Carnegie Mellon University
Electrical & Computer Engineering

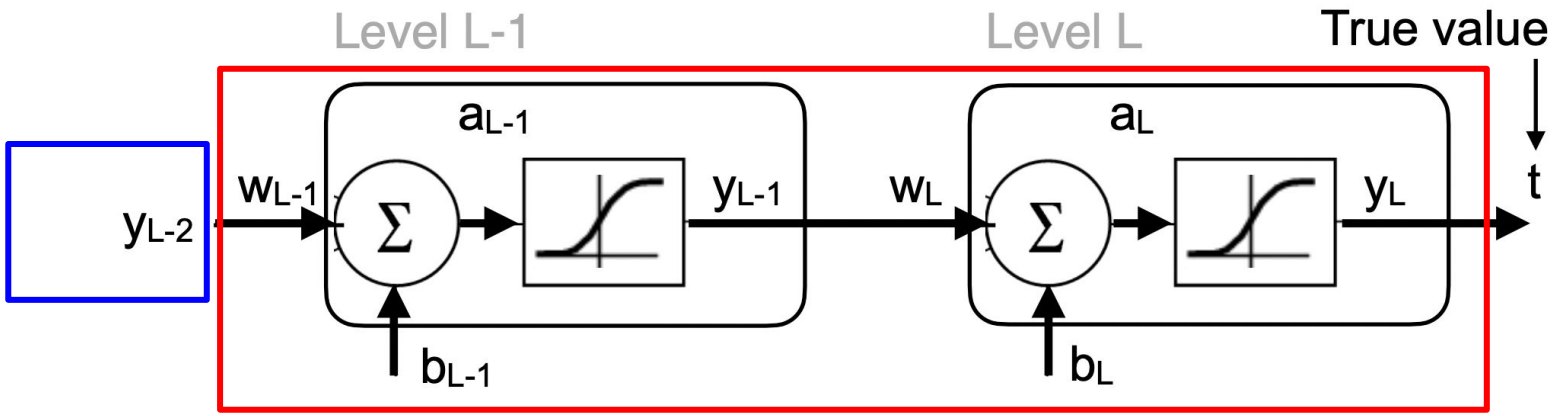
Recitation # 3
Optimization

February 2, 2024

Weiran Lin

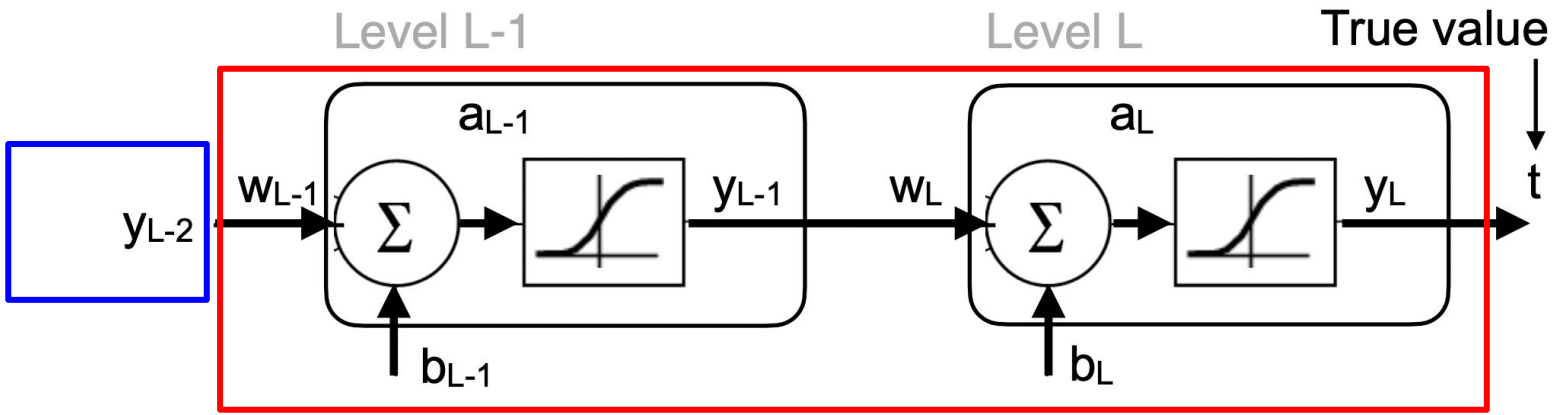
Recap of adversarial attacks:

