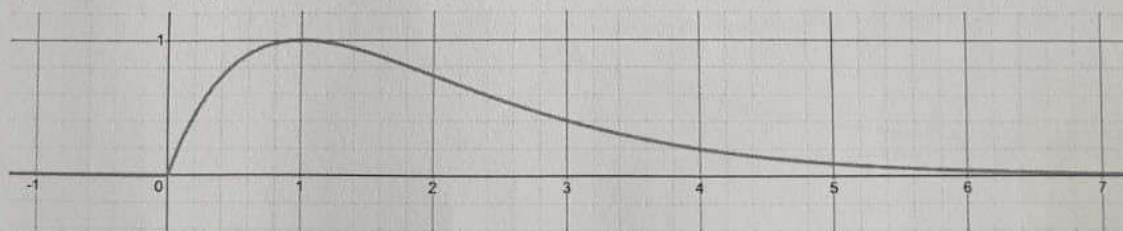
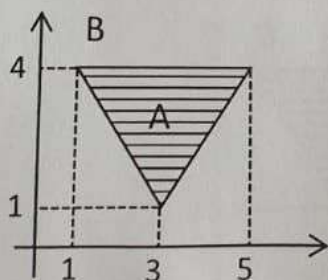


Oficiu (0.5p)

1. (0.25p) Dacă un nod dintr-un graf computațional reprezintă operația  $z = 2xy$ , intrările sunt  $x = -0.5$  și  $y = 2$ , iar gradientul  $\partial L / \partial z = -5$ , atunci gradientii în raport cu intrările  $\partial L / \partial x$  și  $\partial L / \partial y$  sunt:  
 A.  $\partial L / \partial x = -20$  și  $\partial L / \partial y = 5$       B.  $\partial L / \partial x = -40$  și  $\partial L / \partial y = 10$   
 C.  $\partial L / \partial x = 20$  și  $\partial L / \partial y = -5$       D.  $\partial L / \partial x = 40$  și  $\partial L / \partial y = -10$
2. (0.25p) Care este scufundarea asociată funcției nucleu RBF?  
 A.  $f(x) = x$       B.  $f(x) = \sqrt{x}$   
 C.  $k(x, y) = e^{-\frac{\Delta(x, y)}{2\sigma}}$ , unde  $\Delta(x, y)$  este distanța Euclidiană      D. Nu există
3. (0.25p) Cum se descompune eroarea unui model de învățare automată?  
 A. Eroare de modelare, estimare și optimizare      B. Eroare de modelare, învățare și testare  
 C. Eroare de modelare, optimizare și generalizare      D. Eroare de modelare, optimizare și testare
4. (0.25p) La ce se referă teorema „no free lunch”?  
 A. Lansarea procesului de antrenare este costisitoare      B. Rețelele neuronale aproximează orice funcție  
 C. Nu există un singur model bun pe toate task-urile      D. Rețele neuronale sunt în general soluția optimă
5. (0.25p) Când ar putea să apară fenomenul Hughes?  
 A. Când avem prea multe exemple      B. Când avem prea multe trăsături  
 C. Când algoritmul de optimizare nu converge      D. Când avem un model cu capacitate de modelare prea mică
6. (0.25p) Care este numărul de parametri al unui strat convoluțional cu 30 de filtre de dimensiune 5x5 aplicate cu un stride de 2 pe un tensor de dimensiune 64x64x10?  
 A. 7510      B. 7530      C. 750      D. 780
7. (0.25p) Ce determină capacitatea de modelare neliniară a rețelelor neuronale?  
 A. Matricile de ponderi      B. Funcția de pierdere      C. Funcțiile de transfer      D. Ponderile de tip bias
8. (0.25p) Care este dimensiunea tensorului de ieșire al unui strat de pooling cu filtre de 2x2 aplicate la stride de 4 pe un tensor de dimensiune 66x66x40?  
 A. 16x16x40      B. 31x31x40      C. 17x17x40      D. 32x32x40
9. (1p) Considerând funcția kernel  $k(x, y) = 2PQ$ , unde  $P = |\{(i, j): 1 \leq i < j \leq n, (x_i - x_j) \cdot (y_i - y_j) \geq 0\}|$  și  $Q = |\{(i, j): 1 \leq i < j \leq n, (x_i - x_j) \cdot (y_i - y_j) < 0\}|$ , definiți o funcție de scufundare  $\phi$ , care aplicată lui  $x$  și  $y$ , produce:  $k(x, y) = 2PQ = \langle \phi(x), \phi(y) \rangle$ . Exemplificați aplicarea funcției de scufundare pe  $x = (0, 4, 1, 2)$  și  $y = (1, 3, 5, 1)$ , demonstrând egalitatea  $2PQ = \langle \phi(x), \phi(y) \rangle$  pe acest exemplu.
10. (0.5p) Fiind dată funcția de activare  $y = f(x) = \max(0, x) \cdot \exp(-x + 1)$  cu graficul de mai jos, precizați dacă un singur neuron artificial bazat pe această funcție de activare ar putea rezolva problema XOR. Justificați răspunsul printr-un exemplu.



