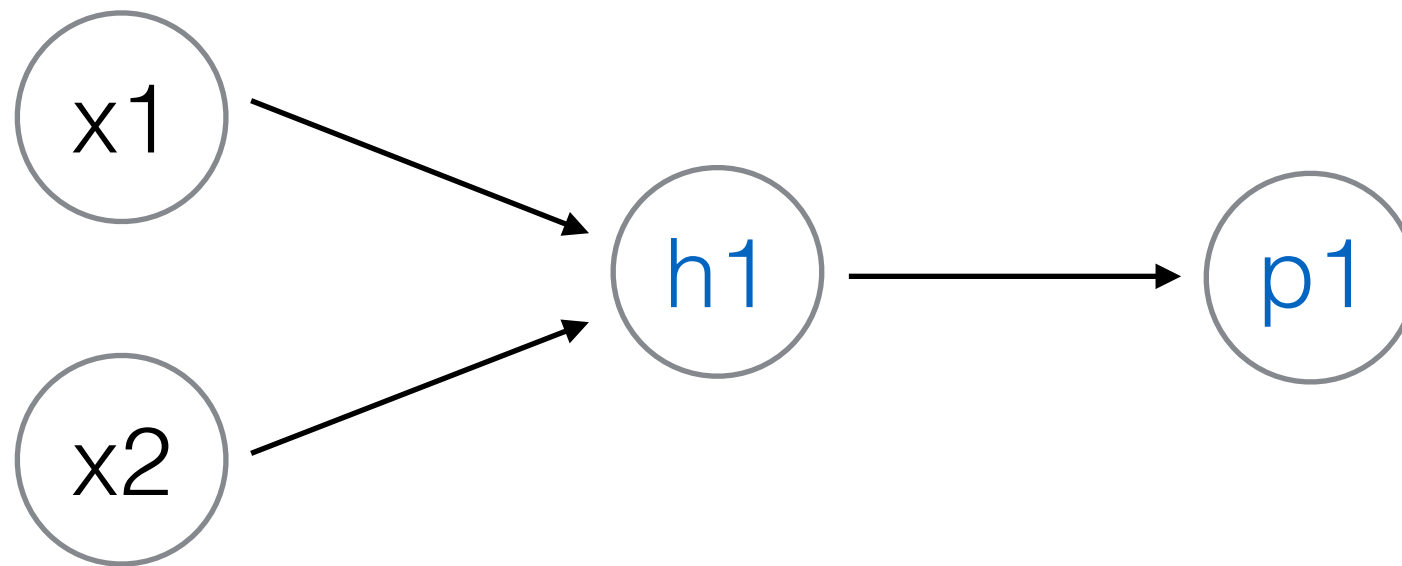


The Computation Graphs in CNN

[*https://github.com/clab/cnn*](https://github.com/clab/cnn)

