SIECI NEURONOWE SNR

semestr zimowy 2021/2022

Pytania / Zadania sprawdzające

1. Skojarzyć dziedziny osiągnięć z nazwiskami: Rosenblatt, Werbos, Goodfellow, Vapnik, Minsky, LeCun, Hinton, Krizhevsky, Fukushima a) perceptron o wyjściach binarnych, b) podstawy sztucznej inteligencji, c) propagacja wsteczna, d) teoria uczenia e) głębokie sieci splotowe, f) SVM, g) GAN.

2. Podstawowy cel bio-modelowania to (wybrać i scharakteryzować jedną lub więcej możliwości): a. tworzenie modeli w celu zrozumienia budowy i działania organizmów żywych i ich organów, b. wykorzystanie biologii do tworzenia nowych urządzeń, c. wykorzystanie biologii do tworzenia nowych metod obliczeniowych

3. Podstawowe cele bio-inżynierii to a. budowa urządzeń zastępujących lub wspomagających niesprawne funkcje człowieka, b. budowa oprzyrządowania medycznego, c. budowa konstrukcji inżynieryjnych wzorowanych na biologii

4. Co to jest sieć neuronowo-glejowa w sieciach biologicznych?

5. Które jony pełnią główną rolę w kształtowaniu potencjału czynnościowego neuronu biologicznego i jaka jest rola każdego z nich

6. Podać przykłady sygnałów w organizmie człowieka, które mogą być modelowane jako tensor:1,2,3,4 i 5 rzędu

7. Podać dwa przykłady wektoryzacji macierzy 3x3 o elementach $a_{ij} = i - j$

8. Wyjaśnić pojęcia: pole recepcyjne, pole aktywacji, obszar agregacji

9. Wyjaśnić pojęcia szybkiej dynamiki i wolnej dynamiki neuronu

10. Wyjaśnić pojęcia: para ucząca, zbiór uczący, ciąg uczący; wskazać na różnicę między nimi

11. Zaznaczyć odpowiedzi prawdziwe, gdzie us: układ statyczny, b: ud: układ dynamiczny, n: nie można ustalić bez dalszych danych:

    a sieć statyczna z uczeniem jest us? ud?, n?

    b sieć dynamiczna z uczeniem jest us? ud? n?

12. Znaleźć tensor powstały w wyniku kontrakcji indeksów i, j tensora $\mathbf{A}_{pij}$, gdzie $A_{pij} = pi + j$, $i, j, p \in \{1, 2\}$.

13. Znaleźć iloczyn zewnętrzny tensorów A_{ij} i B_{pq}, gdzie $A_{ij} = ij$ i $B_{pq} = p + q$

14. Znaleźć iloczyn wewnętrzny tensorów A_{ij} i B_{pq}, gdzie $A_{ij} = ij$ i $B_{pq} = p + q$

15. a) Czy złożenie dwu warstw liniowych można zapisać w postaci jednej warstwy? b) Czy złożenie dwu warstw nieliniowych można zapisać w postaci jednej warstwy? Odpowiedzi szczegółowo uzasadnić.

    Wsk.: Dla standardowych oznaczeń, a) $\boldsymbol{y} = \boldsymbol{\beta} + \boldsymbol{V}(\boldsymbol{b} + \boldsymbol{W}\boldsymbol{u}) = (\boldsymbol{\beta} + \boldsymbol{V}\boldsymbol{b}) + \boldsymbol{V}\boldsymbol{W}\boldsymbol{u}$ czyli otrzymujemy sieć jednowarstwową liniową o obciążeniu $\boldsymbol{b}^* = \boldsymbol{\beta} + \boldsymbol{V}\boldsymbol{b}$ i macierzy wag $\boldsymbol{W}^* = \boldsymbol{V}\boldsymbol{W}$. b) na ogół nie, gdyż $\boldsymbol{y} = \boldsymbol{\gamma}\big(\boldsymbol{\beta} + \boldsymbol{V}\boldsymbol{g}(\boldsymbol{b} + \boldsymbol{W}\boldsymbol{u})\big)$ nie da się ogólnie sprowadzić do postaci $\boldsymbol{y} = \boldsymbol{g}^*(\boldsymbol{\beta}^* + \boldsymbol{V}^*\boldsymbol{u})$

16. Neuron o dwóch wejściach u_1, u_2 i funkcji aktywacji g ma wagi 0.5 i 0.9 i obciążenie 2. Zapisać równanie neuronu w postaci rozszerzonej.

    Wsk.: $\boldsymbol{w}^+ = \begin{bmatrix} 2 \\ 0.5 \\ 0.9 \end{bmatrix}$, $\boldsymbol{u}^+ = \begin{bmatrix} 1 \\ u_1 \\ u_2 \end{bmatrix}$, $y = g({\boldsymbol{w}^+}^\top \boldsymbol{u}^+)$

17. a) Czy każda warstwa o połączeniach lokalnych jest warstwą splotową? b) Czy każda warstwa splotowa jest warstwą o połączeniach lokalnych?

    Wsk.: a) nie, b) formalnie tak, gdy pole recepcyjne obejmuje całą poprzednią warstwę, ale wówczas neuron ten jest jedyny w warstwie

18. Co różni sieci dynamiczne od sieci statycznych?

19. Wartości mapy aktywacji kanału: $u_{x,y} = x - y, x, y = 1, \ldots, 4$. Obliczyć wynik agregacji dla tego kanału w obszarze 2×2 z krokiem 2 dla dwu rodzajów agregacji

20. Dwa kanały warstwy mają postać $u_{x,y,d} = xy + d$, $x, y, d \in \{1, 2\}$, (d - indeks warstwy). Określić warstwę po agregacji w obszarze agregacji 2×2.

