

1. Eroarea de generalizare este aproximată în practică prin:
  - a. Eroarea de estimare
  - b. Eroarea optimă
  - c. Eroarea de optimizare
  - d. Eroarea empirică**

## Erori

- Eroarea de generalizare (generalization error):

$$\mathcal{E}(h) = \int_{X \times Y} V(h(x), y) \rho(x, y) dx dy$$

- Probabilitatea comună  $\rho(x, y)$  este deobicei necunoscută
- Atunci calculăm eroare empirică (empirical error):

$$E(h) = \frac{1}{n} \sum_{i=1}^n V(h(x_i), y_i)$$

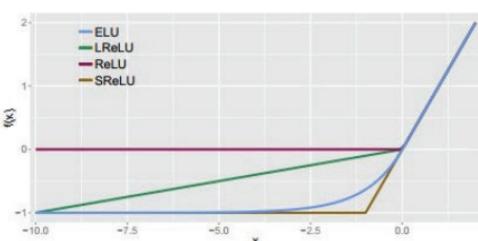
- Estimăm eroarea empirică pe datele de antrenare sau pe cele de test?
- **Nu este corect să raportăm eroarea pe datele de antrenare!**

86

2. Funcția  $f(x) = \begin{cases} x, & x \geq 0; \\ \lambda(e^x - 1), & x < 0 \end{cases}$  corespunde cărei funcții de activare?
  - a. Sigmoidă
  - b. Tangentă hiperbolică
  - c. PReLU
  - d. ELU**

## Funcții de activare

### Exponential Linear Units (ELU)



- Toate beneficiile ReLU
- Nu se saturează
- Output aproape de medie zero
- **Implicită calculul  $\exp()$**

$$f(x) = \begin{cases} x & \text{if } x > 0 \\ \alpha (\exp(x) - 1) & \text{if } x \leq 0 \end{cases}$$

[Clevert et al., 2015]

369

3. În cazul în care folosim  $\beta = 0$  în cadrul măsurii  $F_\beta$ , care din următoarele egalități este adevărată?

- a.  $F_\beta = R$
- b.  $F_\beta = P + R$
- c.  $F_\beta = P$
- d.  $F_\beta = P \cdot R$

$$F_0 = (1 + 0^2) \cdot \frac{P \cdot R}{(0^2 \cdot P) + R} = P$$

## Cum evaluăm un sistem de învățare automată?

- Măsura  $F_\beta$

$$F_\beta = (1 + \beta^2) \cdot \frac{\text{precision} \cdot \text{recall}}{(\beta^2 \cdot \text{precision}) + \text{recall}}$$

- Măsura  $F_1$  este poate cea mai folosită măsură de tipul  $F_\beta$

4. Aplicând algoritmul coborârii pe gradient cu moment  $m = 0.0$  și rata de învățare  $r = 0.1$ , aplicând funcția  $f(x) = x^3 - 2x + 3$ , se ajunge după mai mulți pași la  $x = 3$ . Care este următoarea valoare pentru  $x$ ? (???)

- a. 4.6
- b. 1.6
- c. 1.4
- d. Nu se poate calcula

5. Cum poate fi corectată variația unei rețele neuronale cu 2 straturi?

- a. Eliminând un strat
- b. Adăugând un strat
- c. Folosind Xavier
- d. Crescând rata de învățare

## Bias-Variance Trade-off

- Bias
  - Eroare sistematică care provine din inabilitatea modelului de a învăța adevărată relație dintre trăsături și etichete (underfitting)
  - Poate fi corectată prin creșterea complexității modelului
- Variance
  - Eroare aleatoare care provine din sensibilitatea ridicată la mici fluctuații din date, cauzată de faptul că modelul a învățat și zgromotul din datele de antrenare (overfitting)
  - Poate fi corectată prin adăugarea de exemple de antrenare sau prin scăderea complexității modelului

(cursul 2), la variance elimini un strat (scade complexitatea), la bias adaugi un strat (crește complexitatea)

6. Care din următoarele modele nu își actualizează ponderile pe bază de gradienți?
- Modelul Widrow-Hoff
  - Perceptronul Rosenblatt (???)**
  - Rețeaua neuronală cu 1 strat
  - Rețea neuronală cu cel puțin 2 straturi

#### Scurt istoric

Mașina **Mark I Perceptron** a fost prima implementare a algoritmului perceptronului.

Mașina era conectată la o camera cu  $20 \times 20$  fotocelule de sulfat de cadmio pentru a produce o imagine cu 400 de pixeli.

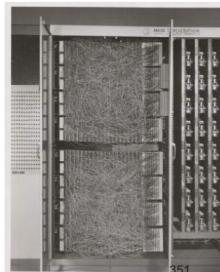
Folosită pentru a recunoaște litere din alfabet.

$$f(x) = \begin{cases} 1 & \text{if } w \cdot x + b > 0 \\ 0 & \text{otherwise} \end{cases}$$

Regula de actualizare

$$w_i(t+1) = w_i(t) + \alpha(d_j - y_j(t))x_{j,i}$$

Frank Rosenblatt, ~1957: *Perceptron*



#### Scurt istoric

Mașina **ADALINE** folosea rezistoare cu memorie capabile să execute operații logice și să stocheze informații.

Funcția de pierdere (suma pătratelor erorilor)

$$\frac{1}{2} \sum_i (d^i - y^i)^2, \text{ unde } y^i = (x^i)^T w + b$$

Regula de actualizare

$$w^{k+1} = w^k + \mu \sum_{i=1}^m (d^i - y^i)x^i$$

$$b^{k+1} = b^k + \mu \sum_{i=1}^m (d^i - y^i)$$

Widrow and Hoff, ~1960: Adaline



352

7. Care din următoarele formule exprimă cota de fals pozitiv?

- FP/(TP+FP)
- FP/(FP+TN)**
- TP/(TP/FP)
- FP/(FP+FN)

Cum evaluăm un sistem de învățare automată?

- Calculul măsurilor TPR și FPR



$$\text{TPR} = \text{TP} / (\text{TP} + \text{FP})$$

$$= 66.67\%$$

$$\text{FPR} = \text{FP} / (\text{FP} + \text{TN})$$

$$= 33.33\%$$

	Predicted YES	Predicted NO
Actual YES	2	1
Actual NO	1	2

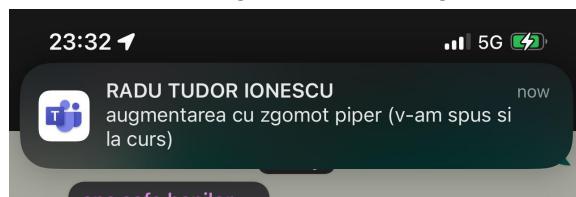
138

$$\text{tpr} = \text{tp} / (\text{tp} + \text{fn})$$

$$\text{fpr} = \text{fp} / (\text{fp} + \text{tn})$$

8. Care din următoarele operații sunt echivalente cu Dropout?

- Blocarea cu filtru Gaussian
- Blurarea cu filtru de medie
- Augmentarea cu zgomot de tip piper (???)**
- Augmentarea cu zgomot Gaussian



9. Se dă mulțimea de exemple de antrenare

$S = \{([-2, 1], 1), ([-3, 1], 1), ([0, 2], -1), ([-3, 2], 1)\}$  și exemplul de testare  $[-1, 2]$ . Rezolvați următoarele subpuncte:

- Calculați funcția de scufundare asociată funcției nucleu polinomiale cu  $d=3$  și  $c = 3$
- Aplicați funcția de scufundare de la punctul a pe ex de test
- Știind că ponderile duale ale unui model SVM bazat pe funcția nucleu polinomială cu  $d=3$  și  $c=3$  sunt  $\alpha_1 = 0.5$ ,  $\alpha_2 = 0.5$ ,  $\alpha_3 = -2$ ,  $\alpha_4 = 1$ , estimați ponderile primale aferente acestui model
- Determinați etichetele exemplului de test folosind clasificatorul SVM primal cu  $w$  calculat la punctul b) și  $b=-3$ . Pentru a obține eticheta se aplică funcția semn

Sursă Chat GPT 9. Se dă mulțimea de exemple de antrenare  $S = \{([-2, 1], 1), ([-3, 1], 1), ([0, 2], -1), ([-3, 2], 1)\}$  și exemplul de testare  $[-1, 2]$ . Rezolvați următoarele subpuncte:

a) Calculați funcția de scufundare asociată funcției nucleu polinomială cu  $d = 3$  și  $c = 3$

Funcția de scufundare asociată funcției nucleu polinomială este dată de:

$$\phi(x) = (1 + x * y)^d$$

Pentru fiecare exemplu din mulțimea de antrenare, calculăm funcția de scufundare:

$$\phi([-2, 1]) = (1 + (-2) * 1)^3 = (-1)^3 = -1$$

$$\phi([-3, 1]) = (1 + (-3) * 1)^3 = (-2)^3 = -8$$

$$\phi([0, 2]) = (1 + 0 * 2)^3 = 1^3 = 1$$

$$\phi([-3, 2]) = (1 + (-3) * 2)^3 = (-5)^3 = -125$$

b) Aplicați funcția de scufundare de la punctul a pe exemplul de test

Aplicăm funcția de scufundare pe exemplul de test  $[-1, 2]$ :

$$\phi([-1, 2]) = (1 + (-1) * 2)^3 = (-1)^3 = -1$$

c) Știind că ponderile duale ale unui model SVM bazat pe funcția nucleu polinomială cu  $d = 3$  și  $c = 3$  sunt  $1 = 0.5$ ,  $2 = 0.5$ ,  $3 = -2$ ,  $4 = 1$ , estimați ponderile primale aferente acestui model

Ponderile primale sunt calculate folosind relația:

$$w = \sum(ai * yi * \phi(xi))$$

Înlocuind cu valorile date, obținem:

$$w = (0.5 * 1 * \phi([-2, 1])) + (0.5 * 1 * \phi([-3, 1])) + (-2 * 1 * \phi([0, 2])) + (1 * 1 * \phi([-3, 2]))$$