Fixed model weights and update inputs



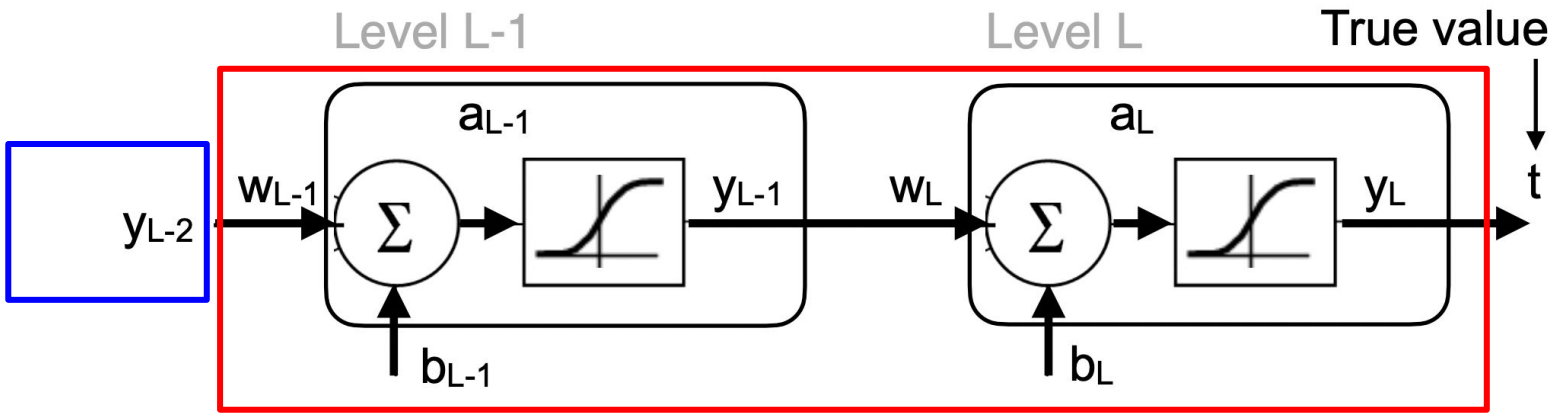
Recap of adversarial attacks:

Fixed model weights and update inputs



- What would happen if we select a wrong class t',
  - and change the inputs?

Recap of adversarial attacks:
Fixed model weights and update inputs



- What would happen if we select a wrong class t',
  - and change the inputs?
  - What would happen if we maximize instead of minimizing the loss function, while still using the true value as t?

Recap of L^p norms:

Why we need them in adversarial attacks?

$$\|x\|_p = \left(|x_1|^p + |x_2|^p + \cdots + |x_n|^p\right)^{1/p}$$

Recap of L^p norms:

Why we need them in adversarial attacks?

$$\|x\|_p = \left(|x_1|^p + |x_2|^p + \cdots + |x_n|^p\right)^{1/p}$$

- Q: If an image has a size of (32,32,3), what is n?

Recap of L^p norms:

Why we need them in adversarial attacks?

$$\|x\|_p = \left(|x_1|^p + |x_2|^p + \cdots + |x_n|^p\right)^{1/p}$$

- Q: If an image has a size of (32,32,3), what is n?
  - n=32*32*3

Recap of L^p norms:

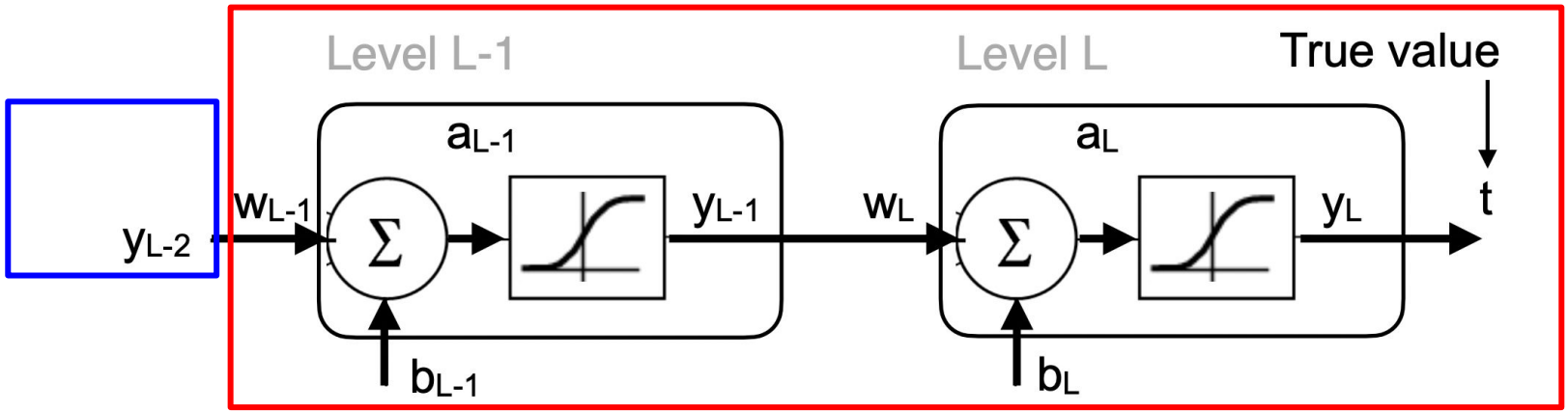
Why we need them in adversarial attacks?