ARK meeting (12/01/2015)
Lingpeng Kong

Logistic Regression

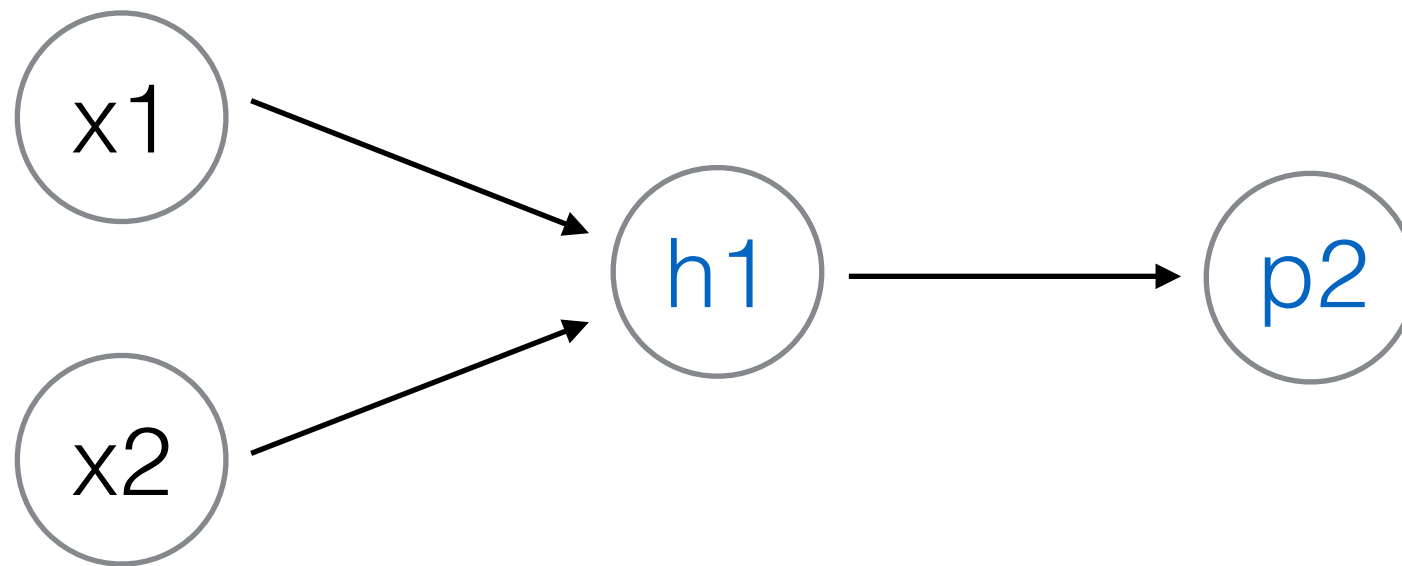


$$h_1 = w_1 * x_1 + w_2 * x_2 + b$$

$$p_1 = 1 / [1 + \exp(-h_1)]$$

note: p_1 — $p(y = 1|x_1, x_2)$

Logistic Regression



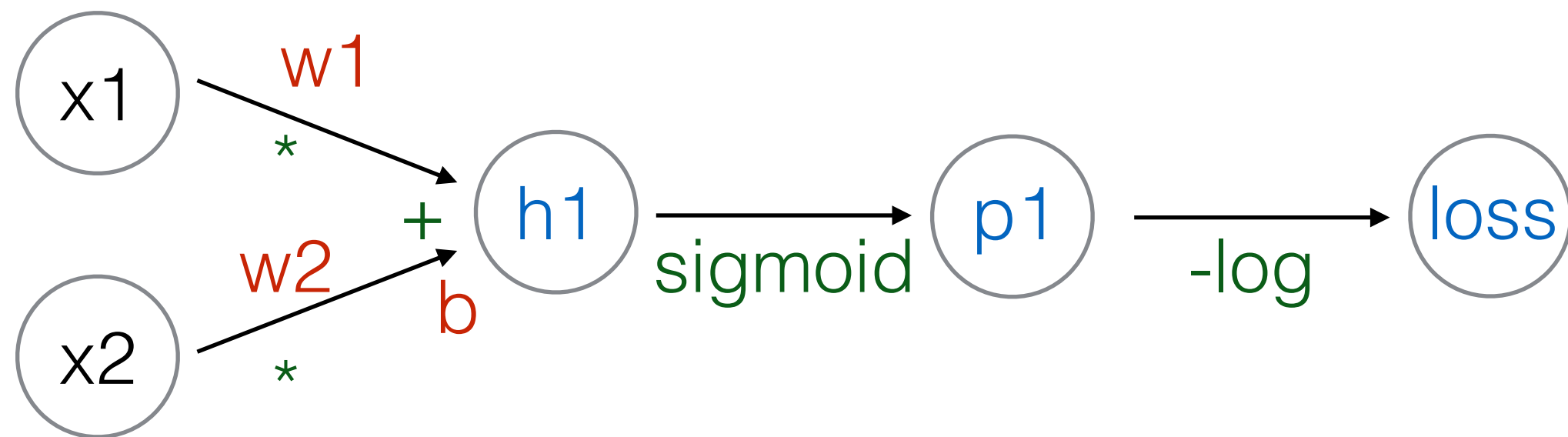
$$h1 = w1 * x1 + w2 * x2 + b$$

$$p2 = 1 - 1 / [1 + \exp(-h1)]$$

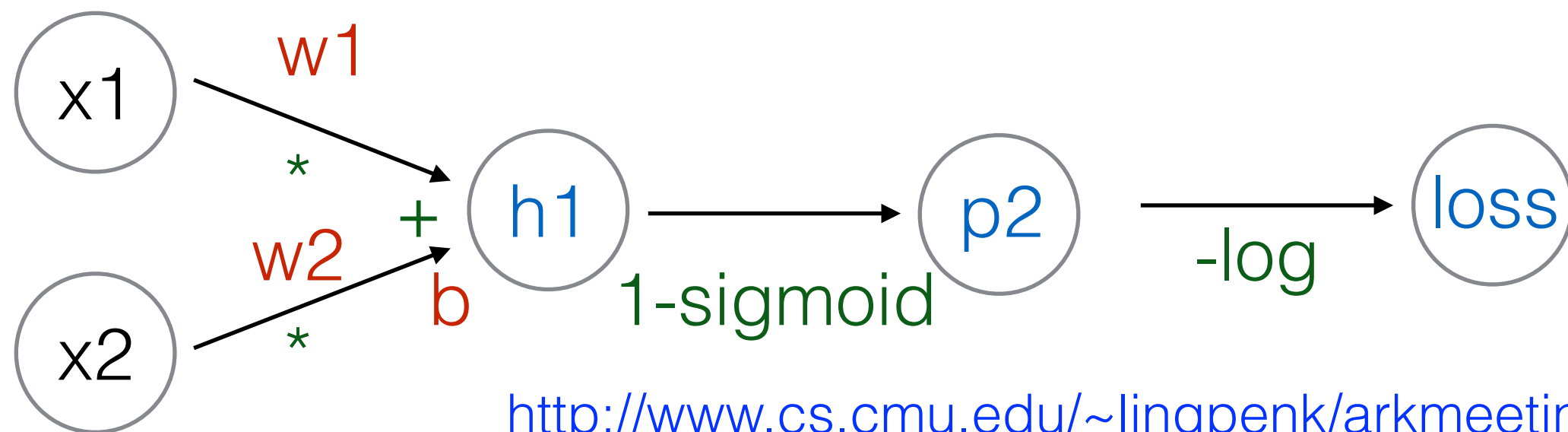