Calculăm valorile:

$$w = (0.5 * 1 * -1) + (0.5 * 1 * -8) + (-2 * 1 * 1) + (1 * 1 * -125)$$

$$w = -0.5 - 4 - 2 - 125$$

$$w = -131.5$$

Deci ponderile primale sunt  $w = -131.5$

d) Determinați etichetele exemplului de test folosind clasificatorul SVM primal cu  $w$  calculat la punctul b) și  $b = -3$ . Pentru a obține eticheta se aplică funcția semn

Calculăm:

$$f([-1, 2]) = w * \phi([-1, 2]) + b$$

$$f([-1, 2]) = -131.5 * -1 + (-3)$$

$$f([-1, 2]) = 131.5 - 3$$

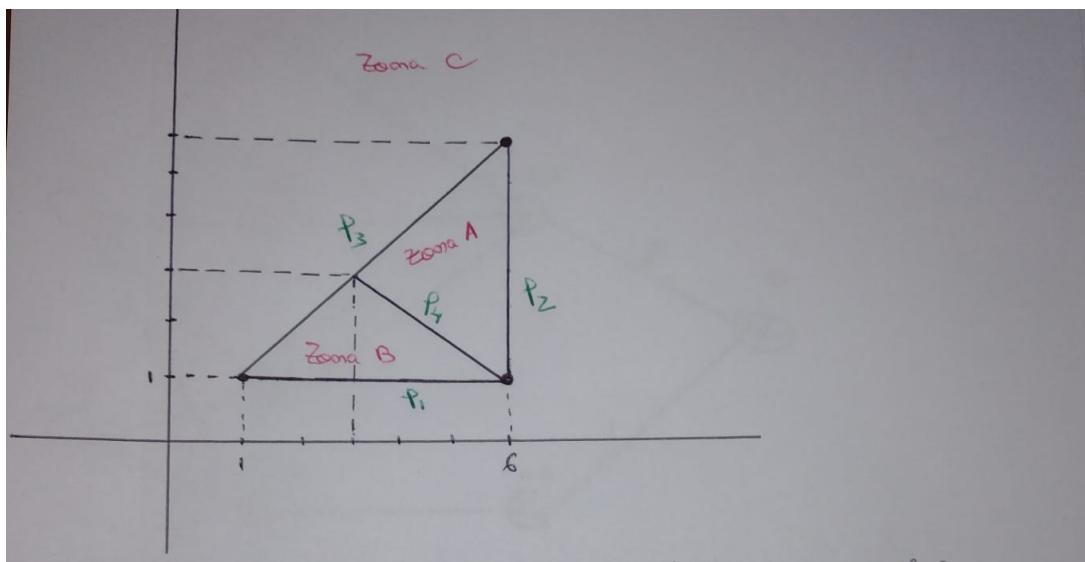
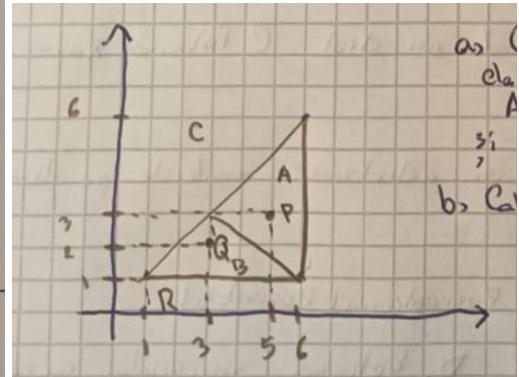
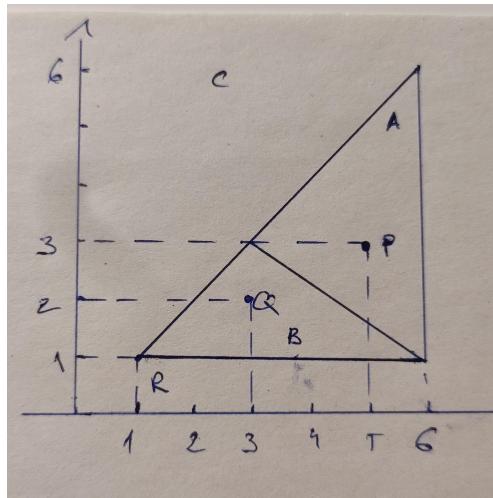
$$f([-1, 2]) = 128.5$$

Pentru a obține eticheta, aplicăm funcția semn:  $semn(f([-1, 2])) = semn(128.5) = 1$

Deci eticheta exemplului de test  $[-1, 2]$  este 1.

10. Fieind date multimile A, B și C din figura de mai jos

- Construiți o rețea neuronală care să clasifice corespunzător punctele din multimea A, B și C. Desenați arhitectura rețelei și calc. ponderile asociate fiecărui strat
- Calc ieșirea rețelei pt pct. P, Q, R din fig



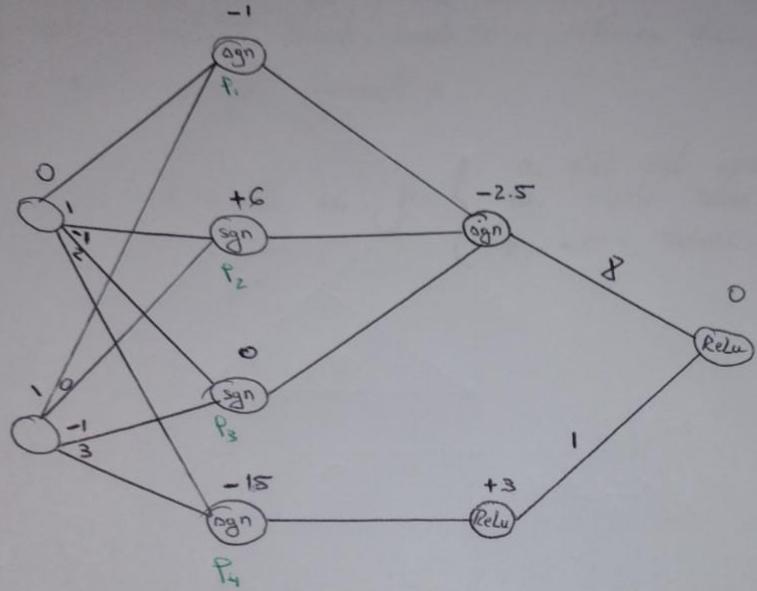
$P_1$  urmărește motonul ale 4 drepte cu  $P_1, P_2, P_3, P_4$  și colinear ecuațiile frecăreia.

$$P_1: y=1 \Rightarrow P_1: y-1=0$$

$$P_2: x=6 \Rightarrow P_2: x-6=0 \Rightarrow P_2: -x+6=0$$

$$P_3: x=y \Rightarrow P_3: x-y=0$$

$$P_4: \frac{x-6}{3-6} = \frac{y-1}{3-1} \Rightarrow P_4: \frac{x-6}{-3} = \frac{y-1}{2} \Rightarrow P_4: 2x+3y-15=0$$



Explicări:

- Pe primul plan se vede că există 2 neuroni care corespund coordonatelor punctului de referință și testot (planul x, y, z, al dorlea)
- Pe al doilea plan se observă 4 neuroni care corespund treptelor drepte. Folosind funcția de activare sign și weight-unite brios-unile de pe figura, în neuroni  $P_1, P_2, P_3$  vor fi voturi de 1, deoarece punctul e în intervalul întunghisului.

SIMULTAN

- Deoarece în planul  $x, y$  neuroni de pe planul 3 e vol喬ea 1, obțin punctul e în intervalul întunghisului, obțin un vot.
- În al doilea plan neuroni de pe planul 3 va fi or
- Vol喬ea 2, ori vol喬ea 4.

- Weight-up și e pus în ora jec simbolic, deoarece punctul e în spatele întunghisului, vol pe care o primește ultimul neuron și încearcă să-l ia, totuși deoarece este în intervalul întunghisului și primește și t, unde t = vol喬ea din al doilea neuron de pe planul 3.

Output-ul rețelei va fi  $\begin{cases} 0, & \text{dacă } \text{jet } \text{operativ zone} C \\ 1, & \text{=} \text{ zone} B \\ 2, & \text{=} \text{ zone} A. \end{cases}$

11. În urma antrenării unui SVM hard-margin pe mulțimea de 4 pot repr. colțurile pătratului unitate cu etichetele date căruia funcția logică OR, se obțin ponderile  $w = [1, 1]$  și  $b = -0.5$ . Cum se modifică poziția hiperplanului de separare prin adăugarea exemplului  $[-1, -1]$  cu eticheta 0? Justificați.

11. În urma antrenării unui SVM hard-margin pe mulțimea de 4 pot repr. colțurile pătratului unitate cu etichetele date căruia funcția logică OR, se obțin ponderile  $w = [1, 1]$  și  $b = -0.5$ . Cum se modifică poziția hiperplanului de separare prin adăugarea exemplului  $[-1, -1]$  cu eticheta 0? Justificați.

~ Radu

asta zice gpt la aia cu svm de la examen. si eu am scris tot ceva de genul ca nu se modifica si ramane la fel si a dat maxim pe ex ala

Actualizarea ponderilor w:

$$\begin{aligned}w' &= w + y^* x \\&= [1, 1] + 0 * [-1, -1] \\&= [1, 1]\end{aligned}$$

Deoarece eticheta este 0, multiplicăm vectorul de intrare cu 0, ceea ce nu schimbă ponderile w.

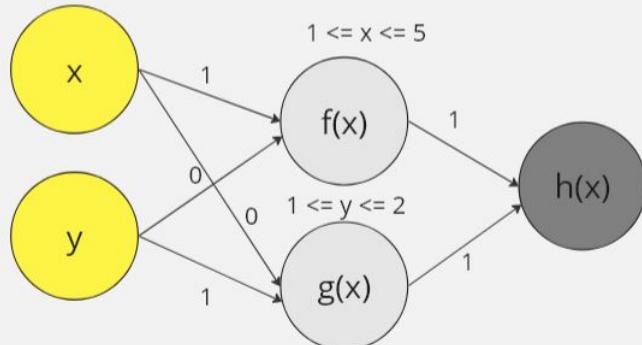
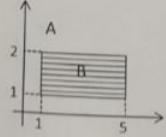
Actualizarea offsetului b:

$$\begin{aligned}b' &= b + y \\&= -0.5 + 0 \\&= -0.5\end{aligned}$$

Deoarece eticheta este 0, adăugăm 0 la offsetul b, ceea ce nu schimbă poziția hiperplanului.