21. Określić wyjścia warstwy softmax dla wejść -1, 0, 1.

22. Znaleźć kontrakcję indeksów tensora 4-wymiarowego $A_{ij} = i + j, i, j = 1, \ldots, 4$

23. Przykład, slajd 125

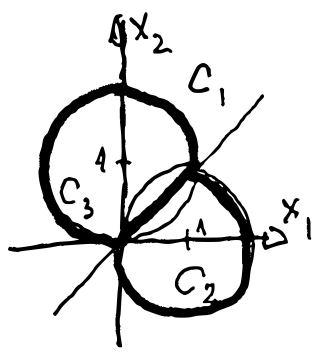
24. Przykład, slajd 127

25. Przykłady, slajdy 131, 132

26. Czy funkcja $f(x_1, x_2) = \begin{bmatrix} x_1^2 + x_2^2 \\ x_1 \\ x_2 \end{bmatrix}$ a) jest pewną funkcją klasyfikującą
b) jest pewną funkcją decyzyjną
c) jest pewną funkcją przynależności
Odpowiedzi uzasadnić.

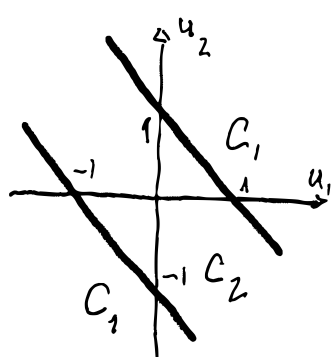
27. Dla funkcji decyzyjnej $f(x_1, x_2) = \begin{bmatrix} x_1^2 + x_2^2 \\ 2\,x_1 \\ 2\,x_2 \end{bmatrix}$ a) narysować zbiór separujący, b) określić funkcję klasyfikującą, c) określić funkcje wskaźnikowe.

Wsk.: a) $\mathcal{C}_2$: $2x_1 - x_1^2 - x_2^2 > 0$ i $x_1 > x_2$ tzn. $(x_1 - 1)^2 + x_2^2 < 1$ i $x_1 > x_2$
$\mathcal{C}_3$: $2x_2 - x_1^2 - x_2^2 > 0$ i $x_2 > x_1$ tzn. $x_1^2 + (x_2 - 1)^2 < 1$ i $x_2 > x_1$



28. Znaleźć funkcję klasyfikującą i wyznaczyć klasy dla funkcji decyzyjnych $d_1(u_1, u_2) = u_1^2 + u_2^2 - 1$, $d_2(u_1, u_2) = -2\,u_1\,u_2$.

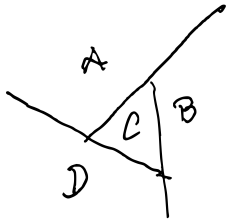
Wsk.: $\mathcal{C}_1$: $u_1^2 + u_2^2 + 2u_1u_2 - 1 > 0$ czyli $(u_1 + u_2)^2 > 1$ a zatem $\mathcal{C}_1 = \{(u_1, u) : u_1 + u_2 < -1 \cup u_1 + u_2 > 1\}$, $\mathcal{C}_2 = \{(u_1, u) : -1 < u_1 + u_2 < -1\}$

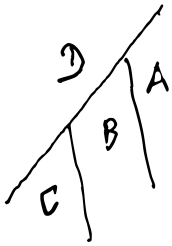


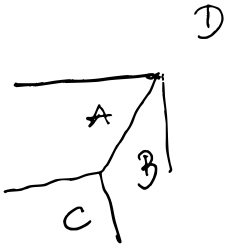
29. Wyznaczyć klasy dla funkcji decyzyjnych $d_1(u_1, u_2) = u_1^2 - u_2^2 - 2u_1$, $d_2(u_1, u_2) = u_1^2 - u_2^2 - 2u_2$. Czy klasyfikacja ta jest liniowa?

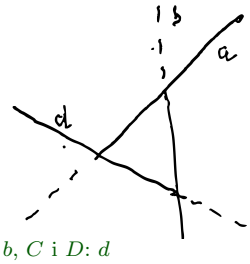
Wsk.: $\mathcal{C}_1 : u_2 - u_1 > 0$; klasyfikacja jest liniowa

30. Czy poniższe klasyfikacje na cztery klasy są liniowe? Odpowiedzi uzasadnić.

a) 

b) 

c) 

Wsk.: a) 

, klasyfikacja jest liniowa, prostymi separującymi klasy są: A i B: a, A i C: a, A i D: d, B i C: b, B i D: b, C i D: d

Wsk.: b)

klasyfikacja jest liniowa, prostymi separującymi klasy są: A i B: a, A i C:a (lub inne warianty), AiD:d, B i C: b, B i D: d, C i D: d

Wsk.: c) klasyfikacja nieliniowa: D nie może być liniowo odseparowany od pozostałych klas)

31. Uzasadnić, że nie wszystkie funkcje logiczne dwóch zmiennych par (Z_1, Z_2) można otrzymać w wyniku klasyfikacji liniowych odpowiadających im obrazów binarnych

Wsk.: funkcjom $\neg Z_1 \cap Z_2$ i $Z_1 \cap \neg Z_2$ odpowiadają klasyfikacje nieliniowe

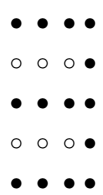
32. Wejściami sieci jest n różnych obrazów o 2×2 pikselach. Na ile sposobów można sklasyfikować te obrazy? Jaka jest największa liczba takich obrazów, które można sklasyfikować liniowo na dowolne dwie klasy?

Wsk.: na 2^n sposobów; dla 4 pikseli jedynie $4 + 1 = 5$ obrazów może być dowolnie sklasyfikowanych liniowo

33. Wejściami sieci są obrazy o formacie kwadratowym 32×32 pikseli. Każdy obraz może być przekształcony losowo do obrazu o większej liczbie pikseli $q \times q$. Jaka wielkość obrazu jest niezbędna, by dowolna klasyfikacja a) 1000 b) 2000 obrazów była liniowa?