$$\|x\|_p = \left(|x_1|^p + |x_2|^p + \cdots + |x_n|^p\right)^{1/p}$$

- Q: If an image has a size of (32,32,3), what is n?
  - n=32*32*3

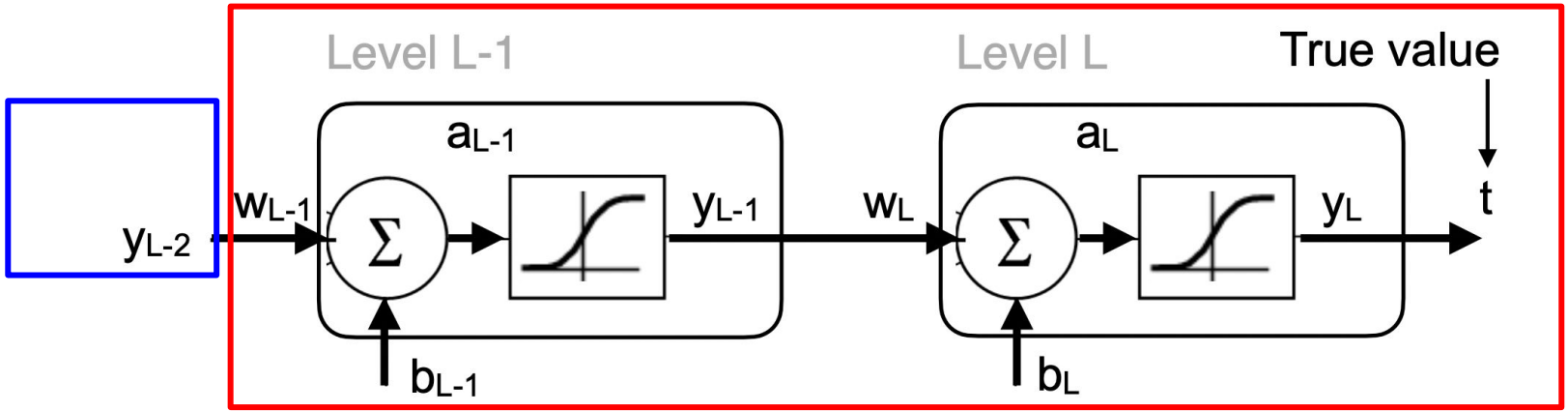
$$\|x\|_\infty = \max\left\{|x_1|, |x_2|, \ldots, |x_n|\right\}$$

Homework 2 tips



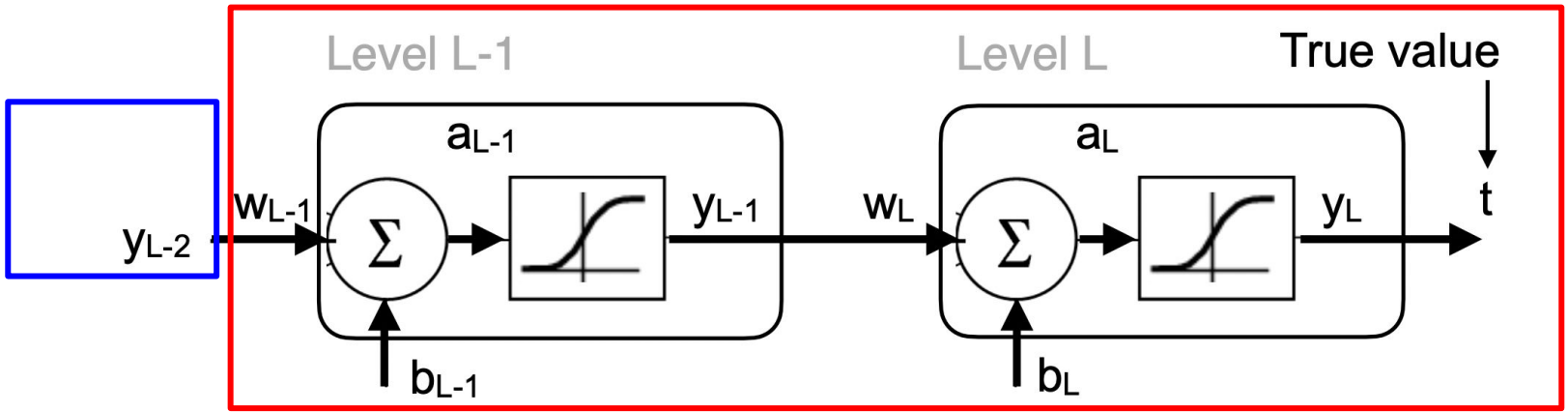
Update the model weights respect to the inputs