15:20

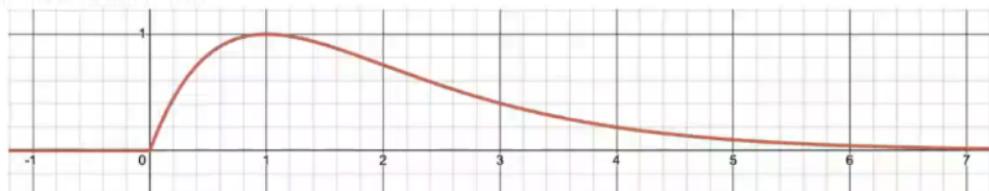
11. (1p) Configurați o rețea neuronală (specificând arhitectura și valorile parametrilor) care să discrimineze între mulțimile de puncte A (eticheta 1) și B (eticheta -1) reprezentate în figura de mai jos.



$$f(x) = 1 \text{ dacă } x \leq 5 \text{ și } x \geq 1 \text{ altfel } 0$$
$$g(x) = 1 \text{ dacă } x \leq 2 \text{ și } x \geq 1 \text{ altfel } 0$$
$$h(x) = -1 \text{ dacă } x = 2 \text{ și } 1 \text{ altfel}$$

## ML - CONSULTATIE 2023

1. (0.25p) Care este numărul de parametri ai unui strat convolutional cu 30 de filtre de dimensiune 5x5 aplicate cu un stride de 2 pe un tensor de dimensiune 64x64x10?
- A. 7530      B. 7510      C. 780      D. 750
2. (0.25p) Dacă un nod dintr-un graf computațional reprezintă operația  $z = 2xy$ , intrările sunt  $x = -0.5$  și  $y = 2$ , iar gradientul  $\partial L/\partial z = -10$ , atunci gradienții 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  $\partial L/\partial y = -10$
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) 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
5. (0.25p) Ce formulă ne dă legătura dintre ponderile primale și cele duale?
- A.  $w = X' \cdot \alpha$       B.  $\alpha = X \cdot w$       C.  $w = X \cdot X' \cdot \alpha$       D.  $w = K \cdot \alpha$
6. (0.25p) Ce combinație de ponderi produce un hiperplan care separă spațiul  $\mathbb{R}^2$  după axa  $Ox$ , a.î. partea de deasupra axei să aibă eticheta pozitivă?
- A.  $w = [-1, 1]$ ,  $b = 1$       B.  $w = [0, -2]$ ,  $b = 0$   
 C.  $w = [0, 0.5]$ ,  $b = 0$       D.  $w = [0, 1]$ ,  $b = 1$
7. (0.25p) Dacă avem o rețea cu un strat cu 12 neuroni, urmat de un strat cu 8 neuroni, atunci dispersia în cazul inițializării Xavier este?
- A. 20      B. 10      C. 0.1      D. 0.2
8. (0.25p) Ce garantează posibilitatea de a ajunge la acuratețe 100% pe antrenare cu un model SVM?
- A. Eliminarea regularizării      B. Parametrul C  
 C. Kernelul RBF      D. Numărul de epoci de antrenare
9. (1p) Considerând funcția kernel  $k(x, y) = 2(P - Q)$ , 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  $2(P - Q) = \langle \phi(x), \phi(y) \rangle$  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.



10. (1p) Fiind dată rețeaua neuronală definită mai jos, să se rezolve următoarele subpunkte:

$$f(x) = \text{hardlim}\left([-1, -1, -1] \cdot \text{sign}\left(\begin{bmatrix} -1 & -1 \\ 0 & 1 \\ 1 & 1 \end{bmatrix} \cdot x + \begin{bmatrix} 0.5 \\ -0.5 \\ -1.5 \end{bmatrix}\right)\right), \text{ unde } \text{hardlim}(x) = \begin{cases} 1, & x \geq 0 \\ 0, & x < 0 \end{cases}$$

- a) (0.5p) Desenați arhitectura rețelei definită mai sus;  
 b) (0.25p) Calculați ieșirea rețelei pe intrarea  $x = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$ ;  
 c) (0.5p) Propuneți două modificări minimale ale arhitecturii a.î. aceasta să rezolve problema XOR;  
 d) (0.25p) Demonstrați că noua rețea (cea de la punctul c) produce rezultatul dorit.

v1 a d a c b b c c

IA - Învățare automată

Examen – Varianta 1

8 iunie 2022

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. Rețelele 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 conoluțional 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?

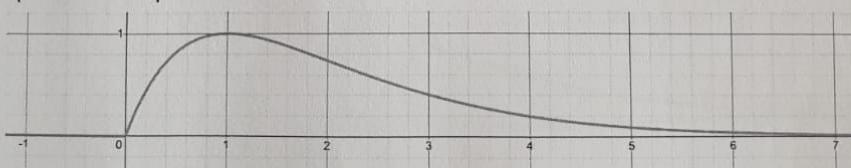
- A. Matricile de ponderi      B. Funcția de pierdere      C. Funcțiile de transfer      D. Ponderile de tip bias

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?

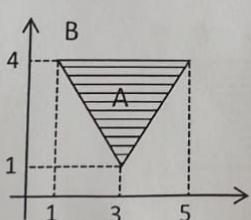
- A. 16x16x40      B. 31x31x40      C. 17x17x40      D. 32x32x40

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.



v2 b c b c d a a a

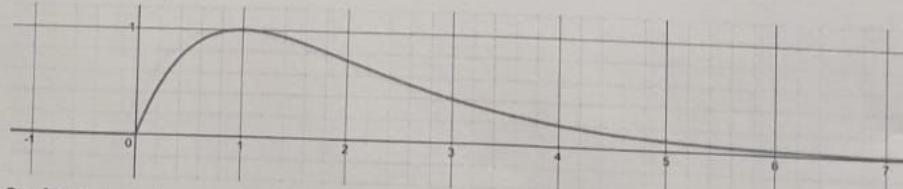
IA - Învățare automată

Examen – Varianta 2

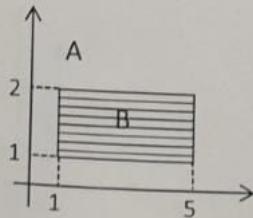
8 iunie 2022

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 gradienții 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 ai unui strat convecțional 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.



v3 b c d a b c c d

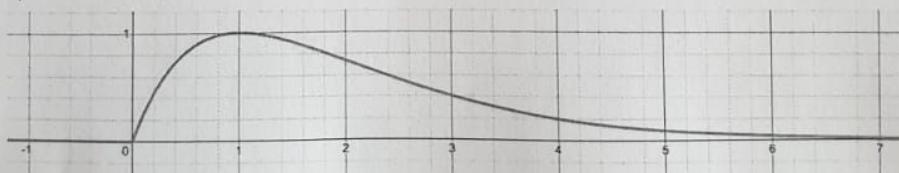
IA - Învățare automată

8 iunie 2022

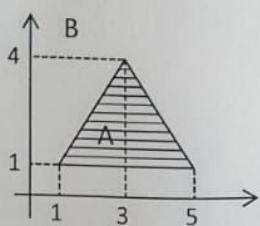
Examen – Varianta 3

Oficiu (0.5p)

1. (0.25p) Cum se descompune eroarea unui model de învățare automată?  
A. Eroare de modelare, optimizare și testare      B. Eroare de modelare, estimare și optimizare  
C. Eroare de modelare, învățare și testare      D. Eroare de modelare, optimizare și generalizare
2. (0.25p) Când ar putea să apară fenomenul Hughes?  
A. Când avem prea multe exemple      B. Când algoritmul de optimizare nu converge  
C. Când avem prea multe trăsături      D. Când avem un model cu capacitate de modelare prea mică
3. (0.25p) Care este numărul de parametri ai unui strat convecțional cu 20 de filtre de dimensiune 3x3 aplicate cu un stride de 3 pe un tensor de dimensiune 128x128x40?  
A. 200      B. 180      C. 7240      D. 7220
4. (0.25p) Care este scufundarea asociată funcției nucleu RBF?  
A. Nu există      B.  $f(x) = x$   
C.  $k(x, y) = e^{-\frac{\Delta(x,y)}{2\sigma}}$ , unde  $\Delta(x, y)$  este distanța Euclidiană      D.  $f(x) = \sqrt{x}$
5. (0.25p) Ce determină capacitatea de modelare neliniară a rețelelor neuronale?  
A. Matricile de ponderi      B. Funcțiile de transfer      C. Funcția de pierdere      D. Ponderile de tip bias
6. (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 gradienții în raport cu intrările  $\partial L/\partial x$  și  $\partial L/\partial y$  sunt:  
A.  $\partial L/\partial x = -40$  și  $\partial L/\partial y = 10$       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.  $\partial L/\partial x = -20$  și  $\partial L/\partial y = 5$
7. (0.25p) Care este dimensiunea tensorului de ieșire al unui strat convecțional cu 40 filtre de 3x3 aplicate la stride de 2 pe un tensor de dimensiune 65x65x20?  
A. 16x16x20      B. 31x31x20      C. 32x32x40      D. 31x31x40
8. (0.25p) La ce se referă teorema „no free lunch”?  
A. Rețelele neuronale sunt în general soluția optimă      B. Lansarea procesului de antrenare este costisitoare  
C. Rețelele neuronale aproximează orice funcție      D. Nu există un singur model bun pe toate task-urile
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  $x = (5, 2, 1, 0)$  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.



v4 b a a d a c b c

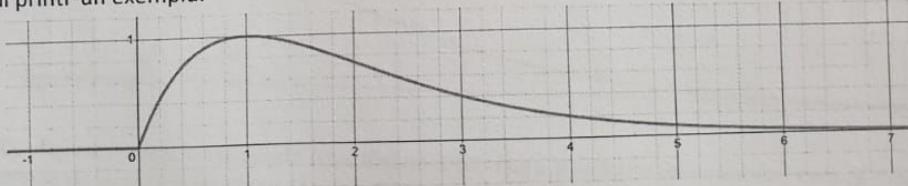
IA - Învățare automată

8 iunie 2022

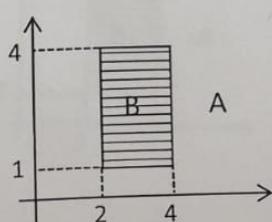
Examen – Varianta 4

Oficiu (0.5p)

1. (0.25p) Care este scufundarea asociată funcției nucleu Hellinger?  
A.  $f(x) = x$       B.  $f(x) = \sqrt{x}$   
C.  $k(x, y) = (\sqrt{x}, \sqrt{y})$ , unde  $\langle \cdot, \cdot \rangle$  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 convecțional 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 convecțional 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 gradienții 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 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  $x = (4, 0, 3, 2)$  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.



