

Egzamin w czasach pandemii koronawirusa w sesji “zdalnej” lato 2019/2020:

- test jednokrotnego wyboru,
- 20 pytań (losowane z puli) na 12 lub 15 minut,
- zaliczenie egzaminu od 12 punktów - 60% (jedno pytanie = jeden punkt, brak punktów ujemnych za złe odpowiedzi).

Prowadzący: prof. UEK dr hab. inż. Janusz Morajda

Przedmiot:

- Inżynieria oprogramowania - “Programowanie systemów rozproszonych”
- Systemy informacyjne - “Metody eksploracji danych”
- Inna nazwa to “Programowanie systemów inteligentnych”

Good luck ;)

pc

1. Aglomeracyjne metody klasteryzacji polegają na:

- a. iteracyjnym dzieleniu klastrow na najbardziej odległe połowy
- b. przyrostowym budowaniu klastrow poprzez dodawanie sąsiednich punktów**
- c. krokowym dodawaniu linii dzielących klastry w optymalnym miejscu
- d. stopniowym ograniczaniu klastrow poprzez usuwanie outlierów

2. Pojęcie tensora występujące często w dziedzinie uczenia głębokiego oznacza:

- a. wyspecjalizowany rodzaj neuronów przetwarzających obraz
- b. warstwę wejściową rozprowadzającą sygnały do sieci
- c. wektor wskazujący kierunek najszybszego spadku błędu sieci
- d. uogólnienie pojęcia macierzy na wiele wymiarów**

3. Algorytmy genetyczne (ewolucyjne) służą generalnie do realizacji zadań:

- a. regresji
- b. eksploracji danych zapisanych w chromosomach roślin i zwierząt
- c. poszukiwania reguł asocjacyjnych
- d. optymalizacji**

4. Wskaż metodę NIEPRZYDATNĄ w rozwiązywaniu zadań predykcyjnych:

- a. algorytm k-średnich**
- b. sieć neuronowa typu RBF
- c. liniowa regresja wieloraka
- d. sieć neuronowa typu PERCEPTRON WIELOWARSTWOWY

5. Technika k-krotnej walidacji krzyżowej służy do:

- a. wyeliminowania wzajemnych zależności między zmiennymi
- b. uniknięcia tendencyjności w ocenie jakości modelu**
- c. klasteryzacji zbioru na k maksymalnie odległych skupień
- d. ustalenia optymalnego momentu przerywania uczenia modelu

6. Algorytm aglomeracyjny rozpoczyna proces analizy od stanu, w którym:

- a. obiekty podzielone są wstępnie przy użyciu innej metody grupowania
- b. wszystkie obiekty tworzą jedno skupienie
- c. obiekty są losowo przydzielone do skupień, a liczba skupień jest z góry określana
- d. każdy obiekt tworzy osobne skupienie**

7. Argumentem funkcji aktywacji neuronu typu RBF jest

- a. suma sygnałów $x_1 + x_2 + \dots$ wektora wejściowego x
- b. iloczyn skalarny wektorów: wejściowego x i wag w
- c. ważona liczba wejść danego neuronu
- d. odległość wektorów: wejściowego x i wag w

8. W pewnym problemie eksploracji danych zmienna wyjściowa (zależna) przyjmuje 3 możliwe wartości: "biały", "czerwony", "niebieski". Dany problem zaliczamy do zadań:

- a. klasyfikacji wzorcowej
- b. klasyfikacji bezwzorcowej (grupowania)
- c. poszukiwania reguł asocjacyjnych
- d. regresyjnych (szacowanie, predykcja)

9. Iloczyn skalarny wektora cech z pewnym wektorem wag można traktować jako formę:

- a. obrotu w przestrzeni cech
- b. redukcji wymiarowości
- c. regresji liniowej
- d. ekstrakcji cech

10. Końcowe trzy etapy eksploracji danych w metodologii CRISP-DM to KOLEJNO:

- a. Wdrożenie - Ewaluacja - Modelowanie
- b. Modelowanie - Ewaluacja - Wdrożenie
- c. Modelowanie - Wdrożenie - Ewaluacja
- d. Ewaluacja - Modelowanie - Wdrożenie

11. Istotną WADĄ sieci neuronowych jako modeli eksploracji danych jest:

- a. brak zdolności wyjaśniania (uzasadniania) wygenerowanych odpowiedzi
- b. nieparametryczność
- c. bardzo długi czas reakcji (odpowiedzi) nauczonej sieci
- d. brak zdolności do modelowania zjawisk nieliniowych