Wsk.: $32 \times 32 + 1 = 1025$ klasyfikacji liniowych; $n \leqslant p + 1 = q^2 + 1$ czyli $q >= \sqrt{n-1}$ a) nie trzeba przekształcać, b) q= 45

34. W przestrzeni obrazów $\mathbb{R}^2$ znajdują się punkty



Jak można interpretować powyższą informację: a) obraz cyfry "3", b) 20 obrazów binarnych c) dwie klasy obrazów, klasyfikacja nieliniowa

35. Narysować rzut kostki 4-wymiarowej na przestrzeń dwuwymiarową

Wsk.: b); 20 obrazów: $(1,1)$, $(1,2)$, $(1,3)$, $(1,4)$, $(2,4)$, $(3,1)$ itd.

36. Miara Lebesgue'a kuli jednostkowej w $\mathbb{R}^m$ a) monotonicznie maleje do zera z m, b) monotonicznie rośnie do ∞ z m, c) początkowo rośnie a potem maleje monotonicznie z m do zera.

37. Przedyskutować różnicę między klasyfikacją nadzorowaną, nienadzorowaną i częściowo nadzorowaną

38. Jaka jest różnica między koncepcjami algorytmu μ-LMS i algorytmu Kaczmarza

39. Porównać Adaline a) z perceptronem Rosenblatta b) z warstwą sieci perceptronowej o sigmoidalnych funkcjach aktywacji. Jakie są wspólne cechy, jakie różnice?

40. Czy Adaline może aproksymować dowolną funkcję ciągłą z dowolną dokładnością?

41. Scharakteryzować własności kilku technik uczenia neuronów liniowych

42. Maszyna liniowa, klasyfikująca obrazy w $\mathbb{R}^3$ na cztery klasy, zakwalifikowała do klasy 2 obraz $\boldsymbol{u} = \begin{bmatrix} 1 & 2 & -2 \end{bmatrix}^\top$ należący do klasy 3. Określić tablicę korekcji wag i wektor korekcji obciążenia. Ile wag ma ta sieć?

Wsk.: Sieć ma 12 wag i 4 obciążenia;

$$\Delta \boldsymbol{W} = \begin{bmatrix} 0 & 0 & 0 \\ -1, & -2 & 2 \\ 1 & 2 & -2 \\ 0 & 0 & 0 \end{bmatrix}, \qquad \Delta \boldsymbol{b} = \begin{bmatrix} 0 \\ -1 \\ 1 \\ 0 \end{bmatrix}$$

43. a) W procesie uczenia perceptronu Rosenblatta, dla obrazu $[1\ 3\ 2]^\top$, należącego do klasy $\mathcal{C}_-$, wyjście perceptronu było równe 1. Określić przyrosty wektora wag każdego neuronu wyjściowego w tym kroku uczenia. b) Określić przyrosty wag równoważnej maszyny liniowej klasyfikującej na dwie klasy $\mathcal{C}_1 = \mathcal{C}_+$ i $\mathcal{C}_2 = \mathcal{C}_-$.

Wsk.: a) $\Delta \boldsymbol{w} = \begin{bmatrix} -1 \\ -3 \\ -2 \end{bmatrix}$; $\Delta b = -1$; b) $\Delta \boldsymbol{W} = \begin{bmatrix} -1 & -3 & -2 \\ 1 & 3 & 2 \end{bmatrix}$, $\Delta \boldsymbol{b} = \begin{bmatrix} -1 \\ 1 \end{bmatrix}$

44. Jaka jest różnica między koncepcjami algorytmu μ-LMS i algorytmu Kaczmarza. Który z tych algorytmów nie może być zastosowany do perceptronu Rosenblatta i daczenia

45. Opisać podstawową sieć perceptronową. Czy dowolna dwuwarstwowa sieć perceptronowa ma identyczne właściwości jak sieć podstawowa.

Wsk.: Tak, o ile funkcja aktywacji warstwy wyjściowej jest monotoniczna

46. Czy nieliniowość typu ReLU spełnia warunki a) tw. Cybenki b) tw. Hornika

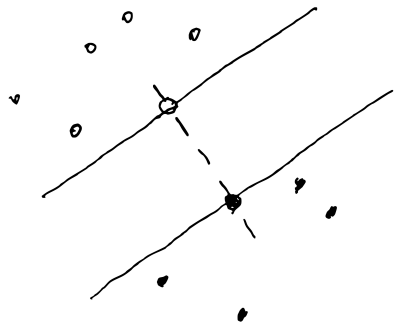
47. Zgodnie z twierdzeniami aproksymacyjnymi, rozmiar warstwy ukrytej powinien być „odpowiednio duży". Co się stanie, jeśli jest on "za mały", czy to źle, jeśli jest "za duży"

48. Przykłady, slajd 326

49. Jaka jest podstawowa wada jednowarstwowego perceptronu Rosenblatta jako klasyfikatora? Czy wadę taką posiada maszyna wektorów podpierających? Odpowiedzi uzasadnić.

50. Jaka może być najmniejsza liczba obrazów podpierających? Jakie warunki muszą Odpowiedź zilustrować w $\mathbb{R}^2$

Wsk.: 2 obrazy; zachodzi, gdy prosta łącząca obrazy podpierające jest ortogonalna do hiperpłaszczyzny separującej



51. Na czym polega ‘kernel trick’? Znaleźć jądro odpowiadające funkcji cech $\boldsymbol{\varphi}(x_1, x_2) = \begin{bmatrix} x_1 \\ x_2 \\ x_1 x_2 \end{bmatrix}$.

    Wsk.: $k(\boldsymbol{x}, \boldsymbol{x}') = (x_1')(x_1'') + (x_2')(x_2'') + (x_1' x_2')(x_1'' x_2'') = \boldsymbol{x}'^{\top} \boldsymbol{x}''$.

