

Odpowiedzi do każdej z części na oddzielnych kartkach (4 różne kartki)

### Część RB

RB. Zad. 1. (8 p.) Mamy środowisko:

0	0	0	0	0	1	W poszczególnych stanach podano
0	0	0	0	0	0	nagrody. Stany o
0	0	0	0	0	0	nagrodzie > 0 są
0	0	0	0	0	0	stanami
0	0	0	0	0	0	absorbującymi. $\beta =$
0	0	0	0	0	0	0,9; $\gamma = 1$ . Stany
0	0	0	0	0	0	będziemy oznaczać
0	0	0	0	0	0	0,5 podając kolejno
						numer wiersza i

kolumny, licząc od zera, zaczynając od górnego lewego rogu. Jak zmieniają się komórki tablicy Q po wykonaniu epizodów przez algorytm SARSA (jeśli w epizodzie nie zmienia się to proszę wpisać „brak zmian”):

Epizod 1) Start ze stanu: 22, wykonane akcje: LRLURUUURU.

Epizod 2) Start ze stanu: 25, wykonane akcje: ULLU.

Epizod 3) Start ze stanu: 35, wykonane akcje: UURU

RB. Zad. 2. (1,5 p.) W algorytmie ewolucyjnym zastosowano krzyżowanie równomierne. Do krzyżowania wybrano osobniki:  $x_1 = [0,1; 0,4; 0,9; 0,3]$  oraz  $x_2 = [0,8; 0,5; 0,7; 0,4]$ . Jaki jest wynik krzyżowania jeśli wynikiem losowania  $U(0; 1) < 0,5$  było [True; True; False; False]?

RB. Zad. 3. (3 p.) Czy strategia ewolucyjna ES(1+1) z siłą mutacji  $\delta=0,5$  uruchomiona z punktem startowym będącym optimum lokalnym funkcji gálara ma w praktyce szansę na znalezienie optimum globalnego? Odpowiedź krótko uzasadnić.

### Część PZ

PZ.1. Mamy 1000-elementowy zbiór danych do klasyfikacji, w którym  $y \in \{A, B, C, D\}$ , a atrybuty wejściowe to  $x_1, \dots, x_m$ . Zadanie próbujesz rozwiązać indukując drzewo za pomocą metody lasu losowego (random forest).

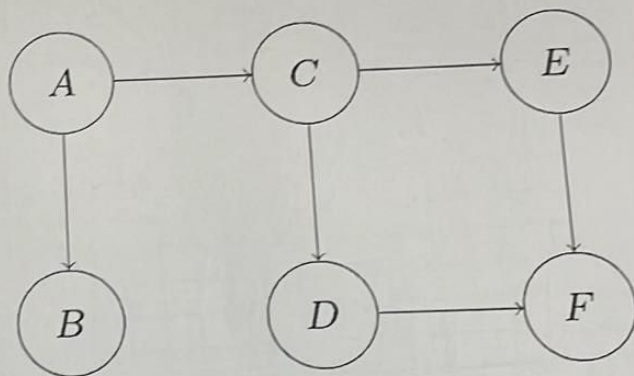
1.1 (2 pkt) Uzasadnij jaka największa liczba drzew może być w lesie?

1.2 (2 pkt) Jakie hiperparametry (czyli parametry algorytmu uczącego) mogą wymagać strojenia w przypadku lasu losowego?

1.3 (1 pkt) Kiedy drzewo będzie mieć silniejsze tendencje do przeuczania się podczas uczenia – gdy jest wyższe, czy niższe? (uzasadnij)

1.4 (0.5 pkt) Od czego zależy maksymalna wysokość pojedynczego drzewa w lesie?

PZ.2. Dany jest graf przedstawiający sieć Bayesa.



2.1 (2 pkt) Uzasadnij, czy  $P(D|E \wedge C) = P(D|C)$ ?

2.2 (2 pkt) Jaki zbiór węzłów zapewni d-separację zmiennych E oraz B? (uzasadnij)

2.3 (1 pkt) Z jakiego rodzaju wnioskowaniem mamy do czynienia odpowiadając na pytanie A na podstawie danych C?

2.4 (2 pkt) Jeżeli dla powyższego grafu uruchomiony zostanie algorytm MCMC z próbkowaniem Gibbsa oraz dowodami A, D, E, to jakie jest p-stwo, że po 6 iteracjach stan E nie był ani razu zmieniany? (uzasadnij)

### Część KR 7 pkt

**KR. Zad. 1 (8 pkt)** Wymień i opisz sposoby zapobiegania przeuczeniu w sieciach neuronowych

**KR. Zad. 2 (5.5 pkt)**

Masz dwa zestawy danych o podobnej wielkości dla zadania klasyfikacji binarnej. Jednak jeden z nich zawiera prawie wyłącznie pozytywne przykłady, a drugi zawiera tylko negatywne przykłady. Chciałbyś użyć obu zestawów do trenowania modelu. Opisz scenariusz, w którym połączenie tych dwóch zestawów danych mogłoby doprowadzić do niepoprawnego uczenia się modelu.

---

### Część JA (12,5 p.)

W klasycznym zadaniu plecakowym zakładamy, że przedmioty są niepodzielne, a każdy z nich ma dodatnią wartość korzyści  $p_i$  i wagi  $w_i$ . Poszukujemy zestawu przedmiotów który maksymalizuje łączną korzyść, przy założeniu że suma wag nie przekroczy limitu  $W$ .