# ML - EXAMEN 2023

Examen 1A

ML - Varianta 2

1. Eroarea de generalizare este aproximată în practică prin:

- A. Eroare de estimare
- B. Eroare optimale
- C. Eroare de optimizare
- D. Eroare empirică

2. Funcția  $f(x) = \begin{cases} x, & x \geq 0 \\ \lambda(e^x - 1), & x < 0 \end{cases}$  corespunde cărei funcții de activare?

- A. Sigmoid
- B. Tangentă hiperbolică
- C. PReLU
- D. ELU

3. În cazul în care folosim  $\beta = 0$  în cadrul măsurii  $F_\beta$ , care din următoarele egalități este adevărată?

- A.  $F_\beta = R$
- B.  $F_\beta = R + \bar{R}$
- C.  $F_\beta = \bar{R}$
- D.  $F_\beta = R \cdot \bar{R}$

4. Aplicând algoritmul coborârii pe gradient cu moment  $m=0.0$ , și rata de învățare  $\epsilon=0.1$ , optimizând funcția  $f(x) = x^3 - 2x + 3$ , se ajunge după mai multe pași la  $x=3$ . Care este următoarea valoare pentru  $x$ ?

- A. 4.6
- B. 1.6
- C. 4
- D. Nu se poate calcula

5. Cum poate fi corectată varianța unei rețele neuronale cu 128 neuroni în 2 straturi

- A. Eliminând un strat
- B. Adăugând un strat
- C. Folosind Xavier
- D. Crescând rata de învățare

6. Care din următoarele moduri nu își actualizează ponderile pe baza de gradienti?

- A. Modelul Widrow-Hoff
- B. Perceptronul Rosenblatt
- C. Rețeaua neuronală cu 1 strat
- D. Rețeaua neuronală cu cel puțin 2 straturi

7. Care din următoarele formule exprimă rată de faza pozitiv

- A.  $FP / (TP + FP)$  B.  $FP / (FP + TN)$  C.  $TP / (TP + FP)$   
D.  $FP / (FP + FN)$

8. Care din următoarele operații sunt echivalente cu Despăgub?

- A. Blurarea cu filtreu Gaussian B. Blurarea cu filtreu de medie  
C. Avergitorarea cu egomorf de tip pipar D. Avergitorarea cu egomorf Gaussian

9. Se dă multimea de exemplu de antrenare  $S = \{([-2, 1], 1), ([-3, 1], 1), ([0, 2], -1), ([-3, 2], 1)\}$  și exemplul de

testare  $[-1, 2]$ . Rezolvă următoarele subpunete

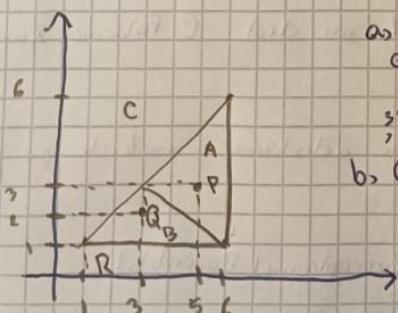
a) Calculați funcția de scufundare asociată funcției nucleu polinomială cu  $d=3$  și  $c=3$

b) Aplicați funcția de scufundare de la punctul a) pe exemplul

c) Să se calculeze ponderile obținute ale unui model SVM bazat pe funcție nucleu polinomială cu  $d=3$  și  $c=3$  sunt  $\alpha_1=0,5$ ,  $\alpha_2=0,5$ ,  $\alpha_3=-2$ ,  $\alpha_4=1$ , estimând ponderile primele aferente acestui model

d) Determinați etichete, exemplului de test folosind clasificatorul SVM primul cu  $w$  calculat la punctul b) și  $b=-3$ . Pentru a obține eticheta se aplică fel semn.

10. Fiind date multimiile A, B și C din fig de maijos:



a) Construiți o rețea neuronală care să clasifice corespunzător punctele din multimi A, B și C. Descrieți arhitectura rețelei și calculează ponderile asociate fiecărui clust.

b) Calculează rezultația pentru P, Q, R din fig

11. În urma antrenării unui SVM hard-margin pe multimea de tip pot se obține că funcția de clasificare este  $f(x) = w^T x + b$ . Funcția logică OR, se obțin ponderile  $w = [1, 1]$  și  $b = -0,5$ . Cum se modifică poziția hiperplanului de separare primătăgore sau exemplului  $[1, 1]$  cu eticheta 0? Justificați.

## ALTE RESURSE

<https://github.com/DimaOanaTeodora/Uni-Work-2019-2022/blob/main/AN%20202%20SEM%202/Inteligenta%20artificiala/ToateRaspunsurile.pdf>

<https://github.com/DariusBuhai/FMI-Unibuc/tree/main/Year%20II/Semester%202/IA/Examen/Examen/Part%202>

1. Functia de scufundare  $(x_1, x_2) \rightarrow (z_1, z_2, z_3, z_4)$ , cu  $z_1 = x_1 * x_2$ ,  $z_2 = x_2 * x_1$ ,  $z_3 = x_1^2$ ,  $z_4 = x_2^4$ , este un caz particular corespunzator functiei kernel:

A. Liniare ( $?>?????$ )

B. Polinomiale

C. Gaussiene

D. Duale

2. Dupa antrenarea unei retele neuronale cu cate 100 de perceptri pe 5 straturi ascunse, timp de 100 de epoci, folosind o rata de invatare de 0.0001, obtinem o acuratete de 42.25% pe multimea de antrenare si 41.23% pe multimea de test. Cum interpretati rezultatele si ce variante putem incerca pentru imbunatatirea performantei?

A. Modelul face underfit. Ar trebui sa crestem rata de invatare.

B. Modelul face overfit. Ar trebui sa utilizam early stopping folosind o multime de validare pentru a seta numarul optim de epoci.

C. Rezultatele obtinute sunt satisfacatoare, putem considera ca modelul nostru este optim.

D. Modelul face overfit. Ar trebui sa scadem treptat rata de invatare in functie de valoarea functiei de cost.

3. Care din urmatoarele functii este o functie de activare?

A. MSE

B. Leave-one-out

C. SGD

D. Tangenta hiperbolica

4. Care din urmatoarele functii de activare aduce output-ul in intervalul [0, infinit)?

A. Leaky ReLU

B. ReLU

C. ELU

D. Tangenta Hiperbolica

5. Cati clasificatori binari trebuie antrenati pentru a rezolva o problema cu 6 clase folosind metoda One vs One?

A. 6

B. 25

C. 15

D. 30

6. Care din urmatoarele metode este folosita pentru imbunatatirea capacitatii de generalizare a unui model?

A. Adaugarea regularizarii L2

B. MLP

- C. Adaugarea mai multor straturi
- D. Reducerea numarului de exemple

7. Pentru documentele de antrenare: ['Ana are mere', 'Ana pregateste placinta cu mere', 'Ana cumpara portocale'] si documentul de test ['Poti face placinta din portocale'], care este reprezentarea documentului 'Ana pregateste placinta cu mere' folosind modelul BOW?

- A. [1 0 1 1 1 1 0 0 0 0]
- B. [1 0 1 1 1 1 0 0]
- C. [0 0 0 1 1 1 1 0 0 0 0 0 0]
- D. [0 0 0 1 1 1 1 1 0 0 0]

8. Care este rezultatul aplicarii functiei de activare ReLU(x) unde  $x = -227.09$ ?

- A. 1
- B. 227.09
- C. 0
- D. -227

9. Avem 10 perechi concordante (P) si 4 perechi discordante (Q),  $n=7$ . Cat este corelatia Kendel Tau?

- A. 0.81
- B. 0.52
- C. 0.28
- D. 0.44

10. In cazul unui model one-versus-all, care este principala diferență în utilizarea funcției de pierdere hinge versus cross-entropy?

- A. Hinge permite continuarea optimizării în cazul în care marginea dintre scoruri este suficient de mare
- B. Cross-entropy permite continuarea optimizării în cazul în care marginea dintre scoruri este prea mică
- C. Cross-entropy permite continuarea optimizării în cazul în care marginea dintre scoruri este suficient de mare
- D. Modele sunt în esență echivalente

(1 raspuns gresit, 1 raspuns necompletat)

Raspunsuri:

- 1. A
- 2. A
- 3. D
- 4. B
- 5. C
- 6. A
- 7. B
- 8. C
- 9. C
- 10. -

<https://github.com/M4731/FMI/tree/master/Year%20II/Artificial%20Intelligence/Machine%20Learning/Exam>

1. Care din urmatoarele functii este o functie de activare?
  - A. MSE
  - B. Tangenta hiperbolica
  - C. SGD
  - D. Leave-one-out
  
2. Care este numarul total de parametri ai unei retele neuronale feedforward cu urmatoarea configuratie: 2 neuroni pe stratul de intrare, 2 straturi ascunse a cate 4 neuroni si 3 neuroni pe stratul de iesire?
  - A. 65
  - B. 47
  - C. 26
  - D. 32
  
3. Care este diferența dintre underfitting și overfitting?
  - A. Overfitting-ul este întotdeauna de preferat pentru că duce la rezultate mai bune pe setul de test.
  - B. Overfitting-ul este un fenomen comun în cazul modelelor foarte simple în timp ce underfitting-ul este un fenomen comun în cazul modelelor complexe.
  - C. Underfitting-ul este întotdeauna de preferat pentru că duce la rezultate mai bune pe setul de test.
  - D. Overfitting-ul este în general un fenomen comun în cazul modelelor complexe în timp ce underfitting-ul este mai comun în cazul modelelor mai simple.
  