52. Znaleźć funkcję cech prowadzącą do jądra $k(\boldsymbol{u}, \boldsymbol{u}') = 1 + u_1 u_1' + (1 + u_2 u_2')^2$, $\boldsymbol{u}, \boldsymbol{u}' \in \mathbb{R}^2$.

    Wsk.: Funkcja cech: Ponieważ $k(\boldsymbol{u}, \boldsymbol{u}') = 1 + u_1 u_1' + \left(1 + 2u_2 u_2' + (u_2 u_2')^2\right)$ więc można przyjąć $\boldsymbol{\varphi}(\boldsymbol{u}) = \sqrt{2} + u_1 + \sqrt{2} u_2 + u_2^2$

53. Czy można skonstruować jednowarstwową maszynę wektorów podpierających, gdy obrazy nie są liniowo rozdzielne? Czy zawsze postępowanie takie jest możliwe? Odpowiedź uzasadnić.

    Wsk.: Tak; omówić wprowadzenie zmiennych dopełniających

54. Jaką rolę odgrywają zmienne dopełniające w SVM?

    Wsk.: Omówić konstrukcję dwuwarstwową maszyny wektorów, gdy obrazy nie są liniowo rozdzielne.

55. Jak dobrać funkcję aktywacji aby perceptron dwuwarstwowy aproksymował dowolnie dokładnie funkcję $f(u) = \operatorname{sign}(u)$ dla $u \in [-1, 1]$. Podać przykład funkcji aktywacji, która **nie spełnia** warunków. Wyjaśnić przyjęte znaczenie sformułowania “aproksymuje dowolnie dokładnie”.

56. Scharakteryzować twierdzenie aproksymacyjne właściwe dla aproksymacji funkcji $f(x) = \frac{1}{\pi(1+x^2)}$, gdzie $\xi \sim \mathcal{N}_{0,1}$ (zatem $f(\xi)$ ma rozkład t_1). Wyjaśnić znaczenie użytych w twierdzeniu terminów, a w szczególności a) “aproksymuje z dowolną dokładnością” b) “dowolnie dokładnie” c) ”odległość”.

57. Omówić pojęcia odległości funkcji, stosowane w twierdzenia aproksymacyjnych dla sieci neuronowych

58. Omówić strukturę perceptronu wielowarstwowego jako złożenia modeli liniowych i nieliniowych

    Wsk.: Przedstawić strukturę jako złożenie LIN1 $\circ$ NONLIN $\circ$ LIN2, definiując każdą z tych struktur dla perceptronu: LIN2: $\boldsymbol{z} = \boldsymbol{W}\boldsymbol{u}$, $\boldsymbol{s} = \boldsymbol{g}(\boldsymbol{z})$, LIN1: $\boldsymbol{y} = \boldsymbol{V}\boldsymbol{s}$

59. Scharakteryzować estymator bayesowski. Czy aproksymator globalnie optymalny może się od niego różnić? Czy koszt bayesowski może nie być zerowy? (odpowiedzi uzasadnić)

60. Dla 5-elementowego zbioru uczącego $\boldsymbol{u}_k = |k - 2|$, $\boldsymbol{y}_k = (k - 3)^k$. Obliczyć

    a empiryczną częstość występowania punktu $\boldsymbol{u} = 3$

    b empiryczną częstość pojawiania się pary $\boldsymbol{u} = 3, \boldsymbol{y} = 0$

    c empiryczne prawdopodobieństwa warunkowe $\boldsymbol{y} \,|\, \boldsymbol{u} = 1$

61. Scharakteryzować źródła losowości kosztu empirycznego

62. Warunkiem koniecznym skończoności luki generalizacji rodziny aproksymatorów jest równoczesne spełnienie warunków:

    a liczność rodziny jest skończona

    b liczność rodziny jest nieskończona

    c wymiar vcjest skończony

    d rozmiar próby jest nieskończony

    Podać zestawy warunków, dla których powyższe stwierdzenie jest prawdziwe.

63. Koszt empiryczny jest

    a jest zawsze *leq* od kosztu teoretycznego

    b jest zawsze *geq*od kosztu teoretycznego

    c może być *leq* lub $\geqslant$ od kosztu teoretycznego

    Prawidłową odpowiedź/odpowiedzi uzasadnić

64. Przykłady, slajd 326

65. Które zjawisko jest niebezpieczne przy ocenie jakości sieci na podstawie kosztu empirycznego

a koszt empiryczny jest $>$ od kosztu teoretycznego

b koszt empiryczny jest $<$ od kosztu teoretycznego

Uzasadnić odpowiedź.

66. Dla rodziny $\mathscr{C}$ wszystkich odcinków w $\mathbb{R}$ zbadać (podać przykłady w przypadku odpowiedzi twierdzących, uzasadnić odpowiedzi przeczące), czy

a istnieje zbiór jednoelementowy w $\mathbb{R}$ rozbijany przez $\mathscr{C}$

b istnieje zbiór dwuelementowy w $\mathbb{R}$ rozbijany przez $\mathscr{C}$

c istnieje zbiór trójelementowy w $\mathbb{R}$ rozbijany przez $\mathscr{C}$

Podać wymiar VC rodziny $\mathscr{C}$.

Wsk.: a) tak, np. $\{2\}$: odcinki $[0,1]$, $[0,3]$; tak, np. $\{2,4\}$: odcinki $[0,1]$, $[0,3]$, $[3,5]$, $[1,5]$; c) nie; dla trzech punktów $a < b < c$ zbiór $\{a,c\}$ nie może być wyodrębniony przez odcinek

67. Co oznacza, że liczba parametrów (wag, obciążeń) sieci powinna być "odpowiednio duża"? Co się stanie, jeśli jest ich "za mało"? Czy to źle, jeśli jest ich "za dużo"? Co to jest błąd generalizacji?

