Assignment Details

202510-Spring 2025-ECGR-4127-001:ECGR-5127-001_Combined

To Submit:

- Two python files with the code portion (hw1_complete.py and person.py) of the assignment.
- A pdf file with your work for the analysis problems.

Analysis Problems

1. Consider a fully-connected network with 120 inputs, two hidden layers with 256 units each, and 10 output classes. Assume that all weights and activations are stored as 8-bit values.
   A. Find the total parameter storage required in bytes.
   B. How many MACs (multiply-and-accumulate) are required to run one inference with this model?
   C. How much temporary storage (SRAM) is required to run this model? Remember you'll need to store the outputs from one layer and the outputs from the next layer at the same time. The consecutive pair of layers with the largest combined requirements sets the amount of SRAM required.
2. Consider a fully-connected network with 1280 inputs, two hidden layers with 512 units each, and 32 output classes. Assume that all weights and activations are stored as 32-bit values.
   A. How much parameter storage is required for this model?
   B. Running this model on an 80MHz processor that can compute 1 MAC every 4 cycles, how long will one inference take (answer in ms).
   C. How much temporary storage (SRAM) is required to run this model?

A

Input layer = $120 \times 256 + 256 = 30,976$

Hidden layer 1 to hidden layer 2 = $256 \times 256 + 256 = 65,792$

hidden layer 2 to output = $256 \times 10 + 10 = 2,570$

$30,976 + 65,792 + 2,570 = 99,838$ bytes

B.

Input layer $120 \times 256 = 30,720$

hidden layer 1 to hidden layer 2 = $256 \times 256 = 65,536$

hidden layer 2 to output = $256 \times 10 = 2,560$

$30,720 + 65,536 + 2,560 =$

98816

C.

Input layer 1 $120 + 256 = 376$

hidden1 - hidden2 $256 + 256 = 512$

hidden 2 + output $256 + 10 = 266$

512 bytes

2

A. Input layer $1280 \times 512 + 512 = 655,872$
Hidden 1 – hidden 2 $= 512 \times 512 + 512 = 262,656$
hidden 2 - output $512 \times 32 + 32 = 16,416$
$655,872 + 262,656 + 16,416 = 934,944$

$934,944 \times 4 = 3,739,776$ bytes

B Input to hidden $128 \times 512 = 655,360$
hidden 1 – hidden 2 $512 \times 512 = 262,144$
hidden 2 - output $512 \times 32 = 16,384$

$655,360 + 262,144 + 16,384 = 933,888$

$(4/80) \times 933,888 = 46,694.4$ ms

C. Input – hidden $1280 + 512 = 1,792$
hidden 1 - hidden 2 $512 + 512 = 1,024$
hidden layer 2 - output $512 + 32 = 544$

$1792 \times 4 = 7168$ bytes