12. Hiperparametrami modelu nazywa się parametry, które:

- a. opisują wagi połączeń pomiędzy całymi warstwami sieci
- b. są dopasowywane podczas uczenia do danych wejściowych
- c. mogą zmieniać się w bardzo szerokim zakresie wartości
- d. sterują przebiegiem procesu uczenia modelu

13. Metoda wzmacniania gradientowego (gradient boosting) służy do:

- a. zwiększania kontrastu pomiędzy blisko leżącymi klastrami
- b. przyspieszania procesu uczenia głębokich sieci neuronowych
- c. budowania mocnego modelu złożonego z wielu słabszych modeli
- d. generowania dodatkowych sztucznych próbek w zbiorach uczących

14. Aby wyeliminować tzw. efekt przeuczenia sieci neuronowej (uczenie nadzorowane), należy uznać za optymalny (ostateczny) zbiór wartości wag, otrzymany w momencie:

- a. zakończenia działania algorytmu uczącego
- b. zidentyfikowania minimum lokalnego funkcji błędu uczącego
- c. minimum błędu dla ciągu uczącego
- d. minimum błędu dla ciągu walidacyjnego

15. Klasyczny (perceptronowy) model neuronu posiada dwa wejścia o wagach $w_1=1$, $w_2=2$, na które podano odpowiednio sygnały $x_1=-1$, $x_2=+1$. Funkcja aktywacji jest funkcją liniową postaci $y=2x$. Sygnał wyjściowy neuronu wynosi:

- a. 1
- b. 2**
- c. -1
- d. 0

16. Na etapie wstępnej eksploracyjnej analizy danych najmniej przydatne jest:

- a. wygenerowanie histogramów zmiennych ciągłych**
- b. użycie metod nienadzorowanej klasteryzacji
- c. wizualne porównanie współzależności par zmiennych
- d. zliczenie wystąpień wartości zmiennych nominalnych

17. Wskaż drugi algorytm (metodę), który służy do rozwiązywania tego samego typu problemów eksploracji danych, co algorytm k-najbliższych sąsiadów:

- a. algorytm PCA (analiza głównych składowych)
- b. algorytm CART (drzewa klasyfikacyjne i regresyjne)**
- c. algorytm k-średnich
- d. algorytm Kohonena (sieć neuronowa typu SOM)

18. Sieć neuronowa Kohonena (SOM) generalnie jest przeznaczona do realizacji:

- a. analizy szeregów czasowych
- b. prognozowania lub szacowania wartości
- c. klasyfikacji wzorcowej
- d. klasyfikacji bezwzorcowej (grupowania)**

19. Problem klasyfikacyjny można zamienić na problem regresyjny poprzez:

- a. zamianę zmiennych kategoriycznych na ciągłe**
- b. zmniejszenie liczby wymiarów do dwóch
- c. prognozowanie stopnia przynależności do klasy
- d. zredukowanie liczby klas do jednej

20. Poszukiwanie zbiorów częstych jest pierwszym etapem:

- a. algorytmu A priori**
- b. algorytmu k-średnich
- c. algorytmu wstecznej propagacji błędów
- d. algorytmu k-najbliższych sąsiadów

21. Algorytm aglomeracyjny rozpoczyna proces analizy od stanu, w którym

- a. obiekty podzielone są wstępnie przy użyciu innej metody grupowania
- b. wszystkie obiekty tworzą jedno skupienie
- c. obiekty są losowo przydzielone do skupień, a liczba skupień jest z góry określana
- d. każdy obiekt tworzy osobne skupienie - chyba to (liczba grup równa się liczbie obserwacji – każda obserwacja stanowi odrębną grupę.)**

22. Problem klasyfikacyjny można zamienić na problem regresyjny poprzez:

- a. zredukowanie liczby klas do jednej
- b. zmniejszenie liczby wymiarów do dwóch
- c. zamianę zmiennych kategoriycznych na ciągłe**
- d. prognozowanie stopnia przynależności do klasy

23.Która z poniższych metod NIE służy do wyznaczania odległości pomiędzy skupieniami:

- a. metoda średnich połączeń
- b. metoda najdalszego sąsiedztwa - chyba
- c. metoda pojedynczego wiązania
- d. metoda k-średnich**

24.Metoda K najbliższych sąsiadów wybiera sąsiadów analizowanego punktu:

- a. najbliższych pod względem cech wejściowych
- b. należących do tego samego klastra
- c. najbliższych pod względem prognozowanej wartości**
- d. leżących w tym samym wymiarze przestrzeni