Homework 2 tips



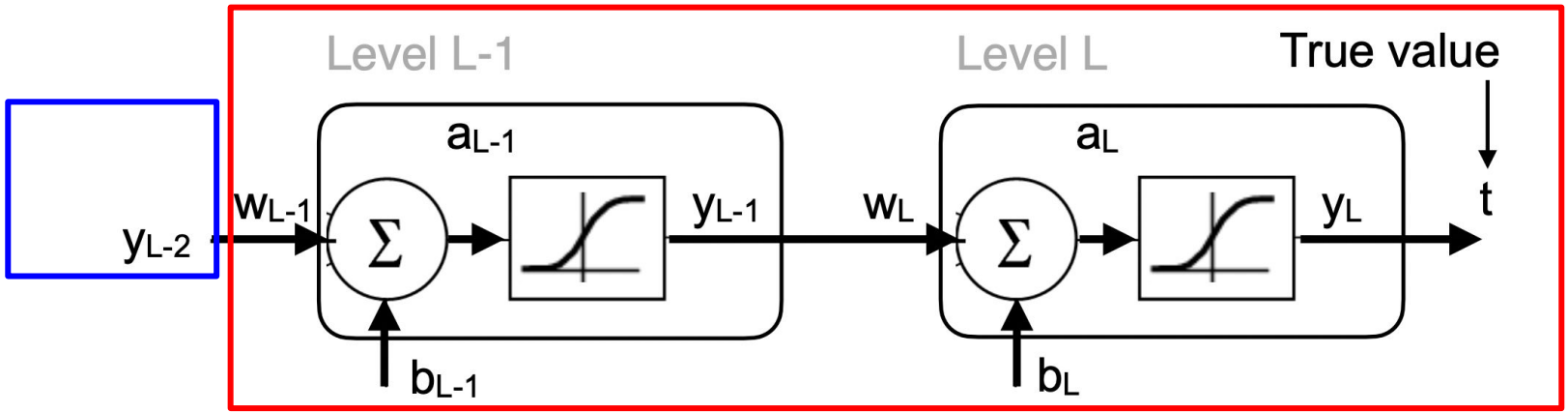
- Each y,a,b is a vector, and each w is a matrix,
  - If y_{L-2} has length M_0, and b_{L-1} has length M_1,
    - What is the length of a_{L-1}? What is the length of y_{L-1}?

Homework 2 tips



- Each y,a,b is a vector, and each w is a matrix,
  - If y_{L-2} has length M_0, and b_{L-1} has length M_1,
    - What is the length of a_{L-1}? What is the length of y_{L-1}?
    - What is the size of W_{L-1}?

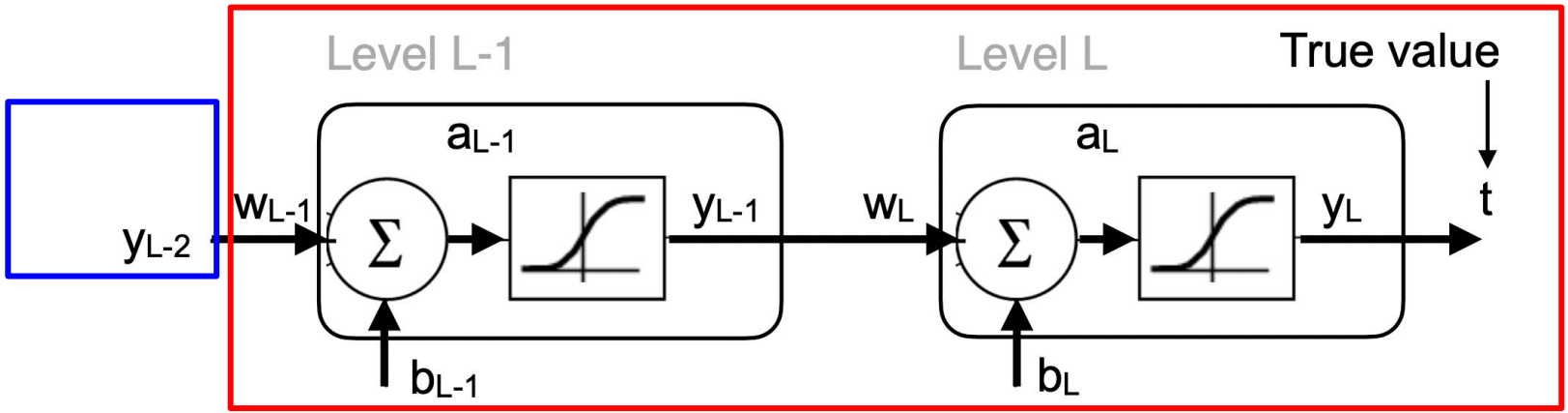
Homework 2 tips



Now we know y_{L-1} has length M_1

- If there are N classes
  - What is the length of a_L? What is the length of y_L?What is the length of b_L?