4. Care este valoarea de iesire a perceptronului daca intrare=[-29.4, 32.4], ponderi=[-0.5, 5.2], bias=-111.43 (functia de activare - sign)?
  - A. -239.28
  - B. -1
  - C. 273.16
  - D. 1
  
5. Avem 10 perechi concordante (P) si 4 perechi discordante (Q), n=7. Cat este corelatia Kendel Tau?
  - A. 0.52
  - B. 0.44
  - C. 0.81
  - D. 0.28
  
6. Care este o metoda de evitarea suprainvatarii?
  - A. adaugare perceptoni pe primul strat
  - B. adaugare de straturi ascunse
  - C. adaugare regularizare
  - D. adaugare perceptoni pe straturile ascunse
  
7. Pentru care dintre urmatoarele probleme putem masura performanta folosind intersectia supra reunii?

- A. Estimarea veniturilor viitoare ale unei familii
  - B. Identificarea limbii unui text
  - C. Transferul stilului dintr-un domeniu în altul (translatia de imagini)
  - D. Detectarea obiectelor dintr-o imagine
8. Căt este valoarea funcției de pierdere hinge loss pentru  $y = [1, -1]$  și  $y_{\text{hat}} = [1.1, -0.7]$ ?
- A. -0.3
  - B. 0.2
  - C. 0.4
  - D. 0.3
9. În urma antrenării unui model de învățare automată, acuratețea pe setul de test este similară cu acuratețea pe setul de antrenare. În ce situație suntem?
- A. Underfitting
  - B. Underfitting și overfitting
  - C. Overfitting
  - D. Optimum
10. Care afirmație este adevarata pentru un niste date normalize cu normalizarea L1?
- A. Suma valorilor absolute pentru un rand normalizat este 1
  - B. Valorile datelor normalize sunt între 0 și 1
  - C. Normalizarea L1 nu produce valori egale sau apropiate de 0
  - D. Normalizarea L1 se aplică pe coloane, nu pe randuri

<https://github.com/CosminHorjea/fmi/blob/master/Sem%20IV/Inteligenta%20artificaiala/ML/Examen/test-219.txt>

1. Care dintre următoarele activări ale neuronilor este rezultatul funcției de activare tangentă hiperbolică?
- A. [-0.1, 0, 1.0, -0.2]
  - B. [0.1, 0.5, 0.3, 0.1]
  - C. [0.1, -0.1, 3.2, -0.001]
  - D. [-0.2, 0, 2.3, 5.1]
2. Dacă avem probabilitățile pentru evenimentele  $P(A)=0.91$ ,  $P(B)=0.82$ ,  $P(B|A)=0.73$ . Calculați probabilitatea  $P(A|B)$  folosind regula lui Bayes.
- A. 0.95
  - B. 0.65
  - C. 0.58
  - D. 0.81
3. Care este rezultatul aplicării funcției de activare ReLU( $x$ ) unde  $x = -227.09$ ?
- A. 1
  - B. 0
  - C. 227.09
  - D. -227

4. Pentru o retea neuronala cu un singur strat si un singur perceptron, ponderile W formeaza:
- A. Un vector paralel cu hiperplanul de separare, aflat de aceeasi parte cu exemplele cu scoruri pozitive
  - B. Un vector paralel cu hiperplanul de separare, aflat de aceeasi parte cu exemplele cu scoruri negative
  - C. Un vector perpendicular pe hiperplanul de separare, aflat de aceeasi parte cu exemplele cu scoruri pozitive
  - D. O matrice de ponderi care mapeaza datele intr-un spatiu liniar
5. Care este forma duala utilizand kernel-ul liniar pe exemplul de test  $X_{\text{test}} = [2.5, 1]$ , daca  $X_{\text{train}} = [[1, 3], [-2, -4], [0.5, -1], [-4, -2]]$ ?  $X_{\text{train}}$  si  $X_{\text{test}}$  sunt in forma primala.
- A.  $X_{\text{test\_dual}} = [5.5, -9, 0.25, -12]$
  - B.  $X_{\text{test\_dual}} = [-3.5, -9, 0.25, 15]$
  - C.  $X_{\text{test\_dual}} = [-5.5, 9, 0.25, -12]$
  - D.  $X_{\text{test\_dual}} = [-5.5, -9, 0.25, 12]$
6. Fie un set de date format din numere reale x din intervalul  $[-5, 5]$  cu eticheta  $\sin(x)$ ,  $\{(x, \sin(x)) \mid x \in [-5, 5]\}$  si un clasificator format dintr-un neuron fara activare. Acest model va:
- A. Suferi de underfitting
  - B. Suferi de overfitting
  - C. Niciuna dintre variante
  - D. Avea rezultate bune atat pe setul de antrenare cat si pe cel de testare
7. Care este, in general, beneficiul adus de utilizarea a mai multe straturi de neuroni (doua sau mai multe) in cazul retelelor neurale?
- A. Utilizarea unui numar mare de straturi de neuroni cu activare liniara duce la o putere de modelare sporita.
  - B. Marirea numarului de parametrii prin utilizarea a mai multe straturi de neuroni duce la prevenirea fenomenului de regularizare.
  - C. Marirea numarului de parametrii prin utilizarea a mai multe straturi de neuroni duce la prevenirea fenomenului de overfitting.
  - D. Utilizarea a mai multe straturi permite invatarea automata a unor reprezentari din ce in ce mai abstracte a inputurilor.
8. Functia kernel PQ provine din:
- A. Calcularea anumitor perechi concordante si discordante
  - B. Diferenta media intre histograme de valori
  - C. Functia de corelatie Kendall Tau
  - D. Functia kernel liniara aplicata histogramelor de valori
9. Care dintre urmatoarele functii de activare este recomandata pentru stratul de iesire cu 2 neuroni al unei retele neuronale ce discrimineaza intre caini si pisici?
- A. Softmax
  - B. ReLU
  - C. Liniara
  - D. Sigmoid

10. Cati neuroni trebuie sa fie pe primul strat al unei retele neurale?

- A. Un numar egal cu dimensiunea inputului.
- B. Numarul depinde de problema si tipul problemei.
- C. Un numar de neuroni mai mic decat dimensiunea inputului.
- D. Un numar de neuroni mai mare decat dimensiunea inputului.

<https://github.com/mehanix/teme-fmi/tree/master/ia/examen>

1. Care dintre urmatoarele abordari nu este o metoda de scalare a caracteristicilor (feature scaling)?

- A. Standardizarea
- B. Normalizarea L2
- C. Scalarea MinMax
- D. Dropout

2. Care este volumul de output al unui strat convolutional de input  $15 \times 15 \times 3$ , avand 5 filtre de dimensiune 3, stride 1, padding 1?

- A.  $8 \times 8 \times 5$
- B.  $15 \times 15 \times 5$
- C.  $4 \times 4 \times 8$
- D.  $7 \times 7 \times 6$

3. Care dintre urmatoarele tehnici nu este o metoda de prevenire a overfittingului?

- A. Utilizarea unei rate de invatare adaptiva
- B. Oprirea timpurie a antrenarii
- C. Utilizarea regularizarii
- D. Initializarea ponderilor cu 0

4. Dupa antrenarea unei retele neuronale cu cate 100 de perceptri pe 5 straturi ascunse, timp de 1000 de epoci, folosind o rata de invatare de 0.1, obtinem o acuratete de 99.95% pe multimea de antrenare si 61.23% pe multimea de test. Cum interpretati rezultatele si ce variante putem incerca pentru imbunatatirea performantei?

- A. Modelul face underfit. Ar trebui sa crestem numarul de perceptri sau numarul de straturi ascunse utilizate.
- B. Modelul face overfit. Ar trebui sa utilizam o rata de invatare mai mica.
- C. Modelul face overfit. Ar trebui sa utilizeaza mai putine exemple de antrenare.
- D. Rezultatele obtinute sunt satisfacatoare, putem considera ca modelul nostru este optim.

5. Avand urmatoarele date de antrenare:  $X_{train} = ((2,2),(2,1),(1,1),(2,0),(-1,0),(-1,2))$   
 $Y_{train} = (1,1,1,-1,-1,-1)$  si datele de test:  $X_{test} = ((1,1), (2,-1), (1,3), (-1,-1))$ ,  $Y_{test} = (1,1,-1,-1)$ , care este acuratetea metodei 3-NN pe multimea de antrenare?

- A. 1
- B. 1/2
- C. 2/3
- D. 5/6

6. Cand este mai eficient sa folosim reprezentarea primala a datelor?

- A. Cand avem o problema de clasificare binara (cu doua clase)
- B. Cand avem o problema de clasificare cu foarte multe clase (mai mult de doua)
- C. Cand numarul de trasaturi este mai mare decat numarul de exemple
- D. Cand numarul de trasaturi este mai mic decat numarul de exemple

7. Fie  $f(x,y,z) = xy + 2z$ , care sunt valorile derivatelor partiale  $df/dx$ ,  $df/dy$ ,  $df/dz$ , pentru  $x = 1$ ,  $y = 2$ ,  $z = 3$ ?