11. (1p) Configurați o rețea neuronală (specificând arhitectura și valorile parametrilor) care să discrimineze între mulțimile de puncte A (eticheta 1) și B (eticheta -1) reprezentate în figura de mai jos.





## Examen – Varianta 4

Oficiu (0.5p)

1. (0.25p) Care este scufundarea asociată funcției nucleu Hellinger?

- A.  $f(x) = x$  B.  $f(x) = \sqrt{x}$   
 C.  $k(x, y) = \langle \sqrt{x}, \sqrt{y} \rangle$ , unde  $\langle \blacksquare, \blacksquare \rangle$  este produsul scalar D. Nu există

2. (0.25p) Ce determină capacitatea de modelare neliniară a rețelelor neuronale?

- A. Funcțiile de activare B. Ponderile de tip bias C. Matricile de ponderi D. Funcția de pierdere

3. (0.25p) Care este dimensiunea tensorului de ieșire al unui strat convoluțional cu 20 filtre de  $3 \times 3$  aplicate la stride de 2 pe un tensor de dimensiune  $65 \times 65 \times 40$ ?

- A.  $32 \times 32 \times 20$  B.  $31 \times 31 \times 20$  C.  $16 \times 16 \times 40$  D.  $31 \times 31 \times 40$

4. (0.25p) La ce se referă teorema de aproximare universală?

- A. Orice model de învățare poate aproxima orice funcție B. Parametrii unei rețele se pot aproxima din date  
 C. Hiperparametrii se pot aproxima pe validare D. Rețelele neuronale pot aproxima orice funcție

5. (0.25p) Eroarea de modelare provine din alegerea:

- A. unui spațiu de ipoteze prea restrâns B. unui model cu capacitate prea mare de modelare  
 C. unui algoritm de optimizare nepotrivit D. unui spațiu de ipoteze circular

6. (0.25p) Care este numărul de parametri al unui strat convoluțional cu 40 de filtre de dimensiune  $3 \times 3$  aplicate cu un stride de 3 pe un tensor de dimensiune  $128 \times 128 \times 20$ ?

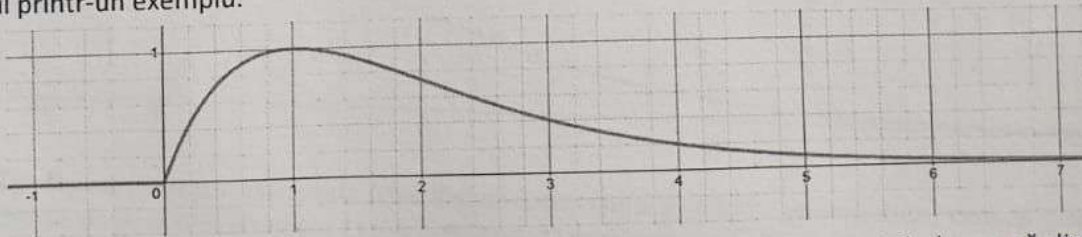
- A. 400 B. 360 C. 7240 D. 7220

7. (0.25p) Când ar putea să apară fenomenul Hughes?