25.Pojęcie tensora występujące często w dziedzinie uczenia głębokiego oznacza:

- a. wektor wskazujący kierunek najszybszego spadku błędu sieci
- b. wyspecjalizowany rodzaj neuronów przetwarzających obraz
- c. uogólnienie pojęcia macierzy na wiele wymiarów**
- d. warstwę wejściową rozprawdzającą sygnały do sieci

26.Końcowe trzy etapy eksploracji danych w metodologii CRISP-DM to KOLEJNO:

- a. Modelowanie - Wdrożenie - Ewaluacja
- b. Modelowanie - Ewaluacja - Wdrożenie**
- c. Wdrożenie - Ewaluacja - Modelowanie
- d. Ewaluacja - Modelowanie - Wdrożenie

27.Sieć neuronowa Kohonena (SOM) generalnie jest przeznaczona do realizacji:

- a. prognozowania lub szacowania wartości
- b. klasyfikacji wzorcowej
- c. analizy szeregów czasowych
- d. klasyfikacji bezwzorcowej (grupowania)**

28.Aglomeracyjne metody klasteryzacji polegają na:

- a. przyrostowym budowaniu klastrów poprzez dodawanie sąsiednich punktów**
- b. krokowym dodawaniu linii dzielących klastry w optymalnym miejscu
- c. iteracyjnym dzieleniu klastrów na najbardziej odległe połowy
- d. stopniowym ograniczaniu klastrów poprzez usuwanie outlierów

29.Wskaż metodę NIEPRZYDATNĄ w rozwiązywaniu zadań predykcyjnych:

- a. sieć neuronowa typu RBF
- b. sieć neuronowa typu PERCEPTRON WIELOWARSTWOWY
- c. algorytm k-średnich**
- d. liniowa regresja wieloraka

30.Poszukiwanie zbiorów częstych jest pierwszym etapem:

- a. algorytmu k-najbliższych sąsiadów
- b. algorytmu k-średnich
- c. algorytmu A priori - chyba to(Algorytm A priori po prostu znajduje częste zbiory zdarzeń)**

31.Wskaż drugi algorytm (metodę), który służy do rozwiązywania tego samego typu problemów eksploracji danych, co algorytm k-najbliższych sąsiadów:

- a. algorytm PCA (analiza głównych składowych)

- b. algorytm k-średnich
- c. algorytm CART (drzewa klasyfikacyjne i regresyjne)**
- d. algorytm Kohonena (sieć neuronowa typu SOM)

32. Problem klasteryzacyjny polega na:

- a. poszukiwaniu granic oddzielających obserwacje różnych klas
- b. identyfikacji skupisk zgodnie z pewnym kryterium podobieństwa**
- c. predykcji przynależności danej obserwacji do różnych klas
- d. grupowaniu cech obserwacji w skorelowane ze sobą zespoły
- d. algorytmu wstecznej propagacji błędów

33. Metoda wzmacniania gradientowego (gradient boosting) służy do:

- a. przyspieszania procesu uczenia głębokich sieci neuronowych
- b. budowania mocnego modelu złożonego z wielu słabszych modeli (Rozwiązania z dziedziny boosting łączą wiele "słabych", prostych rozwiązań w jedno większe)**
- c. generowania dodatkowych sztucznych próbek w zbiorach uczących
- d. zwiększania kontrastu pomiędzy blisko leżącymi klastrami

34. Iloczyn skalarny wektora cech z pewnym wektorem wag można traktować jako formę:

- a. regresji liniowej**
- b. ekstrakcji cech
- c. obrotu w przestrzeni cech
- d. redukcji wymiarowości

35. Na etapie wstępnej eksploracyjnej analizy danych najmniej przydatne jest:

- a. użycie metod nienadzorowanej klasteryzacji
- b. zliczenie wystąpień wartości zmiennych nominalnych [szczelone w to]
- c. wygenerowanie histogramów zmiennych ciągłych**
- d. wizualne porównanie współzależności par zmiennych

36. Przeuczenie modelu (overfitting) można rozpoznać po tym, że:

- a. liczba błędnie zaklasyfikowanych przypadków spada do 0
- b. w procesie uczenia zaczyna rosnąć błąd dla próby testowej - czyha? [ale przeuczenie można rozpoznać po zakończonym uczeniu]**
- c. trafność predykcji modelu przekracza poziom 99.73%
- d. szybkość uczenia się modelu zaczyna spadać w kolejnych iteracjach

