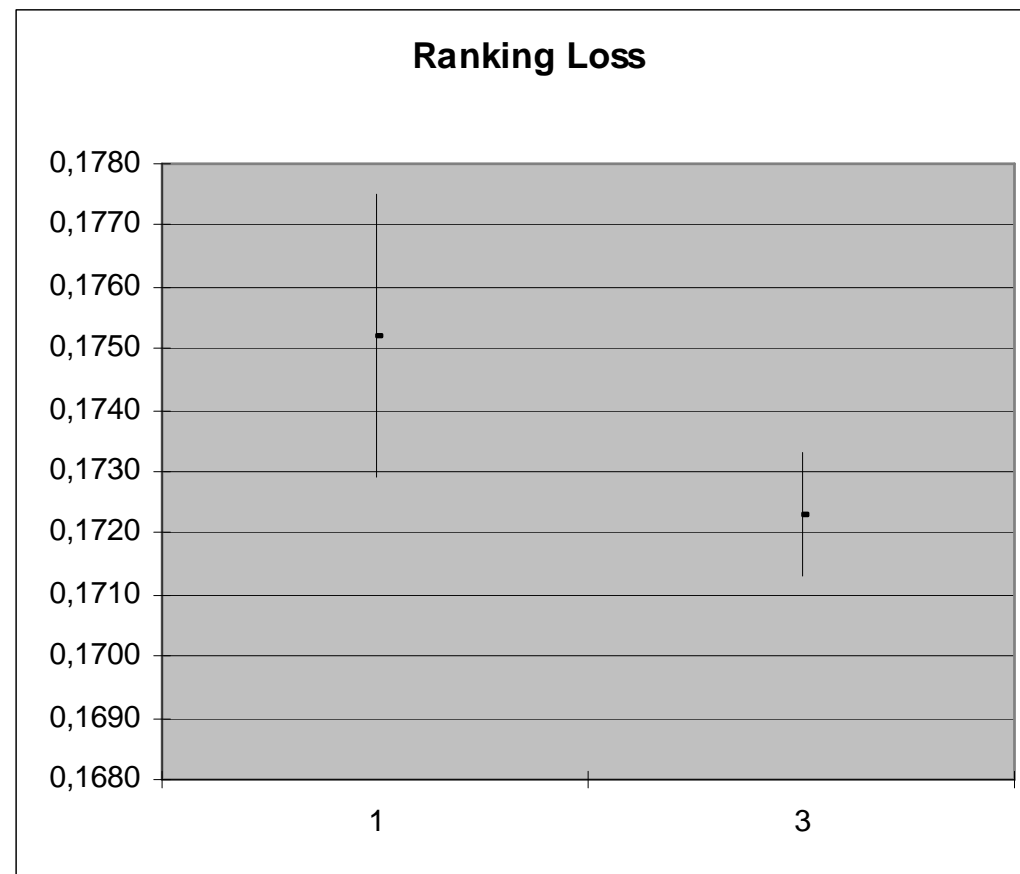
# Wyznaczanie klas funkcjonalnych genomu drożdży – wyniki



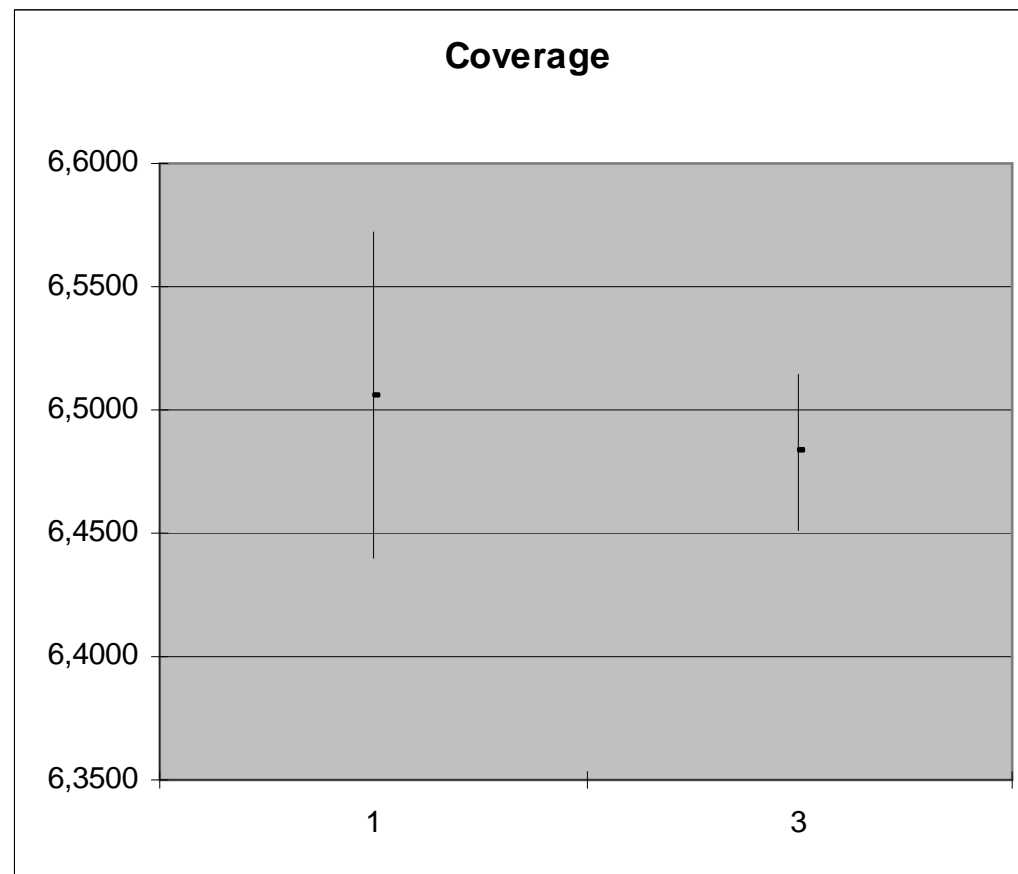
# Wyznaczanie klas funkcjonalnych genomu drożdży – wyniki