- A.  $df/dx = 1$ ,  $df/dy = 2$ ,  $df/dz = 1$
- B.  $df/dx = 2$ ,  $df/dy = 1$ ,  $df/dz = 2$
- C.  $df/dx = 2$ ,  $df/dy = 1$ ,  $df/dz = 3$
- D.  $df/dx = 4$ ,  $df/dy = 2$ ,  $df/dz = 1$

8. Care dintre urmatoarele masuri calculeaza procentajul exemplelor relevante identificate in raport cu numarul total de exemple relevante?

- A. Recall
- B. Precizie
- C. MSE
- D. F1

9. Care dintre urmatoarele afirmatii despre retele neuronale convolutionale este falsa?

- A. Pot fi folosite atat pe imagini, cat si pe date de tip text.
- B. Nu pot fi utilizate pentru detectia de obiecte.
- C. Arhitectura clasica este formata din straturi convolutionale, straturi de pooling si straturi dense.
- D. Straturile de pooling miscoreaza dimensiunea activarilor, limitand cantitatea de informatie pierduta.

10. Care nu este o metoda de evitarea suprainvatarii?

- A. adaugarea regularizarii
- B. adaugarea de perceptri pe straturile ascunse
- C. adaugarea de exemple noi
- D. adaugarea unui strat de dropout

1. Care sunt punctele in care dreapta de separare a perceptronului  $w = [-4, 2]$ ,  $b = [2]$  intersecteaza axele?

- A.  $(0, -1)$  si  $(0.5, 0)$
- B.  $(-1, 0)$  si  $(0, 0.5)$
- C.  $(0, 0.5)$  si  $(0.25, 0)$
- D.  $(0.5, 0)$  si  $(0, -0.25)$

x2. Care din urmatorii algoritmi se foloseste pentru optimizare ponderilor retelelor neuronale?

- A. Dropout
- B. ADALINE
- C. Naive Bayes
- D. SGD cu momentum x

3. Care din urmatoarele metode este folosita pentru a combate bias-ul modelului?

- A. Alegerea unui model mai complex x
- B. Reducerea numarului de exemple
- C. F1-score
- D. Early stopping

4. Care dintre aceste optiuni pot sa ajute in imbunatatirea capacitatii de generalizarea unui model de invatare automata?

- A. Oprirea invatarii atunci cand eroarea pe setul de validare incepe sa creasca
- B. Cresterea numarului de epoci de invatare
- C. Micsorarea setului de antrenare
- D. Cresterea numarului de parametri antrenabili ai modelului.

5. Care este scufundarea asociata functie nucleu RBF?

- A.  $f(x) = x$
- B.  $K(x,y) = \exp(-d(x,y) / (2 * \sigma))$ , unde  $d(x,y)$  denota distanta Euclidiana
- C.  $f(x) = \sqrt{x}$
- D. Nu exista

6. In urma antrenarii unui model de invatare automata avem urmatoarele masuri - True Positive = 30, False Positive = 3. Calculati masura Precision?

- A. 0.74
- B. 0.80
- C. 0.67
- D. 0.91????

7. Cati clasificatori binari vor fi antrenati pentru o problema multi-class cu 7 clase folosind schema de combinare one-vs-one?

- A. 1
- B. 21
- C. 42
- D. 7

8. In urma antrenarii unui model de invatare automata avem urmatoarele masuri - Precision = 0.8, Recall=0.9. Calculati masura F1?

- A. 0.72
- B. 0.89
- C. 0.84 x
- D. 0.64

9. Fiind date etichetele  $y = [23, 14, 30, 45, 18, 31]$  si predictiile aferente  $p = [26, 20, 39, 38, 18, 33]$ , care este masura Kendall Tau?

- A. 0.4
- B. 1.0
- C. 0.6
- D. 0.2

10. Daca avem probabilitatile pentru evenimentele  $P(A)=0.4$ ,  $P(B)=0.8$ ,  $P(B|A)=0.5$ . Calculati probabilitatea  $P(A|B)$  folosind regula lui Bayes.

- A. 0.5
- B. 0.4
- C. 0.25 x
- D. 0.65

5. Pentru punctele  $[(2,2),(1,2),(1,1),(2,1)]$ , avand etichetele  $[0,1,1,0]$ , care este eticheta punctului  $(2,2)$  utilizand metoda 3-NN?

- A. 0
- B. Este necesar un criteriu suplimentar pentru departajarea claselor.
- C. 1
- D. 0 in cazul in care  $C \leq 1$ , 1 altfel

3. Pentru care dintre urmatoarele probleme putem masura performanta unui model de invatare automata folosind media patratelor erorilor?

- A. Estimarea veniturilor viitoare ale unei companii
- B. Identificarea limbii unui text
- C. Detectarea semnificativelor dintr-o imagine
- D. Recunoasterea actiunilor dintr-un clip video

4. Care afirmație este adevarata pentru un niste date normalize cu scalarea min-max?

- A. Valorile datelor sunt intre -1 si 1
- B. Valorile datelor sunt intre 0 si 1  $\Rightarrow$  B (netu asa zice)
- C. Scalarea min-max nu functioneaza pentru date negative
- D. Datele vor fi transformate in numere intregi pozitive

8. Daca un clasificator are acuratete 1% pe o multime de antrenare cu 100 clase, cel mai probabil:

- D. Nici una dintre variante  $\Rightarrow$  D.

9. La ce ajuta adaugarea unei componente de momentum in algoritmul coborarii pe gradient?

- B. Aduce o convergenta mai rapida catre un minim local  $\Rightarrow$  B

5. In urma antrenarii unui model de invatare automata, acuratetea pe setul de antrenare este mult mai mare decat acuratetea pe setul de test. In ce situatie suntem?

- C. Overfitting  $\Rightarrow$  C

6. Care din urmatoarele metode este folosita pentru a combatte bias-ul modelului?

- B. Alegerea unui model mai complex  $\Rightarrow$  B

5. Care este forma duala utilizand kernel-ul liniar pe exemplul de test  $X_{\text{test}} = [-0.5, 0.4, -0.5]$ , daca  $X_{\text{train}} = [[-0.3, 2, 0.3], [-1, -2.6, -4], [0.5, -1, -0.5], [-3, -4, -2.7]]$ ?  $X_{\text{train}}$  si  $X_{\text{test}}$  sunt in forma primala.

- A.  $X_{\text{test\_dual}} = [0.8, 1.46, -0.4]$

- B.  $X_{\text{test\_dual}} = [-1.2, -2.46, 1.4, 1.25]$
- C.  $X_{\text{test\_dual}} = [-0.8, -1.46, 0.4, 1.25]$
- D.  $X_{\text{test\_dual}} = [0.8, 1.46, -0.4, 1.25]$

7. Care sunt valorile urmatoarelor exemple de antrenare si testare dupa ce se aplica normalizarea standard (standardizarea),  $X_{\text{train}} = [[10, 5, 3], [-1, -2, -30], [-0.5, 0.5, 0.3]]$ ,  $X_{\text{test}} = [[0.5, -1, 7], [1.2, 3.4, 5.6]]$ ?
- A.  $X_{\text{train}} = [[-1.41, -1.32, -0.79], [-0.75, -1.09, -1.41], [-0.65, -0.23, 0.61]]$ ;  $X_{\text{test}} = [[-0.46, -0.74, 1.06], [-0.32, 0.77, 0.96]]$
  - B.  $X_{\text{train}} = [[1.41, 1.32], [-0.75, -1.09], [-0.65, -0.23]]$ ;  $X_{\text{test}} = [[-0.46, -0.74, 1.06], [-0.32, 0.77, 0.96]]$  ----- n are cum sunt 2
  - C.  $X_{\text{train}} = [[1.41, 1.32, 0.79], [-0.75, -1.09, -1.41], [-0.65, -0.23, 0.61]]$ ;  $X_{\text{test}} = [[0.46, 0.74, -1.06], [-0.32, 0.77, 0.96]]$
  - D.  $X_{\text{train}} = [[1.41, 1.32, 0.79], [-0.75, -1.09, -1.41], [-0.65, -0.23, 0.61]]$ ;  $X_{\text{test}} = [[-0.46, -0.74, 1.06], [-0.32, 0.77, 0.96]]$  ---- posib

8. Cate ponderi invatabile (inclusiv bias) are un filtru convolutional de dimensiune  $11 \times 11 \times 3$ ?
- A. 364 =====> A?????
  - B. 26
  - C. 25
  - D. 363

1. Cat este eroarea MSE pentru urmatoarele etichete prezise  $y_{\text{pred}} = [0.0, 0.1, 2.1, -1.3]$  si etichetele adevarate  $y_{\text{true}} = [1, 0, 1, 0]$ ?
- A. 2.412
  - B. 0.248
  - C. 0.977
  - D. 0.888
2. Cand este mai eficient sa folosim reprezentarea primala a datelor?
- A. Cand avem o problema de clasificare binara (cu doua clase)
  - B. Cand numarul de trasaturi este mai mic decat numarul de exemple
  - C. Cand avem o problema de clasificare cu foarte multe clase (mai mult de doua)
  - D. Cand numarul de trasaturi este mai mare decat numarul de exemple

3. Care dintre urmatoarele functii nu este o functie kernel?
- A.  $K(x,y) = \text{sum}(3x - 3y)$
  - B.  $K(x,y) = \text{sum}(2x - y)$
  - C.  $K(x,y) = \text{sum}(\min\{x_i, y_i\}) + x \cdot y + 2)^{**2}$
  - D.  $K(x,y) = \text{sum}(\sqrt{x_i \cdot y_i})$
4. Care afirmatie este adevarata pentru un niste date normalize cu scalarea min-max?
- A. Valorile datelor sunt intre -1 si 1
  - B. Datele vor fi transformate in numere intregi pozitive
  - C. Valorile datelor sunt intre 0 si 1
  - D. Scalarea min-max nu functioneaza pentru date negative

5. Care este functia utilizata pentru generarea predictiilor (pasul forward) unei retele neuronale feedforward cu 2 straturi ascunse cu activare ReLU si un strat de iesire cu activare Softmax?