68. Dana jest sieć neuronowa o wymiarze VC równym 20, która klasyfikuje punkty na płaszczyźnie. Czy można skonstruować taki k-elementowy zbiór punktów, dla którego

(a) istnieje taki jego podział na dwie klasy, który nie może być poprawnie zrealizowany przez tę sieć

(b) istnieje taki jego podział na dwie klasy, który może być poprawnie zrealizowany przez sieć

(c) sieć potrafi sklasyfikować ten zbiór na dowolnie wybrane dwie klasy

Odpowiedzi udzielić dla (i) k=19, (ii) k=20, (iii) k=21

Wsk.: (a) niezależnie od vc, zawsze może istnieć pewna konfiguracja punktów, która nie może być sklasyfikowana przez sieć; tzn. TAK, niezależnie od k

(b) TAK dla każdego k, bo niezależnie od k, zawsze może istnieć pewna konfiguracja punktów, która może być prawidłowo sklasyfikowana;

(c) vc $= 20$, zatem TAK dla $k = 19$ i $k = 20$, NIE dla $k = 21$.

69. Określić wymiar VC rodziny wszystkich kół na płaszczyźnie$\mathscr{C}$, a następnie zbadać (podać przykłady dla odpowiedzi twierdzących, uzasadnić odpowiedzi przeczące), czy

(a) istnieje zbiór 3-elementowy na płaszczyźnie rozbijany przez $\mathscr{C}$

(b) istnieje zbiór 3-elementowy na płaszczyźnie nie rozbijany przez $\mathscr{C}$

(c) istnieje zbiór 4-elementowy na płaszczyźnie rozbijany przez $\mathscr{C}$

(d) istnieje zbiór 4-elementowy na płaszczyźnie nie rozbijany przez $\mathscr{C}$

Wsk.: Dowolny zbiór 3-elementowy w $\mathbb{R}^2$ może być rozbity przez koła (rys. i); nie istnieje zbiór 4-elementowy, który może być rozbity przez koła na płaszczyźnie: w konfiguracji wielokąta wypukłego (rys. ii) nie można wyodrębnić par punktów naprzeciwległych, a w konfiguracji trójkąta (rys. iii) nie można wyodrębnić 3 punktów zewnętrznych. W rezultacie vc $= 3$, a więc (a) TAK, (c), NIE; niezależnie od vc mogą istnieć zbiory nierozbijane przez $\mathscr{C}$, więc (b) TAK, rys. iv, (d) TAK.

70. Określić wymiar VC rodziny $\mathscr{C}$ zbiorów postaci $(-\infty, x)\times(-\infty, y)$ w $\mathbb{R}^2$ i zbadać (podać przykłady dla odpowiedzi twierdzących, uzasadnić odpowiedzi przeczące), czy a) istnieje zbiór dwuelementowy w $\mathbb{R}^2$ rozbijany $\mathscr{C}$, b) istnieje zbiór dwuelementowy w $\mathbb{R}^2$ nie rozbijany przez $\mathscr{C}$ (rys.), c) istnieje zbiór trójelementowy w $\mathbb{R}^2$ rozbijany przez $\mathscr{C}$,

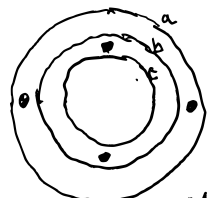
Wsk.: vc = 2. a) TAK, b) TAK, c) NIE

71. Zbadać, czy rodzina odcinków skończonych w $\mathbb{R}$ rozbija zbiory a) dwuelementowe b) trójelementowe c) czteroelementowe. Określić wymiar VC tej rodziny.

    Wsk.: a) TAK, b) NIE, c) NIE; $vc = 2$

72. Określić wymiar VC rodziny $\mathscr{C}$ pierścieni $\mathscr{C}$ w $\mathbb{R}^2$ (zbiory postaci $r_1^2 \leqslant (x - x_0)^2 + (y - y_0)^2 \leqslant r_2^2$ dla dowolnych x_0, y_0, $0 < r_1 < r_2$).

    Wsk.: vc = 4; wyodrębnienie dwu przeciwległych punktów (rys.) jest tu możliwe: pierścienie (a, b) i (b, c). Łatwo pokazać możliwość wyodrębniania pozostałych punktów, ich par i trójek.



73. W jaki sposób pojęcie wymiaru VC jest związane z klasyfikatorami neuronowymi.

    Wsk.: Wyjścia sieci $\boldsymbol{y} \in \mathbb{R}^q$ definiują funkcje decyzyjne, a zatem określają klasyfikacje w $\mathbb{R}^q$. Wymiar VC jest największą licznością zbioru punktów rozbijanego przez wyjścia sieci.

74. Omówić techniki zmniejszania błędu nadmiernego dopasowania

75. Do czego służy regularyzacja sieci?. Omówić krótko techniki regularyzacji

76. Omówić technikę dropout. Jaki jest cel jej stosowania?

77. Na czym polega współdzielenie wag, jak można usprawiedliwić jego stosowanie i jaki jest jego cel?

78. Na czym polega technika tłumienia wag i do czego jest stosowana

79. Scharakteryzować różnicę pomiędzy aproksymatorem liniowym a aproksymatorem liniowym względem parametrów

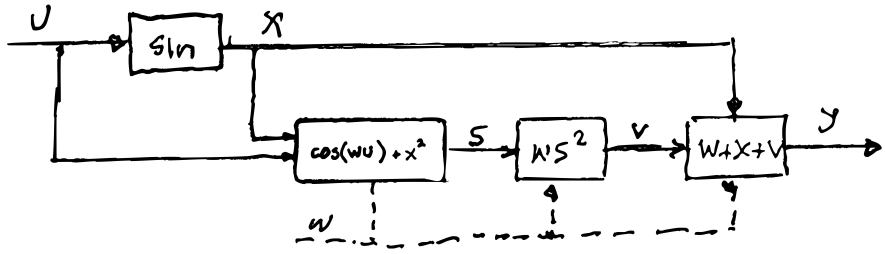
80. Dla wszystkich klasyfikatorów danej klasy, dowolnego rozmiaru próby $N \geqslant 1000$ i dowolnego poziomu ufności $1 - \delta^o$, luka generalizacji nie przekracza $\epsilon^o = 100/\delta^0$. Uzasadnić, że dla dowolnego poziomu ufności $1 - \delta^o$, luka generalizacji ϵ jest dowolnie mała ($\epsilon < \epsilon^o$), o ile próba jest odpowiednio duża $N > 1000$. Określić ϵ^0.