37. Argumentem funkcji aktywacji neuronu typu RBF jest

- a. ważona liczba wejść danego neuronu
- b. odległość wektorów: wejściowego x i wag w [raczej to]**
- c. suma sygnałów $x_1 + x_2 + \dots$ wektora wejściowego x
- d. iloczyn skalarny wektorów: wejściowego x i wag w

38. Oryginalna zmienna x przyjmuje trzy wartości: -2, 1, 4. Po przeprowadzeniu normalizacji tej zmiennej wg metody min-max do przedziału [0, 1], oryginalnej wartości 1 odpowiada znormalizowana wartość:

- a. 0,25
- b. 0,5**
- c. 0,75
- d. 1

39. Problem klasyfikacyjny polega na:

- a. ustaleniu optymalnej liczby klas, do których należą obserwacje
- b. poszukiwaniu przypadków najbardziej reprezentacyjnych dla klas
- c. rozróżnianiu obserwowanych cech na wejściowe i wyjściowe
- d. **prognozowaniu kategorii obserwacji na podstawie jej cech**

40. Technika k-krotnej walidacji krzyżowej służy do:

- a. klasteryzacji zbioru na k maksymalnie odległych skupień
- b. wyeliminowania wzajemnych zależności między zmiennymi
- c. ustalenia optymalnego momentu przerywania uczenia modelu
- d. **uniknięcia tendencyjności w ocenie jakości modelu**

41. Rolą pojedynczego neuronu w warstwie perceptronu użytego do klasyfikacji jest:

- a. **podział przestrzeni wejść na dwie półprzestrzenie**
- b. rozpoznawanie jednej ze znanych klas w zbiorze
- c. rozpoznawanie jednego przypadku w zbiorze
- d. klasyfikacja pojedynczej cechy wejściowej

42. Algorytmy genetyczne (ewolucyjne) służą generalnie do realizacji zadań:

- a. poszukiwania reguł asocjacyjnych
- b. eksploracji danych zapisanych w chromosomach roślin i zwierząt
- c. **optymalizacji [raczej to]**
- d. regresji

43. Przyjęcie metryki L1 (metryki Manhattan) sprawia, że okręgi przyjmują kształt:

- a. trójkątów
- b. kardiod
- c. elips
- d. **rombów**

44. Przycinanie drzew decyzyjnych stosuje się w celu:

- a. eliminacji zmiennych nieistotnych
- b. **uniknięcia przeuczenia i skomplikowania modelu (?) tak**
- c. redukcji liczby klas
- d. ponownego przeprowadzenia podziału w miejscu cięcia

45. Problem regresyjny polega na:

- a. dopasowaniu współczynników modelu do obserwacji uczących
- b. prognozowaniu wartości na podstawie obserwowanych cech
- c. **poszukiwaniu korelacji między zmiennymi wejściowymi**
- d. znajdowaniu prostej najlepiej dopasowanej do obserwacji

46. Typowy zbiór danych używany do trenowania modeli uczenia maszynowego składa się z:

- a. instancji zgrupowanych w kategorie
- b. **obserwacji obejmujących cechy**
- c. obiektów posiadających atrybuty
- d. klas zawierających instancje

47. Hiperparametrami modelu nazywa się parametry, które:

- a. są dopasowywane podczas uczenia do danych wejściowych
- b. mogą zmieniać się w bardzo szerokim zakresie wartości

c. sterują przebiegiem procesu uczenia modelu

d. opisują wagi połączeń pomiędzy całymi warstwami sieci

48. W pewnym problemie eksploracji danych zmienna wyjściowa (zależna) przyjmuje 3 możliwe wartości: "biały", "czerwony", "niebieski". Dany problem zaliczamy do zadań:

a. klasyfikacji bezwzorcowej (grupowania)

b. regresyjnych (szacowanie, predykcja)

c. poszukiwania reguł asocjacyjnych

d. klasyfikacji wzorcowej

49. Istotną WADĄ sieci neuronowych jako modeli eksploracji danych jest:

a. brak zdolności wyjaśniania (uzasadniania) wygenerowanych odpowiedzi

b. bardzo długi czas reakcji (odpowiedzi) nauczanej sieci

c. nieparametryczność

d. brak zdolności do modelowania zjawisk nieliniowych

50. Aby wyeliminować tzw. efekt przeuczenia sieci neuronowej (uczenie nadzorowane), należy uznać za optymalny (ostateczny) zbiór wartości wag, otrzymany w momencie:

a. zakończenia działania algorytmu uczącego

b. minimum błędu dla ciągu walidacyjnego

c. zidentyfikowania minimum lokalnego funkcji błędu uczącego

d. minimum błędu dla ciągu uczącego