- A.  $\text{relu}(\text{softmax}(x * w + b) + |w|)$
- B.  $\text{softmax}(2 * \text{relu}(x * w_1 + b_1))$
- C.  $\text{softmax}(\text{relu}(\text{relu}(x * w_1 + b_1) * w_2 + b_2) * w_3 + b_3)$
- D.  $\text{softmax}(2 * \text{relu}(x * w + b) * \text{relu}(x * w + b))$

6. Cati clasificatori binari vor fi antrenati pentru o problema multi-class cu 4 clase folosind schema de combinare one-vs-all?

- A. 32
- B. 16
- C. 4
- D. 1

7. Care este valoarea de iesire a perceptronului daca intrare=[11.4, -5.4], ponderi=[3.2, 0.5], bias=-23.3 (functia de activare - sign)?

- A. 9.53
- B. -1
- C. 1
- D. -9.52

8. Daca un nod dintr-un graf computational reprezinta operatia  $z = 2 * x * y^2$ , intrarile sunt  $x = -0.5$  si  $y = 2$ , iar gradientul  $dL/dz = -5$ , atunci gradientii in raport cu intrarile  $dL/dx$  si  $dL/dy$  sunt:

- A.  $dL/dx = -20$  si  $dL/dy = 5$
- B.  $dL/dx = 40$  si  $dL/dy = -5$
- C.  $dL/dx = -40$  si  $dL/dy = 10$
- D.  $dL/dx = 10$  si  $dL/dy = -5$

9. Daca spatiul de intrare al trasaturilor are dimensiunea 5 atunci, pentru a nu suferi de "blestemul dimensionalitatii":

- A. Ar trebui ca multimea de antrenare sa aiba aproximativ  $10^5$  exemple.
- B. Ar trebui ca multimea de antrenare sa aiba aproximativ  $5^5$  exemple
- C. Ar trebui ca multimea de antrenare sa aiba aproximativ  $5 \times 5$  exemple
- D. Ar trebui ca multimea de antrenare sa aiba aproximativ 500 exemple

10. Care dintre urmatoarele tehnici nu este o metoda de prevenire a overfittingului?

- A. Oprirea timpurie a antrenarii
- B. Scaderea ratei de invatare
- C. Utilizarea regularizarii
- D. Normalizarea datelor

1. Care este scufundarea asociata functie nucleu liniare?

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

2. Care este rezultatul aplicarii functiei de activare  $\text{ReLU}(x)$  unde  $x = 0.63$ ?

- A. 1
- B. 0
- C. -0.63
- D. 0.63

3. Care este scufundarea asociata functie nucleu RBF?

- A.  $K(x,y) = \exp(-d(x,y) / (2 * \sigma))$ , unde  $d(x,y)$  denote distanta Euclidiana
- B.  $f(x) = x$
- C. Nu exista
- D.  $f(x) = \sqrt{x}$

4. Care din urmatoarele functii de activare aduce output-ul in intervalul (-1, 1)?

- A. Sigmoida
- B. ELU
- C. Tangenta Hiperbolica
- D. Leaky ReLU

5. Care din urmatoarele metode este folosita pentru imbunatatirea capacitatii de generalizare a unui model?

- A. Adaugarea mai multor straturi
- B. MLP
- C. Reducerea numarului de exemple
- D. Dropout

6. In cazul analizei liniar discriminante, hiperplanul pe care se proiecteaza punctele este:

- A. Orientat astfel incat punctele sa fie cat mai departate
- B. Perpendicular pe hiperplanul de separare
- C. Orientat astfel incat punctele sa fie cat mai apropiate
- D. Paralel cu hiperplanul de separare, distanta fiind controlata prin bias

7. Care este scufundarea asociata functie nucleu Hellinger?

- A. Nu exista
- B.  $f(x) = \sqrt{x}$
- C.  $f(x) = x$
- D.  $K(x,y) = \langle \sqrt{x}, \sqrt{y} \rangle$ , unde  $\langle , \rangle$  denote produsul scalar

8. Care din urmatoarele functii este o functie de activare?

- A. NeLU
- B. SGD
- C. Leave-one-out
- D. Leaky ReLU

9. Cati clasificatori binari vor fi antrenati pentru o problema multi-class cu 4 clase folosind schema de combinare one-vs-all?

- A. 4
- B. 1

- C. 16  
D. 32

10. Care dintre urmatoarele afirmatii despre retele neuronale convolutionale este falsa?
- A. Pot fi folosite atat pe imagini, cat si pe date de tip text.
  - B. Straturile convolutionale sunt ne-antranabile, fiind alcătuite din filtre presetate de utilizator.
  - C. Straturile de pooling miscoreaza dimensiunea activarilor, limitand cantitatea de informatie pierduta.
  - D. Automatizeaza extragerea informatiilor relevante.
1. Care sunt valorile urmatoarelor exemple de antrenare si testare dupa ce se aplica normalizarea standard (standardizarea),  $X_{train} = [[10, 5, 3], [-1, -2, -30], [-0.5, 0.5, 0.3]]$ ,  $X_{test} = [[0.5, -1, 7], [1.2, 3.4, 5.6]]$ ?
- A.  $X_{train} = [[-1.41, -1.32, -0.79], [-0.75, -1.09, -1.41], [-0.65, -0.23, 0.61]]$ ;  $X_{test} = [[-0.46, -0.74, 1.06], [-0.32, 0.77, 0.96]]$
  - B.  $X_{train} = [[1.41, 1.32, 0.79], [-0.75, -1.09, -1.41], [-0.65, -0.23, 0.61]]$ ;  $X_{test} = [[0.46, 0.74, -1.06], [-0.32, 0.77, 0.96]]$
  - C.  $X_{train} = [[1.41, 1.32, 0.79], [-0.75, -1.09, -1.41], [-0.65, -0.23, 0.61]]$ ;  $X_{test} = [[-0.46, -0.74, 1.06], [-0.32, 0.77, 0.96]]$
  - D.  $X_{train} = [[1.41, 1.32], [-0.75, -1.09], [-0.65, -0.23]]$ ;  $X_{test} = [[-0.46, -0.74, 1.06], [-0.32, 0.77, 0.96]]$
2. In urma antrenarii unui model de invatare automata avem urmatoarele masuri - Precision = 0.4, Recall=0.8. Calculati masura F1?
- A. 0.72
  - B. 0.64
  - C. 0.53
  - D. 0.89
3. Pentru care dintre urmatoarele probleme putem masura performanta unui model de invatare automata folosind media patratelor erorilor?
- A. Estimarea veniturilor viitoare ale unei companii
  - B. Identificarea limbii unui text
  - C. Detectarea semafoarelor dintr-o imagine
  - D. Recunoasterea actiunilor dintr-un clip video
4. Care afirmatie este adevarata pentru un niste date normalize cu scalarea min-max?
- A. Valorile datelor sunt intre -1 si 1
  - B. Valorile datelor sunt intre 0 si 1
  - C. Scalarea min-max nu functioneaza pentru date negative
  - D. Datele vor fi transformate in numere intregi pozitive
5. In urma antrenarii unui model de invatare automata, acuratetea pe setul de antrenare este mult mai mare decat acuratetea pe setul de test. In ce situatie suntem?
- A. Underfitting
  - B. Underfitting is overfitting
  - C. Overfitting
  - D. Optimum

6. Care din urmatoarele metode este folosita pentru a combate bias-ul modelului?
- A. F1-score
  - B. Alegerea unui model mai complex
  - C. Reducerea numarului de exemple
  - D. Early stopping
7. Cati parametri antrenabili (ponderi + bias) are o retea neuronala cu stratul de intrare avand 6 componente, un strat ascuns de 24 neuroni si stratul de iesire cu 3 neuroni?
- A. 243
  - B. 24
  - C. 300
  - D. 64
8. Daca un clasificator are acuratete 1% pe o multime de antrenare cu 100 clase, cel mai probabil:
- A. Este in regim de suprainvatare
  - B. Acurateata este potrivita
  - C. Este in regim optim de antrenare
  - D. Nici una dintre variante
9. La ce ajuta adaugarea unei componente de momentum in algoritmul coborarii pe gradient?
- A. Permite regularizarea modelului
  - B. Aduce o convergenta mai rapida catre un minim local
  - C. Determina convergenta catre minimul global
  - D. Permite utilizarea unei arhitecturi mai adanci
10. In urma antrenarii unui model de invatare automata avem urmatoarele masuri - True Positive = 64, False Positive = 6. Calculati masura Precision?
- A. 0.74
  - B. 0.91
  - C. 0.67
  - D. 0.80
1. Care este scufundarea asociata functie nucleu RBF?
- A.  $k(x,y) = \exp(-d(x,y) / (2 * \sigma))$ , unde  $d(x,y)$  denote distanta Euclidiana
  - B. Nu exista
  - C.  $f(x) = x$
  - D.  $f(x) = \sqrt{x}$
2. Pentru punctele  $[(2,2), (1,2), (1,1), (2,1)]$ , avand etichetele  $[0, 1, 1, 0]$ , care este eticheta punctului  $(2,2)$  utilizand metoda 3-NN?
- A. Este necesar un criteriu suplimentar pentru departajarea claselor.
  - B. 0 in cazul in care  $C \leq 1$ , 1 altfel
  - C. 0
  - D. 1