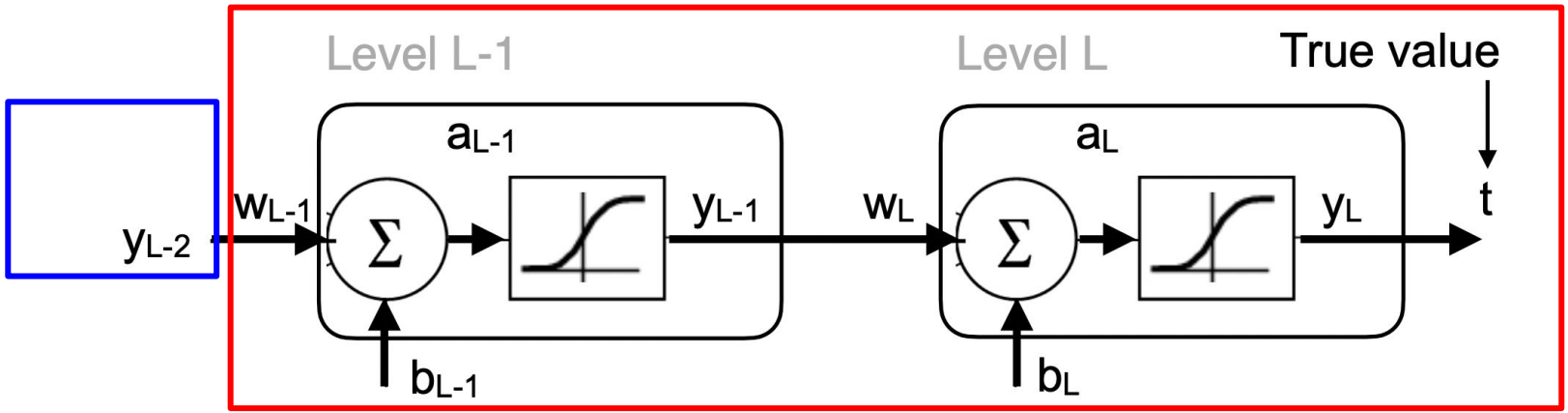
Homework 2 tips



Now we know y_{L-1} has length M_1

- If there are N classes
  - What is the length of a_L? What is the length of y_L?What is the length of b_L?
  - What is the size of W_L?

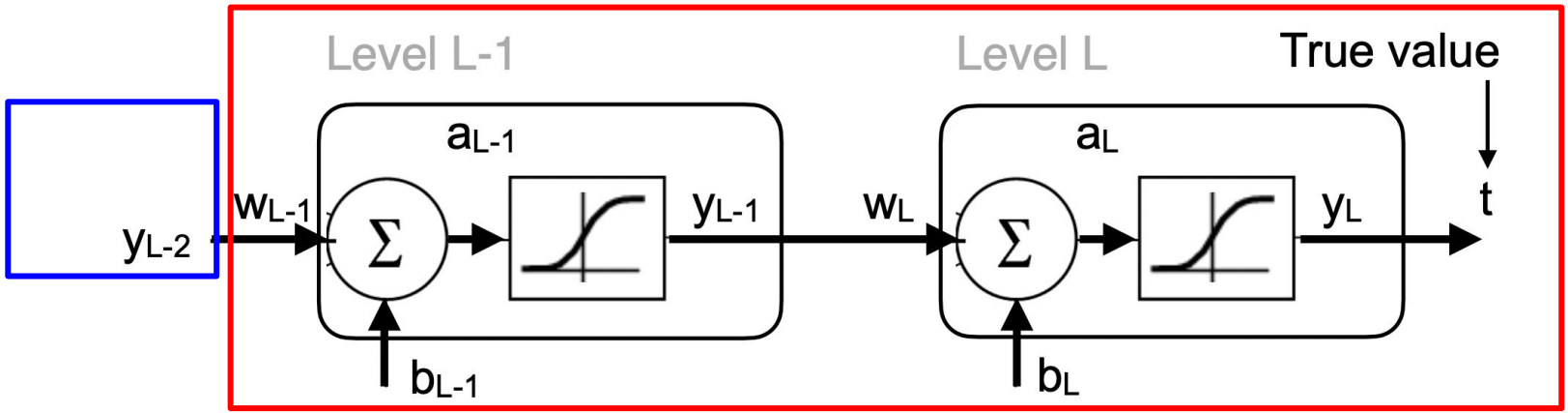
Homework 2 tips



Now we know y_{L-1} has length M_1

- If there are N classes
  - What is the length of a_L? What is the length of y_L?What is the length of b_L?
  - What is the size of W_L?
  - What is the size of the Loss E?