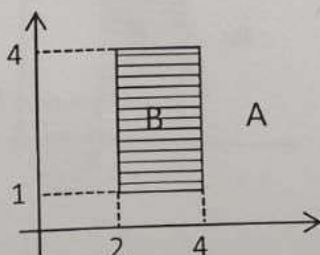
- A. Când algoritmul de optimizare nu converge B. Când avem prea puține exemple  
 C. Când avem un model cu capacitate de modelare prea mică D. Când avem prea puține trăsături

8. (0.25p) Dacă un nod dintr-un graf computațional reprezintă operația  $z = 2xy^2$ , intrările sunt  $x = -0.5$  și  $y = 2$ , iar gradientul  $\partial L / \partial z = -5$ , atunci gradientii în raport cu intrările  $\partial L / \partial x$  și  $\partial L / \partial y$  sunt:

- A.  $\partial L / \partial x = 20$  și  $\partial L / \partial y = -5$  B.  $\partial L / \partial x = -20$  și  $\partial L / \partial y = 5$   
 C.  $\partial L / \partial x = -40$  și  $\partial L / \partial y = 10$  D.  $\partial L / \partial x = 40$  și  $\partial L / \partial y = -10$

9. (1p) Considerând funcția kernel  $k(x, y) = \sum_i \min\{x_i, y_i\}$ , unde  $x$  și  $y$  sunt histogramme cu frecvența de apariție a unor cuvinte, având frecvența maximă egală cu 5, definiți o funcție de scufundare  $\phi$ , care aplicată lui  $x$  și  $y$ , produce:  $k(x, y) = \sum_i \min\{x_i, y_i\} = \langle \phi(x), \phi(y) \rangle$ . Exemplificați aplicarea funcției de scufundare pe  $x = (4, 0, 3, 2)$  și  $y = (3, 2, 5, 4)$ , demonstrând egalitatea  $\sum_i \min\{x_i, y_i\} = \langle \phi(x), \phi(y) \rangle$  pe acest exemplu.10. (0.5p) Fiind dată funcția de activare  $y = f(x) = \max(0, x) \cdot \exp(-x + 1)$  cu graficul de mai jos, precizați dacă un singur neuron artificial bazat pe această funcție de activare ar putea rezolva problema XOR. Justificați răspunsul printr-un exemplu.

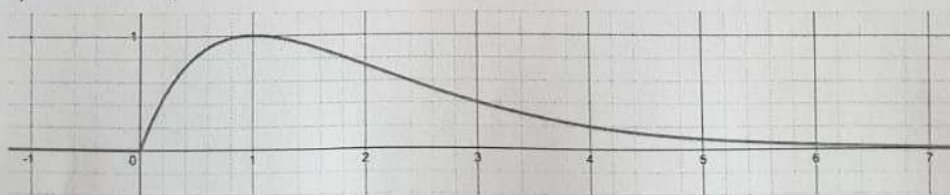
11. (1p) Configurați o rețea neuronală (specificând arhitectura și valorile parametrilor) care să discrimineze între mulțimile de puncte A (eticheta 1) și B (eticheta -1) reprezentate în figura de mai jos.



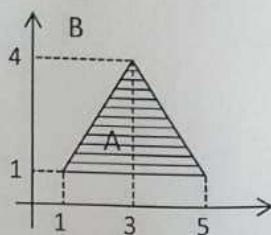


Oficiu (0.5p)