3. In urma antrenarii unui model de invatare automata avem urmatoarele masuri - Precision = 0.4, Recall=0.8. Calculati masura F1?
- A. 0.53
  - B. 0.89
  - C. 0.64
  - D. 0.72
4. Care din urmatoarele activari sunt rezultatul aplicarii functiei de activare sigmoid?
- A. [[0.3, 0.1], [0.2, 0.8]]
  - B. [[0.01, -0.99], [0.99, 0.02], [0.7, 0.3]]
  - C. [[-1, 2], [0.4, 0.6]]
  - D. [1.1, 0.99]
5. Care este functia utilizata pentru generarea predictiilor (pasul forward) unei retele neuronale feedforward cu 2 straturi ascunse cu activare ReLU si un strat de iesire cu activare Softmax?
- A. softmax(2\*relu(x \* w + b)\*relu(x \* w + b))
  - B. softmax(relu(relu(x \* w1 + b1) \* w2 + b2) \* w3 + b3)
  - C. relu(softmax(x \* w + b) + |w|)
  - D. softmax(2\*relu(x \* w1 + b1))
6. Care este valoarea de iesire a perceptronului daca intrare=[11.4, -5.4], ponderi=[3.2, 0.5], bias=-23.3 (functia de activare - sign)?
- A. 1
  - B. 9.53
  - C. -9.52
  - D. -1
7. Care este output-ul unui model SVM cu ponderile  $W = [1, -0.5, 0.2, 0.9]$ ,  $b = -1.5$  pentru urmatoarele exemple  $X = [[0.5, -1.5, -2.5, 0.7], [-2.1, 5.3, -1.2, 6.4]]$ ?
- A. [-1, -1]
  - B. [1, -1]
  - C. [1, 1]
  - D. [-1, 1]
8. Care dintre urmatoarele afirmatii despre retele neuronale convolutionale este falsa?
- A. Pot fi folosite atat pe imagini, cat si pe date de tip text.
  - B. Automatizeaza extragerea informatiilor relevante.
  - C. Straturile de pooling miscoreaza dimensiunea activarilor, limitand cantitatea de informatie pierduta.
  - D. Straturile convolutionale sunt ne-antranabile, fiind alcătuite din filtre presețate de utilizator.
9. Care este forma duala utilizand kernel-ul liniar pe matricea  $X$ , daca forma primala este  $X = [[0.5, 2.7, 3], [4, 1.5, 6]]$ ?
- A.  $X_{\text{dual}} = [[16.54, 24.05], [24.55, 4.25]]$
  - B.  $X_{\text{dual}} = [[16.14, -24.05], [24.05, 59.25]]$
  - C.  $X_{\text{dual}} = [[16.54, 24.05], [24.05, 54.25]]$
  - D.  $X_{\text{dual}} = [[16.54, -2.05], [25.05, 54.25]]$

10. Cand este mai eficient sa folosim reprezentarea duala a datelor?
- A. Cand avem o problema de clasificare binara (cu doua clase)
  - B. Cand numarul de trasaturi este mai mic decat numarul de exemple
  - C. Cand avem o problema de clasificare cu foarte multe clase (mai mult de doua)
  - D. Cand numarul de trasaturi este mai mare decat numarul de exemple
1. Daca avem probabilitatile pentru evenimentele  $P(A)=0.2$ ,  $P(B)=0.3$ ,  $P(B|A)=0.5$ . Calculati probabilitatea  $P(A|B)$  folosind regula lui Bayes.
- A. 0.67
  - B. 0.42
  - C. 0.33
  - D. 0.51
2. Daca un clasificator are acuratete 1% pe o multime de antrenare cu 100 clase, cel mai probabil:
- A. Acurateata este potrivita
  - B. Este in regim optim de antrenare
  - C. Este in regim de suprainvatare
  - D. Nici una dintre variante
3. Care este numarul total de parametri ai unei relete neuronale feedforward cu urmatoarea configuratie: 3 neuroni pe statul de intrare, 2 straturi ascunse a cate 5 neuroni si 2 neuroni pe statul de iesire?
- A. 26
  - B. 65
  - C. 62
  - D. 50
4. Care din urmatoarele metode este folosita pentru a combate bias-ul modelului?
- A. F1-score
  - B. Early stopping
  - C. Alegerea unui model mai complex
  - D. Reducerea numarului de exemple
5. Care din urmatoarele clasificatoare au o decizie liniara?
- A. MLP fara functii de transfer
  - B. Naive Bayes
  - C. SVM cu kernel RBF
  - D. KNN
6. Care dintre aceste optiuni pot sa ajute in imbunatatirea capacitatii de generalizarea unui model de invatare automata?
- A. Continuarea antrenarii pana se ajunge la overfitting
  - B. Micsorarea setului de antrenare
  - C. Adaugarea unui termen de regularizare
  - D. Cresterea numarului de parametri antrenabili ai modelului

7. Cati clasificatori binari vor fi antrenati pentru o problema multi-class cu 4 clase folosind schema de combinare one-vs-all?

- A. 4
- B. 1
- C. 32
- D. 16

8. Fiind urmatoarele puncte  $[(2, 5), (6, 5), (3, 0), (3, 1), (1, 5), (2, 3)]$  si etichetele corespunzatoare  $[2, 2, 1, 0, 1, 1]$ . Folosind metoda kNN,  $k=3$  si distanta Euclidea, care este eticheta prezista pentru punctul  $(2, 3)$ ?

- A. 1
- B. 2
- C. 0
- D. Nu se poate determina

9. Care este valoarea de iesire a perceptronului daca intrare= $[11.4, -5.4]$ , ponderi= $[3.2, 0.5]$ , bias=-23.3 (functia de activare - sign)?

- A. -1
- B. -9.52
- C. 9.53
- D. 1

10. Fiind date etichetele  $y = [23, 14, 30, 45, 18, 31]$  si predictiile aferente  $p = [26, 20, 39, 38, 18, 33]$ , care este masura Kendall Tau?

- A. 0.4
- B. 1.0
- C. 0.2
- D. 0.6

1. Cati clasificatori binari vor fi antrenati pentru o problema multi-class cu 4 clase folosind schema de combinare one-vs-all?

D. 4 =====> D

2. Cati clasificatori binari vor fi antrenati pentru o problema multi-class cu 4 clase folosind schema de combinare one-vs-one?

A. 6 =====> A

3. Daca valoarea functiei cost este aproximativ la fel pentru setul de testare cat si pentru cel de antrenare, cel mai probabil:

C. Este in regim optim de antrenare =====> C

4. Care este output-ul unui perceptron cu ponderile  $W = [0.5, 0.3, 0.1]$ ,  $b = -0.5$ , functia de activare semn pe datele de intrare  $X = [[-3, -2, 1], [2, 0.5, 0.1]]$ ?

D.  $[-1, 1]$  =====> D

5. Care este forma duala utilizand kernel-ul liniar pe exemplul de test  $X_{\text{test}} = [-0.5, 0.4, -0.5]$ , daca  $X_{\text{train}} = [[-0.3, 2, 0.3], [-1, -2.6, -4], [0.5, -1, -0.5], [-3, -4, -2.7]]$ ?  $X_{\text{train}}$  si  $X_{\text{test}}$  sunt in forma primala.

D.  $X_{\text{test\_dual}} = [0.8, 1.46, -0.4, 1.25]$  =====> D

6. Care dintre urmatoarele metode corespunde unei metodologii de invatare nesupervizata?  
D. K-means clustering =====> D

7. Care sunt valorile urmatoarelor exemple de antrenare si testare dupa ce se aplica  
normalizarea standard (standardizarea),  $X_{\text{train}} = [[10, 5, 3], [-1, -2, -30], [-0.5, 0.5, 0.3]]$ ,  
 $X_{\text{test}} = [[0.5, -1, 7], [1.2, 3.4, 5.6]]$ ?

D.  $X_{\text{train}} = [[1.41, 1.32, 0.79], [-0.75, -1.09, -1.41], [-0.65, -0.23, 0.61]]$ ;  $X_{\text{test}} = [[-0.46,$   
 $-0.74, 1.06], [-0.32, 0.77, 0.96]]$  =====> D

9. Ce dimensiune spatiala va avea un activation map dupa aplicarea unui strat de pooling de  
11x11 cu stride=4 pe o intrare de 227x227?

B. 55x55 =====> B

10. Care din urmatoarele functii este potrivita pentru masurarea performantei unui  
clasificator aplicat pe date cu distributie balansata?

B. Acuratetea =====> B

1. In cazul analizei liniar discriminante, hiperplanul pe care se proiecteaza punctele este:  
B. Orientat astfel incat punctele sa fie cat mai appropriate =====> B

2. Care din urmatoarele metode este folosita pentru a combatte bias-ul modelului?

D. Alegerea unui model mai complex =====> D

3. Catii clasificatori binari trebuie antrenati pentru a rezolva o problema cu K clase folosind  
metoda One vs One?

C.  $(K * (K - 1)) / 2$  =====> C

4. Care din urmatorii algoritmi se foloseste pentru optimizare ponderilor retelelor neuronale?

C. SGD =====> C

5. Catii clasificatori binari vor fi antrenati pentru o problema multi-class cu 4 clase folosind  
schema de combinare one-vs-one?

C. 6 =====> C

6. Care din urmatoarele functii de activare aduce output-ul in intervalul (-1, 1)?

C. Tangenta Hiperbolica =====> C

7. Care este matricea de confuzie a unui clasificator, pentru setul de test cu etichetele  
[1,2,3,1,1,2,2,1,3,2] si predictiile [2,1,3,1,1,1,2,2,2]?

A.  $[[2 \ 2 \ 0], [3 \ 1 \ 0], [0 \ 1 \ 1]]$  =====> A

8. Care este cel mai probabil efect pe care il putem obtine pentru modelul de regresie ridge  
cand setam parametrul de regularizare alpha cu o valoarea foarte mica?

B. Model se comporta ca un model de regresie liniara =====> B

9. Care este scufundarea asociata functie nucleu liniare?

B.  $f(x) = x$  =====> B

10. Daca un model antrenat pe 12 clase are acuratete 97% pe multimea de testare, cel mai probabil:

A. Este in regim de overfitting =====> A

Model majeri:

1. knn cu dist manhattan
2. bow (bag of words pentru un dictionar)
3. aplicare min-max pe  $[[0.1, 0.4], [0.2, 0.5], [0.3, 0.6]]$  (3 examples, 2 features)
4. cati neuroni are layerul ascuns cu un singur layer ascuns: cuv cheie: input layer, number of features
5. 4-6 2-1 dimensiunea celui de al doilea layer neuronal
6. e-mai spam classification task - performanta cea mai buna
7. clasificator linear
8. recall pentru un clasificator unde  $y = [0, 1, 1, 0, 0, 0, 0, 1]$
9. valid loss pentru retea neuronala cu gradient descendente
10. forma hartii activate 2x2 max pooling