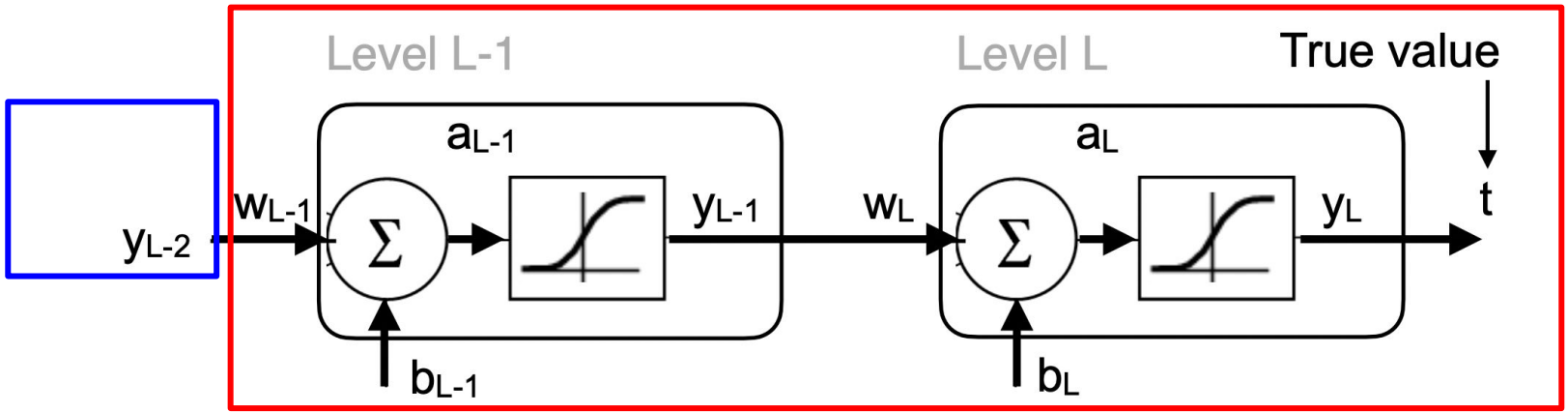
Homework 2 tips



Now we know y_{L-2} has length M_0, y_{L-1} has length M_1 and there are N classes,

- What is the size of dE/dy_L? What is the size of dE/dy_{L-1}?What is the size of dE/dy_{L-2}?

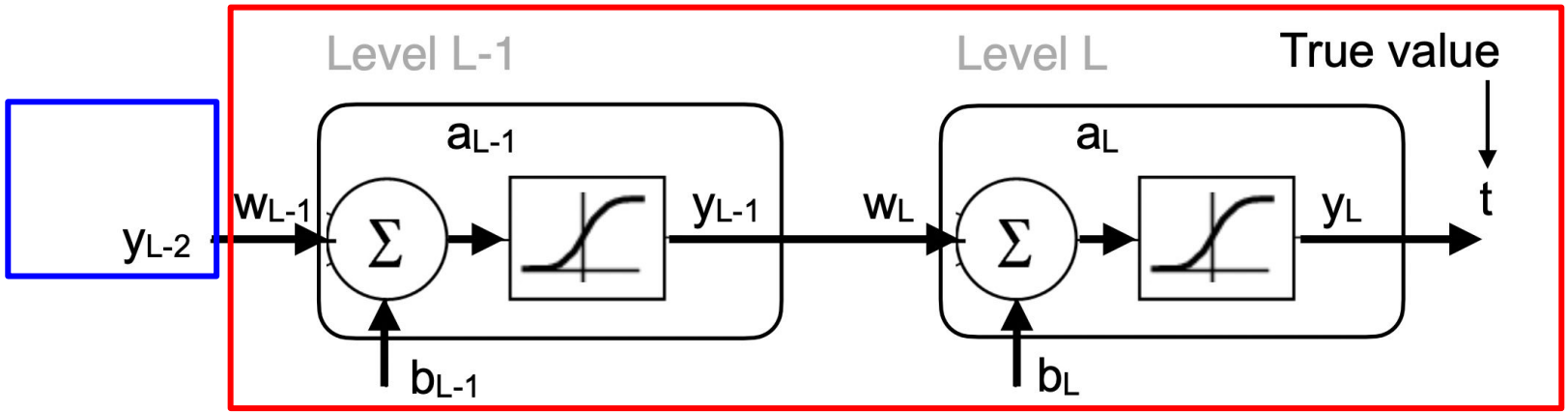
Homework 2 tips



Now we know y_{L-2} has length M_0, y_{L-1} has length M_1 and there are N classes,

- What is the size of dE/dy_L? What is the size of dE/dy_{L-1}?What is the size of dE/dy_{L-2}?
- What is the size of dE/dW_L? What is the size of dE/dW_{L-1}?