81. Sformułować podstawowe różnice pomiędzy algorytmami propagacji wstecznej i propagacji prostej.

82. Narysować graf funkcji i zapisać algorytm propagacji wstecznej służący do obliczania pochodnej (a) $\dfrac{\mathrm{d}y}{\mathrm{d}u}$ dla ustalonego w (b) $\dfrac{\mathrm{d}y}{\mathrm{d}w}$ dla ustalonego u, dla funkcji y określonej wzorami

$$
\begin{aligned}
x &= \sin u \\
s &= \cos(wu) + x^2 \\
v &= ws^2 \\
y &= x + v + w
\end{aligned}
$$

Wsk.: Należy najpierw sporządzić graf funkcji (rys).



Algorytm musi uwzględniać obliczanie funkcji a następnie obliczanie pochodnej.

(a) Obliczenie $\frac{\mathrm{d}y}{\mathrm{d}u}$

obliczanie wartości funkcji : $x = \sin(u), \quad s = x^2 + \cos(wu), \quad v = ws^2 \quad y = v + w + x$

obliczanie wartości pochodnych : $D_y = 1, \quad D_v = D_y, \quad D_s = 2wsD_v, \quad D_x = 2xD_s + D_y, \quad D_u = -w\sin(wu)D_s + \cos(u)D_x, \quad \frac{\mathrm{d}y}{\mathrm{d}u} = D_u$

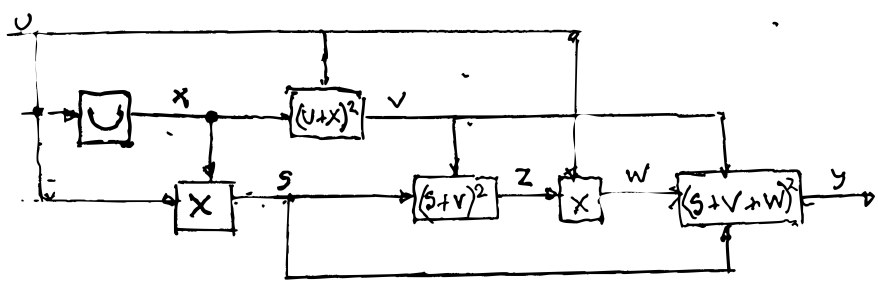
(b) Obliczenie $\frac{\mathrm{d}y}{\mathrm{d}w}$:

$$D_y = 1, \quad D_v = D_y, \quad D_s = 2wsD_v, \quad D_w = s^2D_v + D_y - u\sin(wu)D_s, \quad \frac{\mathrm{d}y}{\mathrm{d}w} = D_w$$

83. Zapisać algorytm obliczania pochodnej $\frac{\mathrm{d}y}{\mathrm{d}u}$ przy użyciu propagacji wstecznej dla systemu

$$\begin{aligned} y &= (s + w + v)^2 \\ s &= xu \\ w &= uz \\ v &= (u + x)^2 \\ z &= (s + v)^2 \\ x &= u^2 \end{aligned}$$

Wsk.: Należy najpierw sporządzić graf funkcji (rys). Symbol $\times$ na rysunku oznacza mnożenie, symbol $\bigcup$ funkcję kwadratową.



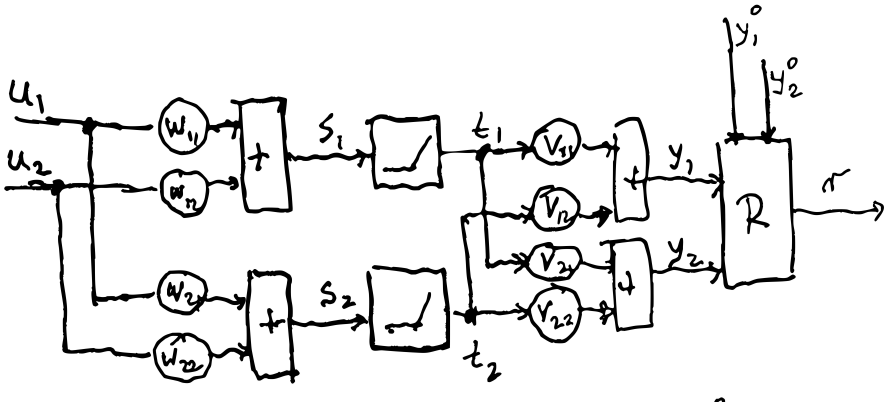
Algorytm musi uwzględniać obliczanie funkcji a następnie obliczanie pochodnej.

obliczanie wartości funkcji : $x = u^2, \quad s = xu, \quad v = (x + u)^2, \quad z = (s + v)^2, \quad w = uz, \quad y = (v + w + s)^2$

obliczanie wartości pochodnych : $D_y = 1, \quad D_w = 2(s + v + w)D_y, \quad D_z = uD_w, \quad D_s = 2(v + s)D_z + 2(s + v + w)D_y,$

$D_v = 2(v + s)D_z + 2(s + v + w)D_y, \quad D_x = 2(u + x)D_v + uD_s, \quad D_u = 2uD_x + 2(u + x)D_v + zD_w + xD_s, \quad \frac{\mathrm{d}y}{\mathrm{d}u} = D_u$

