# 1. Înțelegerea Word2Vec (15 puncte)

(a) (2 puncte)  
Întrucât doar un singur termen y\_o este diferit de zero (egal cu 1), entropia încrucișată se simplifică la -log(ŷ\_o). Astfel, expresia completă -∑\_w y\_w log(ŷ\_w) devine echivalentă cu -log(ŷ\_o).

(b) (6 puncte)  
i. Derivata funcției de pierdere în raport cu v\_c este: ∂J/∂v\_c = U(ŷ - y).  
  
ii. Gradientul devine nul atunci când distribuția prezisă ŷ corespunde exact distribuției reale y.  
  
iii. Termenul Uŷ împinge v\_c spre predicțiile greșite, în timp ce Uy atrage v\_c spre vectorul corect u\_o. Actualizarea v\_c pe baza diferenței dintre aceste două forțe îl orientează spre cuvântul corect și îl îndepărtează de celelalte.

(c) (1 punct)  
Normalizarea L2 elimină informația despre amplitudinea vectorilor. Dacă doi vectori sunt proporționali (ex: u\_x = αu\_y), normalizarea îi face identici, ceea ce poate duce la pierderea unor informații semnificative, cum ar fi intensitatea unei trăsături.

(d) (5 puncte)  
Pentru cuvântul corect (w = o): ∂J/∂u\_o = (ŷ\_o - 1) \* v\_c.  
Pentru celelalte cuvinte (w ≠ o): ∂J/∂u\_w = ŷ\_w \* v\_c.  
În general, se poate exprima: ∂J/∂u\_w = (ŷ\_w - y\_w) \* v\_c.

(e) (1 punct)  
Gradientul față de matricea U este dat de: ∂J/∂U = v\_c (ŷ - y)^T.

# 2. Învățare automată și Rețele Neuronale (8 puncte)

(a) (4 puncte)  
i. Impulsul (momentum) ajută la netezirea traiectoriei de învățare, reducând oscilațiile și accelerând convergența prin integrarea gradientelor trecute.  
  
ii. Când gradientul unui parametru este mic, impulsul (v\_{t+1}) mic duce la actualizări relativ mai mari, permițând tuturor parametrilor să evolueze într-un mod echilibrat.

(b) (4 puncte)  
i. Coeficientul de scalare γ trebuie să fie 1/(1 - p\_drop) pentru a păstra aceeași valoare așteptată a ieșirii după aplicarea dropout-ului.  
  
ii. Dropout-ul este folosit exclusiv în timpul antrenamentului pentru a preveni supraînvățarea. La evaluare, dropout-ul este dezactivat pentru a obține predicții consistente.

# 3. Parsing Tranzițional Bază pe Rețele Neuronale (54 puncte)

(a) (4 puncte)  
Configurație inițială:  
- Stivă: ['ROOT']  
- Buffer: ['I', 'presented', 'my', 'findings', 'at', 'the', 'NLP', 'conference']  
  
Pași de parsare:  
SHIFT ➔ ['ROOT', 'I']  
SHIFT ➔ ['ROOT', 'I', 'presented']  
LEFT-ARC (presented → I) ➔ ['ROOT', 'presented']  
SHIFT ➔ ['ROOT', 'presented', 'my']  
SHIFT ➔ ['ROOT', 'presented', 'my', 'findings']  
LEFT-ARC (findings → my)  
RIGHT-ARC (presented → findings)  
SHIFT ➔ ['ROOT', 'presented', 'at']  
SHIFT ➔ ['ROOT', 'presented', 'at', 'the']  
SHIFT ➔ ['ROOT', 'presented', 'at', 'the', 'NLP']  
SHIFT ➔ ['ROOT', 'presented', 'at', 'the', 'NLP', 'conference']  
RIGHT-ARC (conference → NLP)  
LEFT-ARC (NLP → the)  
LEFT-ARC (the → at)  
RIGHT-ARC (at → presented)  
RIGHT-ARC (presented → ROOT)

(b) (2 puncte)  
O propoziție cu n cuvinte necesită exact 2n pași pentru parsare: n pași SHIFT și n pași ARC (LEFT sau RIGHT).

(f) (12 puncte)  
1. Tip de eroare: Eroare de atașare a frazei prepoziționale  
 - Dependență incorectă: risks → involved  
 - Dependență corectă: risks → citing  
  
2. Tip de eroare: Eroare de atașare a unui modificator  
 - Dependență incorectă: crucial → most  
 - Dependență corectă: elements → most  
  
3. Tip de eroare: Eroare de atașare a frazei verbale  
 - Dependență incorectă: comment → declined  
 - Dependență corectă: comment → to  
  
4. Tip de eroare: Eroare de atașare în coordonare  
 - Dependență incorectă: plants → in  
 - Dependență corectă: one → in

(g) (2 puncte)  
Etichetele de parte de vorbire (POS tags) contribuie la dezambiguizarea rolurilor sintactice ale cuvintelor (ex: 'book' ca substantiv sau verb), îmbunătățind acuratețea predicțiilor de tranziție.