- (0.25p) Cum se descompune eroarea unui model de învățare automată?
  - Eroare de modelare, optimizare și testare
  - Eroare de modelare, estimare și optimizare
  - Eroare de modelare, învățare și testare
  - Eroare de modelare, optimizare și generalizare
- (0.25p) Când ar putea să apară fenomenul Hughes?
  - Când avem prea multe exemple
  - Când algoritmul de optimizare nu converge
  - Când avem prea multe trăsături
  - Când avem un model cu capacitate de modelare prea mică
- (0.25p) Care este numărul de parametri al unui strat convoluțional cu 20 de filtre de dimensiune 3x3 aplicate cu un stride de 3 pe un tensor de dimensiune 128x128x40?
  - 200
  - 180
  - 7240
  - 7220
- (0.25p) Care este scufundarea asociată funcției nucleu RBF?
  - Nu există
  - $f(x) = x$
  - $k(x, y) = e^{-\frac{\Delta(x, y)}{2\sigma}}$ , unde  $\Delta(x, y)$  este distanța Euclidiană
  - $f(x) = \sqrt{x}$
- (0.25p) Ce determină capacitatea de modelare neliniară a rețelelor neuronale?
  - Matricile de ponderi
  - Funcțiile de transfer
  - Funcția de pierdere
  - Ponderile de tip bias
- (0.25p) Dacă un nod dintr-un graf computațional reprezintă operația  $z = 2xy$ , intrările sunt  $x = -0.5$  și  $y = 2$ , iar gradientul  $\partial L / \partial z = 5$ , atunci gradientii în raport cu intrările  $\partial L / \partial x$  și  $\partial L / \partial y$  sunt:
  - $\partial L / \partial x = -40$  și  $\partial L / \partial y = 10$
  - $\partial L / \partial x = 40$  și  $\partial L / \partial y = -10$
  - $\partial L / \partial x = 20$  și  $\partial L / \partial y = -5$
  - $\partial L / \partial x = -20$  și  $\partial L / \partial y = 5$
- (0.25p) Care este dimensiunea tensorului de ieșire al unui strat convoluțional cu 40 filtre de 3x3 aplicate la stride de 2 pe un tensor de dimensiune 65x65x20?
  - 16x16x20
  - 31x31x20
  - 32x32x40
  - 31x31x40
- (0.25p) La ce se referă teorema „no free lunch”?
  - Rețele neuronale sunt în general soluția optimă
  - Lansarea procesului de antrenare este costisitoare
  - Rețelele neuronale aproximează orice funcție
  - Nu există un singur model bun pe toate task-urile
- (1p) Considerând funcția kernel  $k(x, y) = 2PQ$ , unde  $P = |\{(i, j): 1 \leq i < j \leq n, (x_i - x_j) \cdot (y_i - y_j) \geq 0\}|$  și  $Q = |\{(i, j): 1 \leq i < j \leq n, (x_i - x_j) \cdot (y_i - y_j) < 0\}|$ , definiți o funcție de scufundare  $\phi$ , care aplicată lui  $x$  și  $y$ , produce:  $k(x, y) = 2PQ = \langle \phi(x), \phi(y) \rangle$ . Exemplificați aplicarea funcției de scufundare pe  $x = (5, 2, 1, 0)$  și  $y = (1, 3, 5, 1)$ , demonstrând egalitatea  $2PQ = \langle \phi(x), \phi(y) \rangle$  pe acest exemplu.
- (0.5p) Fiind dată funcția de activare  $y = f(x) = \max(0, x) \cdot \exp(-x + 1)$  cu graficul de mai jos, precizați dacă un singur neuron artificial bazat pe această funcție de activare ar putea rezolva problema XOR. Justificați răspunsul printr-un exemplu.



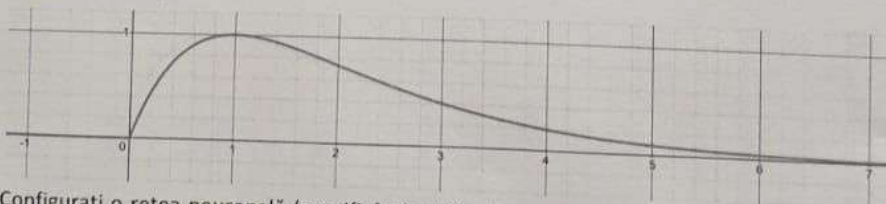
- (1p) Configurați o rețea neuronală (specificând arhitectura și valorile parametrilor) care să discrimineze între mulțimile de puncte A (eticheta 1) și B (eticheta -1) reprezentate în figura de mai jos.