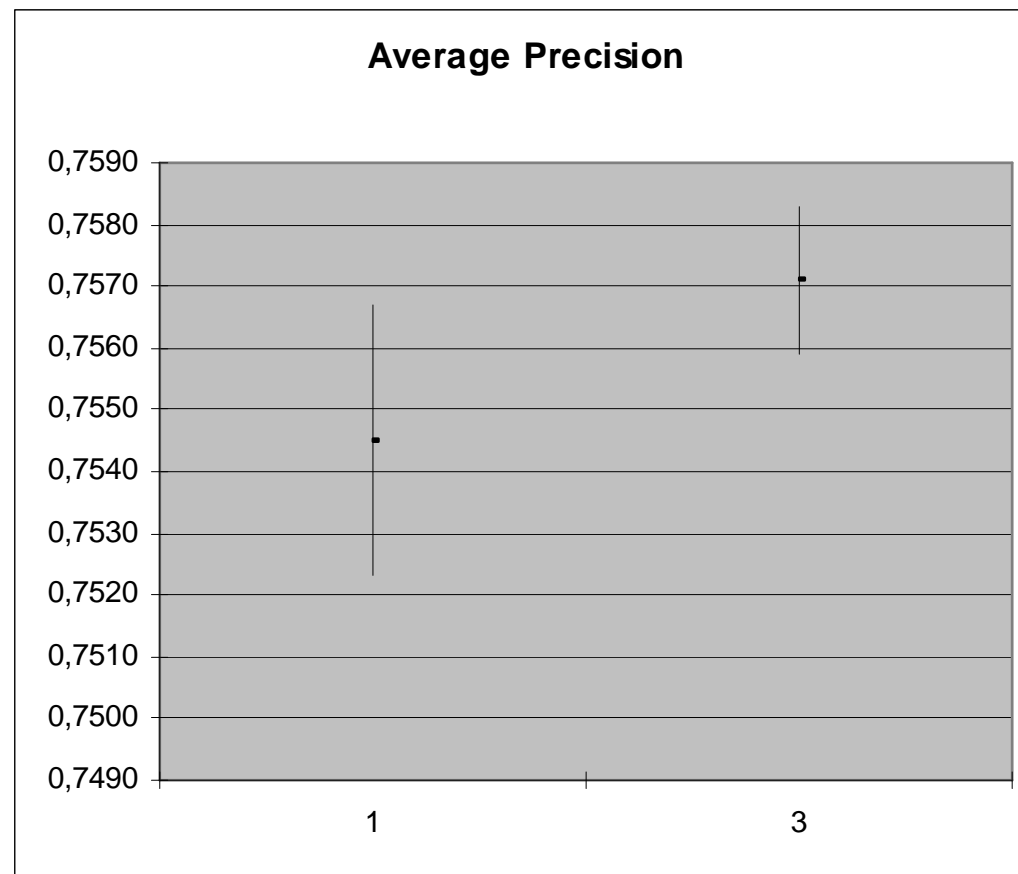
# Wyznaczanie klas funkcjonalnych genomu drożdży – wyniki



# Wyznaczanie klas funkcjonalnych genomu drożdży – wyniki



# Wyznaczanie klas funkcjonalnych genomu drożdży – wyniki



# Wyznaczanie klas funkcjonalnych genomu drożdży – wyniki

- Podsumowanie

- Wyraźna poprawa wyników (w przypadku zastosowania 3 sieci neuronowych) dla miar
  - Hamming Loss
  - Ranking Loss
  - Average Precision
- Porównywalne wyniki dla miar
  - One-error
  - Coverage



# 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, pp. 1338-1351, 2006
- R. Grodzicki, J. Mańdziuk, L. Wang, „Improved Multilabel Classification with Neural Networks”, Lecture Notes in Computer Science, vol. 5199, pp. 409-416, Springer-Verlag, 2008
- A. Elisseeff, J. Weston, „A Kernel Method for Multi-Labelled Classification”, Advances in Neural Information Processing Systems, vol. 14, pp. 681-687, 2002
- LIBSVM Data: Multi-label Classification:  
<http://www.csie.ntu.edu.tw/~cjlin/libsvmtools/datasets/multilabel.html>