84. Zapisać **algorytm** obliczania pochodnej $\frac{\mathrm{d}r}{\mathrm{d}w_{11}}$ z zastosowaniem propagacji wstecznej dla następującego systemu (przyjąć $r = (y_1 - y_1^o)^2 + (y_2 - y_2^0)^2$)

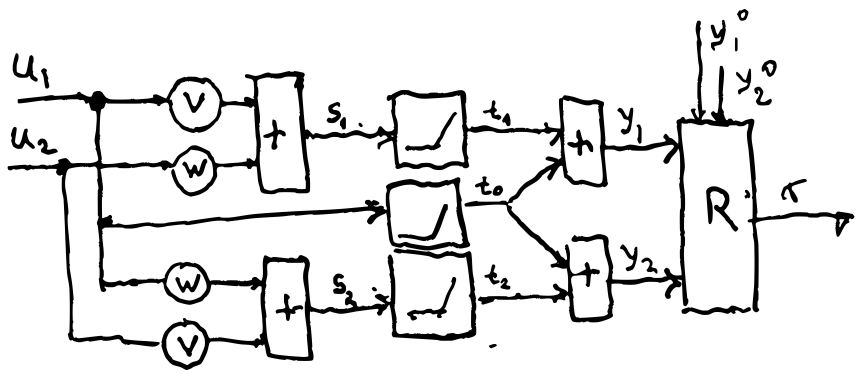


Wsk.: Zapisać najpierw algorytm obliczania funkcji (np. jak w zad. **??**), a następnie:

$$D_{y_1} = 2(y_1 - y_1^0), \quad D_{y_2} = 2(y_2 - y_2^0), \quad D_{t_1} = v_{11}D_{y_1} + v_{21}D_{y_2} \quad ((t_1 \text{ wpływa na } y_1 \text{ i } y_2),$$

$$D_{s_1} = s_1\mathbf{1}(s_1)D_{t_1}, \quad D_{s_2} = s_2\mathbf{1}(s_2)D_{t_2}, \quad D_{w_{11}} = u_1 D_{s_1} \quad (\text{pochodna iloczynu } w_{11}u_1)), \quad \frac{\mathrm{d}r}{\mathrm{d}w_{11}} = D_{w_{11}}$$

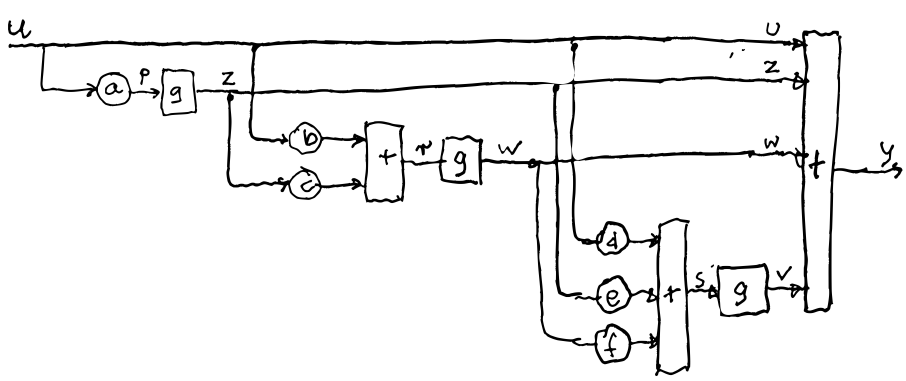
85. Zapisać **algorytm** obliczania pochodnej (a) $\dfrac{\mathrm{d}r}{\mathrm{d}u_1}$ (b) $\dfrac{\mathrm{d}r}{\mathrm{d}v}$ z zastosowaniem propagacji wstecznej dla następującego systemu



Wsk.: W obu przypadkach zapisać najpierw algorytm obliczania funkcji, a następnie: (a)
$D_{y_1} = 2(y_1 - y_1^0)$, $D_{y_2} = 2(y_2 - y_2^0)$, $D_{t_0} = D_{y_1} + D_{y_2}$, $D_{t_1} = D_{y_1}$, $D_{t_2} = D_{y_2}$, $D_{s_1} = s_1\mathbf{1}(s_1)D_{t_1}$, $D_{s_2} = s_2\mathbf{1}(s_2)D_{t_2}$,
$D_{u_1} = vD_{s_1} + wD_{s_2} + u_1\mathbf{1}(u_1)D_{t_0}$, $\dfrac{\mathrm{d}r}{\mathrm{d}u_1} = D_{U_1}$
(b) Należy tylko zastąpić ostatnią linię przez

$$D_v = u_1 D_{s_1} + u_2 D_{s_2}, \quad \frac{\mathrm{d}r}{\mathrm{d}v} = D_v$$

86. Zapisać **algorytm** obliczania pochodnej $\dfrac{\mathrm{d}y}{\mathrm{d}u}$ z zastosowaniem propagacji zwrotnej dla następującego systemu



Wsk.:

$D_y = 1, \quad D_v = D_y, \quad D_s = g'(s)D_v, \quad D_w = D_y + fD_s, \quad D_r = g'(r)D_w, \quad D_z = D_y + cD_r + eD_s, \quad D_p = g'(p)D_z, \quad D_u = D_y + aD_p + bD_r$

87. Omówić co najmniej trzy metody obliczeniowe minimalizacji błędu aproksymacji perceptronu dwuwarstwowego przy założeniu, że znane są pochodne funkcji celu względem wag.

88. Na czym polega i gdzie w sieciach neuronowych ma zastosowanie wygładzanie inercyjne? Omówić warianty wygładzania inercyjnego

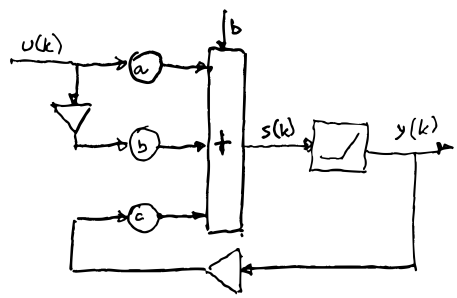
89. W jaki sposób propagacja wsteczna może być zastosowana do obliczania drugich pochodnych?