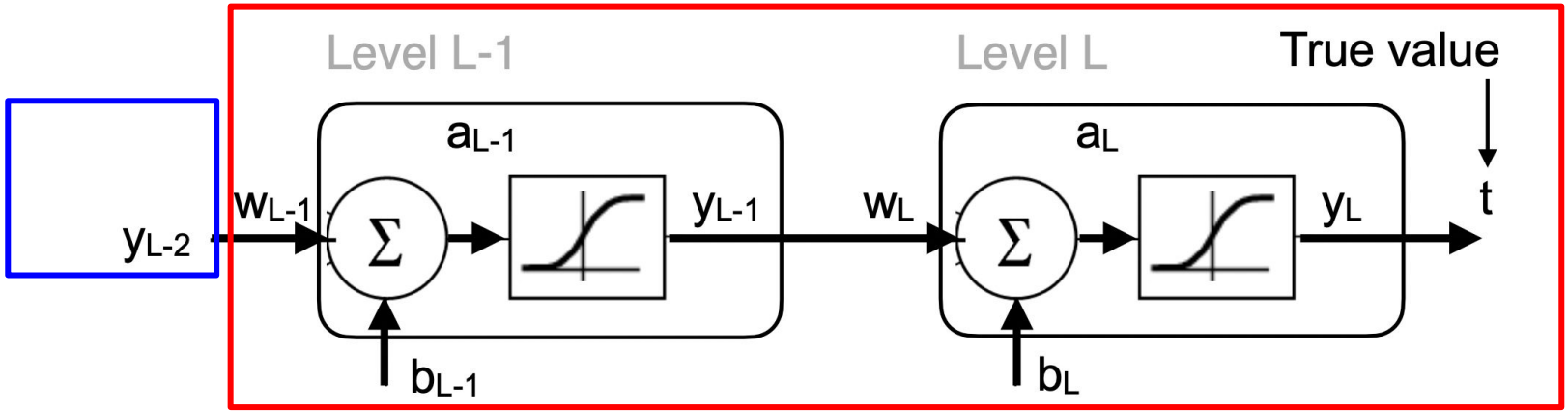
Homework 2 tips



Now we know y_{L-2} has length M_0, y_{L-1} has length M_1 and there are N classes,

- What is the size of dE/dy_L? What is the size of dE/dy_{L-1}?What is the size of dE/dy_{L-2}?
- What is the size of dE/dW_L? What is the size of dE/dW_{L-1}?
- What is the size of dE/da_L? What is the size of dE/da_{L-1}?

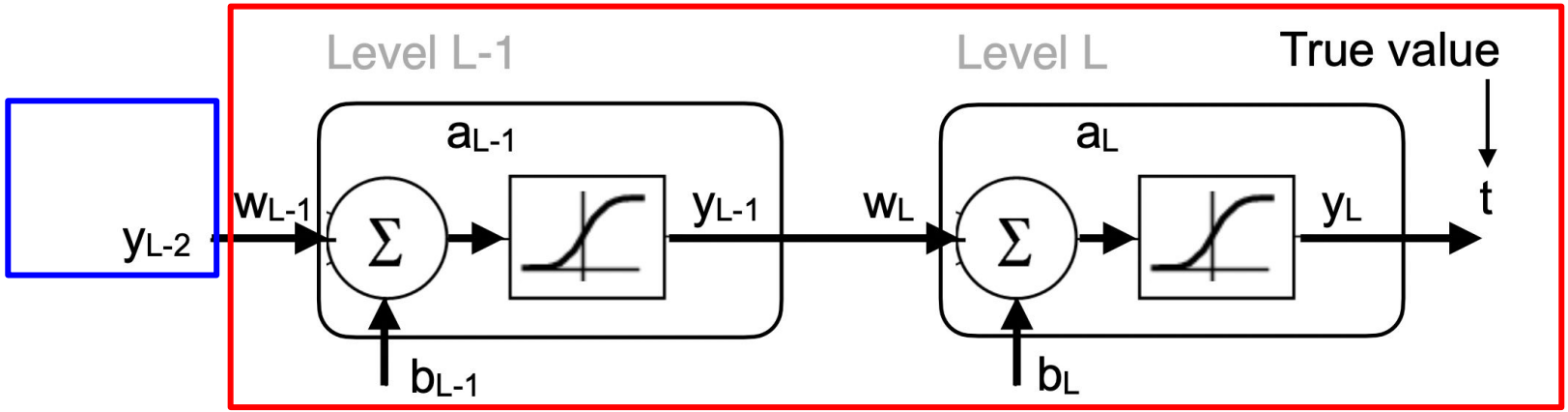
Homework 2 tips



Now we know y_{L-2} has length M_0, y_{L-1} has length M_1 and there are N classes,

- What is the size of dE/dy_L? What is the size of dE/dy_{L-1}?What is the size of dE/dy_{L-2}?
- What is the size of dE/dW_L? What is the size of dE/dW_{L-1}?
- What is the size of dE/da_L? What is the size of dE/da_{L-1}?
- What is the size of dE/db_L? What is the size of dE/db_{L-1}?