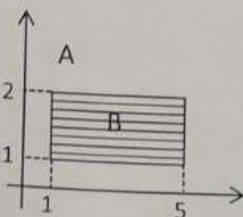


Oficiu (0.5p)

1. (0.25p) Eroarea de modelare provine din alegerea:
  - A. unui spațiu de ipoteze prea larg
  - B. unui model cu capacitate slabă de modelare
  - C. unui algoritm de optimizare nepotrivit
  - D. unui spațiu de ipoteze circular
2. (0.25p) La ce se referă teorema de aproximare universală?
  - A. Orice model de învățare poate aproxima orice funcție
  - B. Parametrii unei rețele se pot aproxima din date
  - C. Rețelele neuronale pot aproxima orice funcție
  - D. Hiperparametrii se pot aproxima pe validare
3. (0.25p) Dacă un nod dintr-un graf computațional reprezintă operația  $z = 2xy^2$ , intrările sunt  $x = -0.5$  și  $y = 2$ , iar gradientul  $\partial L / \partial z = -5$ , atunci gradientii în raport cu intrările  $\partial L / \partial x$  și  $\partial L / \partial y$  sunt:
  - A.  $\partial L / \partial x = 20$  și  $\partial L / \partial y = -5$
  - B.  $\partial L / \partial x = -40$  și  $\partial L / \partial y = 10$
  - C.  $\partial L / \partial x = -20$  și  $\partial L / \partial y = 5$
  - D.  $\partial L / \partial x = 40$  și  $\partial L / \partial y = -10$
4. (0.25p) Care este numărul de parametri al unui strat convoluțional cu 10 de filtre de dimensiune 5x5 aplicate cu un stride de 2 pe un tensor de dimensiune 64x64x30?
  - A. 750
  - B. 7530
  - C. 7510
  - D. 760
5. (0.25p) Ce determină capacitatea de modelare neliniară a rețelelor neuronale?
  - A. Funcția de pierdere
  - B. Ponderile de tip bias
  - C. Matricile de ponderi
  - D. Funcțiile de activare
6. (0.25p) Care este dimensiunea tensorului de ieșire al unui strat de pooling cu filtre de 4x4 aplicate la stride de 2 pe un tensor de dimensiune 66x66x40?
  - A. 32x32x40
  - B. 31x31x40
  - C. 16x16x40
  - D. 17x17x40
7. (0.25p) Care este scufundarea asociată funcției nucleu Hellinger?
  - A.  $f(x) = \sqrt{x}$
  - B. Nu există
  - C.  $k(x, y) = \langle \sqrt{x}, \sqrt{y} \rangle$ , unde  $\langle \cdot, \cdot \rangle$  este produsul scalar
  - D.  $f(x) = x$
8. (0.25p) Când ar putea să apară fenomenul Hughes?
  - A. Când avem prea puține exemple
  - B. Când algoritmul de optimizare nu converge
  - C. Când avem prea puține trăsături
  - D. Când avem un model cu capacitate de modelare prea mică
9. (1p) Considerând funcția kernel  $k(x, y) = \sum_i \min\{x_i, y_i\}$ , unde  $x$  și  $y$  sunt histograme cu frecvența de apariție a unor cuvinte, având frecvența maximă egală cu 5, definiți o funcție de scufundare  $\phi$ , care aplicată lui  $x$  și  $y$ , produce:  $k(x, y) = \sum_i \min\{x_i, y_i\} = \langle \phi(x), \phi(y) \rangle$ . Exemplificați aplicarea funcției de scufundare pe  $x = (0, 4, 1, 2)$  și  $y = (3, 2, 5, 3)$ , demonstrând egalitatea  $\sum_i \min\{x_i, y_i\} = \langle \phi(x), \phi(y) \rangle$  pe acest exemplu.
10. (0.5p) Fiind dată funcția de activare  $y = f(x) = \max(0, x) \cdot \exp(-x + 1)$  cu graficul de mai jos, precizați dacă un singur neuron artificial bazat pe această funcție de activare ar putea rezolva problema XOR. Justificați răspunsul printr-un exemplu.



11. (1p) Configurați o rețea neuronală (specificând arhitectura și valorile parametrilor) care să discrimineze între mulțimile de puncte A (eticheta 1) și B (eticheta -1) reprezentate în figura de mai jos.



11. (1p) Configurați o rețea neuronală (specificând arhitectura și valorile parametrilor) care să discrimineze între mulțimile de puncte A (eticheta 1) și B (eticheta -1) reprezentate în figura de mai jos.