

Deep Learning – homework 3

- התרגיל יבוצע בזוגות. את התוצרים יעלה רק אחד מבני הזוג בציון שמות + ת.ז. של שותפי ההגשה.
 - תוצר התרגיל הוא מסמך PDF / WORD ויועלה ל MOODLE
 - הפתרון המצטיין יפורסם לכלל הסטודנטים
 - ניתן להשתמש במודלי AI לצורך הפתרון. ניתן להיעזר בסטודנטים אחרים. **אין להעתיק**
 - ככל שיש צורך – הנח הנחות הרלוונטיות לפתרון
-

1.) Consider an LSTM network processing a sequence of length $T=5$:

a) Write the complete mathematical formulation for the forget gate (ft), input gate (it), cell state (Ct), and output gate (ot) at any time step t .

b) Explain mathematically how LSTM addresses the vanishing gradient problem that occurs in simple RNNs. Support your explanation with derivatives.

c) If we have a batch size of 32, input dimension of 100, and hidden state dimension of 64, calculate:

- The dimension of each weight matrix
- The total number of parameters
- The memory requirements during training

2.) For a CNN with the following architecture:

- Input: $32 \times 32 \times 3$
- Conv1: 16 filters of size 5×5 , stride 1
- MaxPool1: 2×2 , stride 2
- Conv2: 32 filters of size 3×3 , stride 1
- MaxPool2: 2×2 , stride 2
- Fully connected: 128 neurons
- Output: 10 classes

a) Calculate the output dimensions after each layer

b) Derive the total number of trainable parameters

c) If we use batch normalization after Conv1, write the mathematical equations for:

- Forward pass normalization
- Parameter update during backpropagation

3.) Given a training process with the following metrics:

- Training accuracy: 98%
- Validation accuracy: 85%
- Training loss: 0.02
- Validation loss: 0.4

a) For each of these scenarios, explain mathematically what's happening and propose solutions:

- Learning rate is too high
- Learning rate is too small
- Batch size is too high
- Batch size is too small

b) Derive the relationship between

- Precision and Recall
- F1 score and Accuracy