90. Zastosować metodę Kumara do określenia rozkładu początkowego wag dla funkcji aktywacji "soft-plus" $g(x) = \ln(1 + \exp x)$

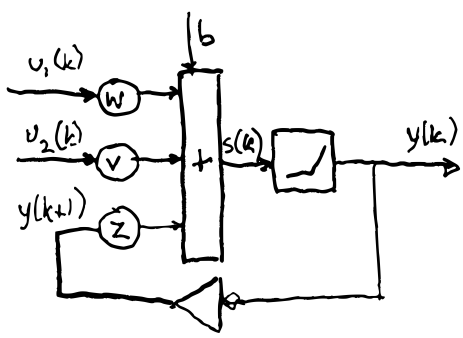
91. Przy klasyfikacji na trzy klasy na wyjściu softmax sieci dla obrazu w klasie 2 pojawiły się wartości 0.1, 0.7, 0.2. Jaka jest wartość kosztu chwilowego (entropia)?

    Wsk.: $-\ln(0.7)$

92. Rozwinąć w czasie sieć jednowarstwową L ze sprzężeniem zwrotnym (poniżej) do postaci sieci głębokiej.



a)



b)

Ile różnych wag ma taka sieć głęboka, jeśli $\boldsymbol{u} \in \mathbb{R}^3$, $\boldsymbol{y} \in \mathbb{R}^4$. Jak zależy liczba różnych wag od głębokości sieci rozwiniętej?

93. Na czym polega mapowanie resztowe? Opisać jego działanie dla dwu warstw splotowych $6 \times 6 \times 30$ z filtrami 3×3.

94. W sieci splotowej (ze współdzieleniem wag i bez uzupełniania obrazu) do obrazu RGB 20×20 zastosowano

    a) 5 filtrów splotowych nieliniowych 1×1
    b) 7 filtrów splotowych nieliniowych 9×9
    c) 8 filtrów splotowych nieliniowych 20×20.

    Dla każdego przypadku: i) określić rozmiar warstwy wyjściowej, ii) określić liczbę parametrów (wag, obciążeń) jakie należy wyznaczyć,

    Wsk.: a): i) $20 \times 20 \times 5$, ii) $\cdot(1 \cdot 1 \cdot 3 + 1) \cdot 5 = 20$, b): i) $(14 \times 14 \times 7$, ii) $(9 \cdot 9 \cdot 3 + 1) \cdot 7 = 1708$, c): i) $1 \times 1 \times 10$, ii) $(20 \cdot 20 \cdot 3 + 1) \cdot 8 = 9608$,

95. Jakie powinny być wymiary i liczba filtrów splotowych zastosowanych do obrazu $32 \times 32 \times 3$, aby utworzyć warstwę 50 neuronów połączonych kompletnie z tym obrazem? Ile jest różnych parametrów (wag, obciążeń) tych filtrów? Odpowiedź uzasadnić.

    Wsk.: 50 filtrów $32 \times 32 \times 3$; $(32 \cdot 32 \cdot 3 + 1) \cdot 50 = 153650$ parametrów

96. Załóżmy, że wartości pikseli obrazu w skali szarości (jedna warstwa) o rozmiarze 4×4 są równe $x_{ij} = i + j$. Nie uwzględniając efektów brzegowych, obliczyć:

    a) splot obrazu z pięcioma filtrami 2×2 o dowolnie ustalonych wagach

    b) wynik agregacji (*max-pooling*) wyniku operacji a) z oknem 2×2 i krokiem 1;

    c) splot wyniku operacji (b) z jednym filtrem 1×1 o wagach $(-1, 1)$;

    d) rezultat operacji ReLU na wyniku operacji c)

    Wsk.: a) warstwa wynikowa (obliczyć) ma rozmiar $3 \times 3 \times 5$; b) wynikiem jest warstwa $2 \times 2 \times 5$; c) wynikiem jest warstwa $2 \times 2 \times 1$; d) rozmiar jak c);
    uwaga: dla rozstrzygnięcia, czy ewentualny nieprawidłowy wynik jest tylko efektem problemów arytmetycznych, pokazywać również wyniki pośrednie

97. Obraz RGB ma rozmiar 64×64. Nie uwzględniając efektów brzegowych, określić:

    a) rozmiar splotu tego obrazu z czterema filtrami 5×5

    b) rozmiar wynikający ze agregacji (*max-pooling*) wyniku operacji a) z oknem 3×3 i krokiem 1;

    c) rozmiar warstwy będącej splotu wyniku operacji b) z filtrem 1×1

    Wsk.: a) warstwa $60 \times 60 \times 4$, b) warstwa $58 \times 58 \times 4$, c) warstwa $58 \times 58 \times 1$

98. Omówić istotne cechy architektury sieci typu VGG możliwe przyczyny ich sukcesu

99. Omówić istotne cechy architektury sieci typu resnet i możliwe przyczyny ich sukcesu

100. Wyjaśnić strukturę i działanie sieci NiN

101. Omówić istotne cechy architektury sieci typu inception i możliwe przyczyny ich sukcesu

102. Czym różni się uczenie syjamskie i uczenie trójkami od uczenia wykorzystującego koszt pojedynczych obrazów

103. Dlaczego sieci GAN należą do sieci a) antagonistycznych b) generatywnych ?

104. Na czym polega różnica pomiędzy oceną krytyka a oceną dyskryminatora w sieciach typu GAN

105. Jaka jest różnica w architekturze i działaniu sieci DCGAN i sieci CGAN?

106. Wyjaśnić działanie sieci CycleGAN

107. Na czym polega różnica między stratą / wskaźnikiem kosztu sieci DCGAN i sieci WGAN?

108. Wyjaśnić działanie sieci WGAN