Rozważmy nieco bardziej złożony problem, w którym zdefiniowane są dwa dodatkowe ograniczenia. Przedmioty o numerach parzystych muszą mieć łączną wagę nie większą niż  $W_1$ , a przedmioty o numerach nieparzystych – nie większą wagę niż  $W_2$ . Łączna waga przedmiotów nie może przekroczyć  $W$ , przy czym  $W \leq W_1 + W_2$ ; wartości wszystkich ograniczeń wagowych są dodatnie.

Zakładając, że zostanie użyty algorytm  $A^*$ , proszę zaproponować:

- sposób reprezentacji rozwiązania,
- funkcję zysku
- funkcję heurystyczną.

Funkcja heurystyczna powinna się charakteryzować niewielkim nakładem obliczeń, powinna być dopuszczalna i monotoniczna.

Odpowiedź proszę uzasadnić, podając uzasadnienie monotoniczności i dopuszczalności. Funkcja heurystyczna, której wartość po zsumowaniu z zyskiem będzie jednakowa dla każdego węzła przestrzeni przeszukiwań, nie będzie uznana za odpowiedź prawidłową.

---

Szanowni Państwo,

poniżej prawidłowe odpowiedzi do moich zadań z terminu 2 egzaminu: (pytań nie udostępniam)

Zad. 1.

Ep. 1) brak zmian

Ep. 2)  $Q(15, \wedge) = 0,9$

Ep. 3)  $Q(25, \wedge) 0 + 0,9(0 + 1 \cdot 0,9 - 0) = 0,81$

$Q(15, \wedge) 0,9 + 0,9 \cdot (1 + 1 \cdot 0 - 0,9) = 0,99$

Zad. 2.

0,1; 0,4; 0,7; 0,4; uznam również: 0,8; 0,5; 0,9; 0,3

Zad 3

Nie, w przedziale  $[\mu - 3\sigma, \mu + 3\sigma]$  znajduje się 99,7 % wszystkich obserwacji, a mamy generator udający losowość. Aby znalazła musiałaby strzelić w okolicę 0 czyli na odległość około 7 czyli na odległość około  $14\sigma$ . Co gorsza  $\sigma$  będzie maleć co 'a' iteracji.

Pozdrawiam

Rafał Biedrzycki

1. Mamy 1000-elementowy zbiór danych do klasyfikacji, w którym  $y \in \{A, B, C, D\}$ , a atrybuty wejściowe to  $x_1, \dots, x_m$ . Zadanie próbujesz rozwiązać indukując drzewo za pomocą metody lasu losowego (random forest).

1. (2 pkt) Uzasadnij jaka największa liczba drzew może być w lesie?

- Drzewa tworzone są niezależnie od siebie, więc może ich być w zasadzie dowolna liczba (choć od pewnego poziomu nie będzie poprawiać się model)

2. (2 pkt) Jakie hiperparametry (czyli parametry algorytmu uczącego) mogą wymagać strojenia w przypadku lasu losowego?

- Np. liczba drzew, liczba atrybutów próbkowanych dla każdego drzewa, wielkość poszczególnych podzbiorów, maksymalna wysokość drzewa.

3. (1 pkt) Kiedy drzewo będzie mieć silniejsze tendencje do przeuczania się podczas uczenia – gdy jest wyższe, czy niższe? (uzasadnij)

- Gdy jest wyższe – wtedy będzie mogło być bardziej rozbudowane i może wymagać przycinania.

4. (0.5 pkt) Od czego zależy maksymalna wysokość pojedynczego drzewa w lesie?

- Najważniejsze ograniczenie wynika z liczby atrybutów

2. Dany jest graf przedstawiający sieć Bayesa.

1. (2 pkt) Uzasadnij, czy  $P(D|E \wedge C) = P(D|C)$ ?

- Prawda. zapis oznacza warunkową niezależność D i E przy zaobserwowaniu C – i faktycznie, zaobserwowanie C przecina jedyną otwartą ścieżkę inf.  $D \leftarrow C \rightarrow E$

2. (2 pkt) Jaki zbiór węzłów zapewni d-separację zmiennych E oraz B? (uzasadnij)

-  $\{A\}$  lub  $\{C\}$  – mamy dwie ścieżki:

$B \leftarrow A \rightarrow C \rightarrow E$  - blokują ją  $\{A\}$  lub  $\{C\}$ ,

$B \leftarrow A \rightarrow C \rightarrow D \rightarrow F \leftarrow E$  (ta zablokowana przez F)

3. (1 pkt) Z jakiego rodzaju wnioskowaniem mamy do czynienia odpowiadając na pytanie A na podstawie danych C?

- Stan rodziców na podstawie dzieci  $\rightarrow$  diagnostyczne

4. (2 pkt) Jeżeli dla powyższego grafu uruchomiony zostanie algorytm MCMC z próbkowaniem Gibbsa oraz dowodami A,D,E, to jakie jest p-stwo, że po 6 iteracjach stan E nie był ani razu zmieniany? (uzasadnij)

- Stan E nie będzie zmieniany nigdy, bo to dowód

