|  |
| --- |
| **Deep Learning for Perception(CS4045)** |
| Date: April 12th 2025 |
| **Course Instructor(s)** |
| Mr. Syed Irtaza Muzaffar |

|  |  |
| --- | --- |
| **Sessional-II Exam** | |
| **Total Time (Hrs):** | **1** |
| **Total Marks:** | **30** |
| **Total Questions:** | **3** |

|  |
| --- |
| \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_  Roll No Section Student Signature |

**Attempt all the questions.**

***CLO2 #: Understand and design the structure of deep neural networks.***

**Question 1 [4+4+2 marks]**

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

You are training a deep neural network using the ADAM optimizer. At time step τ=2, consider the following settings for a single weight wi​:

* Previous moment estimates:

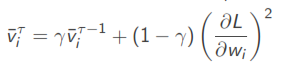
m̄ᵢ1 = 0.1, v̄ᵢ2=0.4

* Current gradient:
* Hyperparameters:

1. Compute the updated biased moment estimates m̄ᵢ²​ and v̄ᵢ² at time step τ=2.

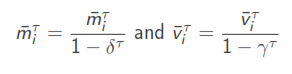


m̄ᵢ²=0.9(0.1) + (1-0.9)0.2 = 0.09 + 0.02= **0.11**



v̄ᵢ²=0.999 \* 0.4 + (1-0.999) \* (0.2)2 =0.3996+0.00004= **0.39964**

1. Compute the bias-corrected estimates m̄ᵢ²​ and v̄ᵢ²​.



m̄ᵢ²=0.11 / (1-0.9) = 1.1

v̄ᵢ²= 0.39964 / (1-0.999) = 399.64

**c)** Given the previous weight wi1=0.5, compute the updated weight wi2​.



wi2 = 0.5 – (10-3 \* 1.1) / 19.9909+10-8 =0.5 – 0.0011/19.9909=0.5 - 0.00005502751=0.4999

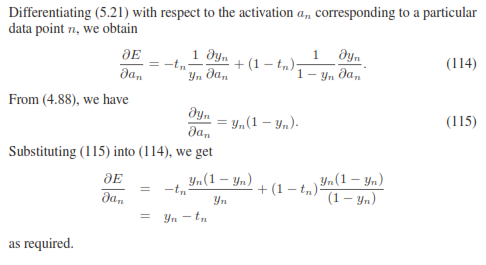
***CLO3 #: Understand the different layers and their operations.***

**Question 2 [3+3+2+2 marks]**

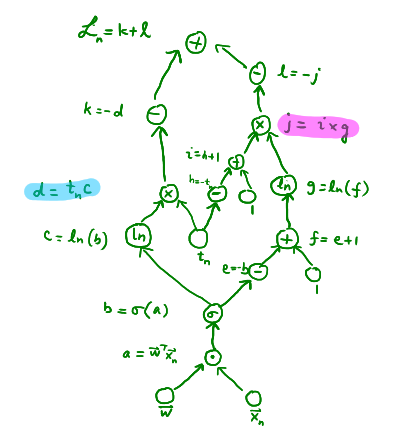
**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

1. Show the derivative of the error function (equ 1) with respect to the activation ak for an output unit having a logistic sigmoid activation function satisfies (equ 2).

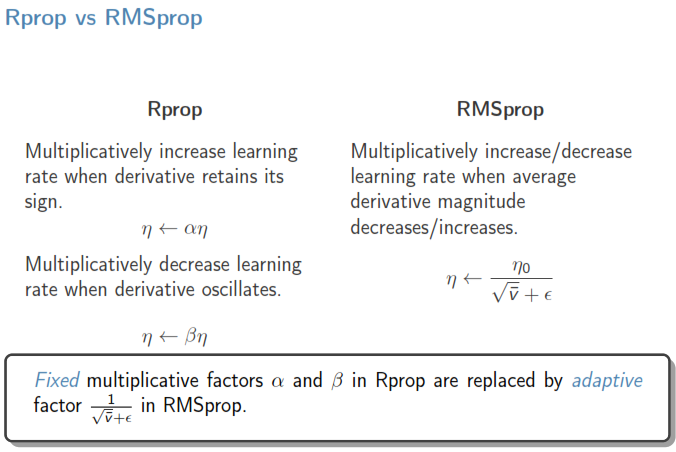
Solution:



1. Make a computation graph for the logistic regression loss function via automatic differentiation.



1. Difference between RMSprop and Rprop.



1. A CNN uses 5x5 filter, stride=1, padding=2. Input image = 64x64. Compute output dimensions. Repeat with stride=2, padding=0.

***CLO1 #: Understand the theoretical foundations of deep learning, including neural networks and***

***optimization techniques.***

**Question 3: Write the answer of objective part on sheet. [10 marks]**

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

1. \_Adam\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ optimizer combines momentum and adaptive learning rates?
2. Rprop converges much faster than gradient descent.[**True**/False]
3. In Rprop, each direction is handled dependently. [True/**False**]
4. Quickprop approximates \_\_\_\_\_\_\_\_ Second-order \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_derivative?
5. The primary issue with vanilla gradient descent is\_\_\_\_\_\_\_\_\_\_ Overfitting \_\_\_\_\_\_\_\_\_\_\_\_\_.
6. Label smoothing helps reduce \_\_OverFitting\_ by distributing probability mass across classes.
7. \_\_ Dropout \_\_\_\_\_\_\_\_ regularization technique randomly deactivates neurons during training?
8. Which optimization method uses second-order derivatives?  
   (a) SGD (b) Newton's Method (c) RMSProp (d) Adam
9. Data augmentation regularization method modifies the dataset by transforming input samples?
10. Padding in CNNs allows:  
    (a) Increased feature dimensions (b) **Preserving input size** (c) More pooling (d) None of the above