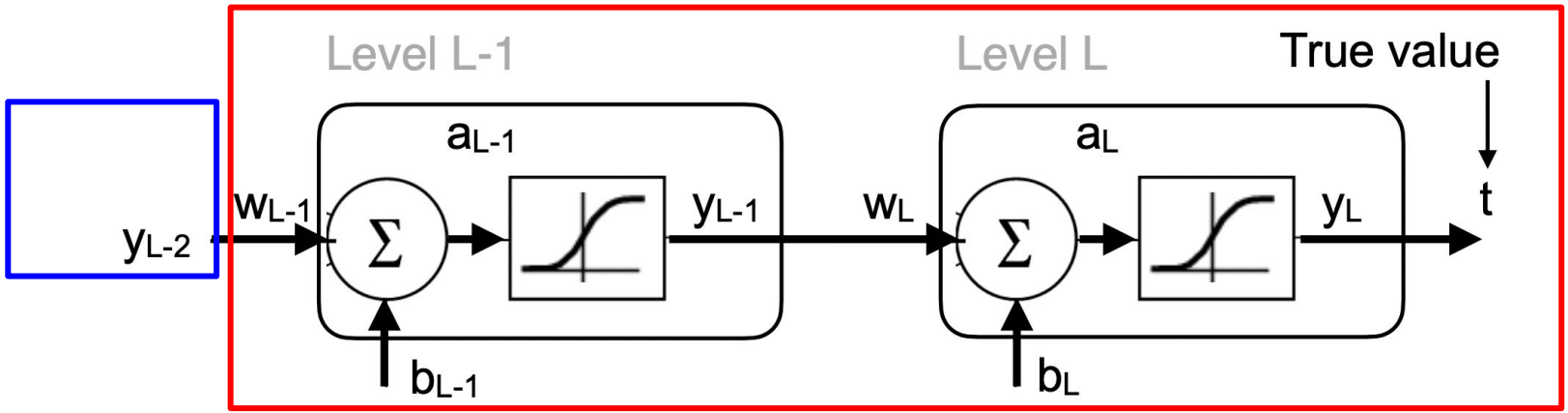
Homework 2 tips



What we expect you to implement:

1. Setup: initialize all w and b according to given dimensions (M and N)

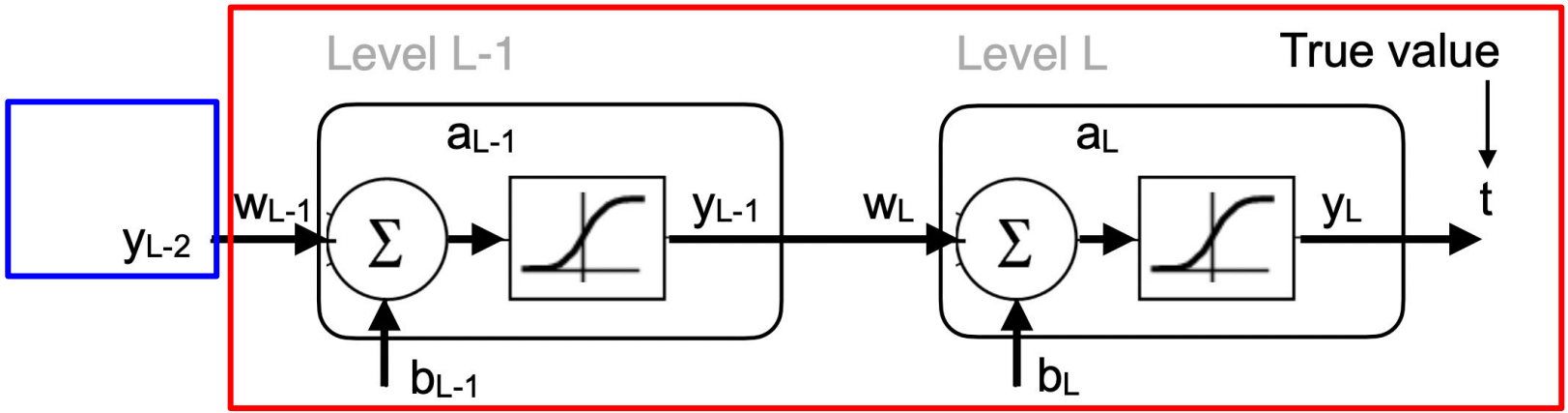
Homework 2 tips



What we expect you to implement:

1. Setup: initialize all w and b according to given dimensions (M and N)
2. Forward pass: given y_{L-2}, compute y_{L-1}, y_L and loss function E

Homework 2 tips



What we expect you to implement:

1. Setup: initialize all w and b according to given dimensions (M and N)
2. Forward pass: given y_{L-2}, compute y_{L-1}, y_L and loss function E
3. Backward pass: given E and current w and b,
   a. Compute dE/dw and dE/db
   b. Update w and b accordingly.