note: $p2 \text{ — } p(y = 0|x1, x2)$

Loss Function

$y = 1$



$y = 0$



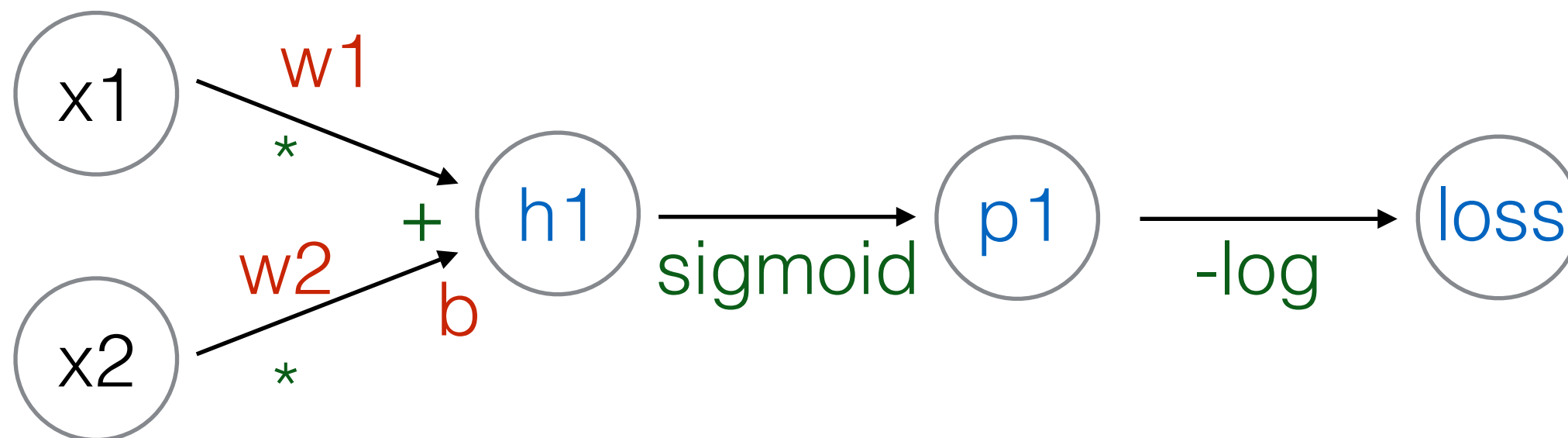
Computation Graph

Input x1 x2

Parameter w1 w2 b

Expression h1 p1 loss
Special

Operation * + sigmoid -log



How to minimize? (automatic differentiation)

Neural networks are our friends — Wang Ling

<http://www.cs.cmu.edu/~lingwang/lectures/deepnn.pdf>

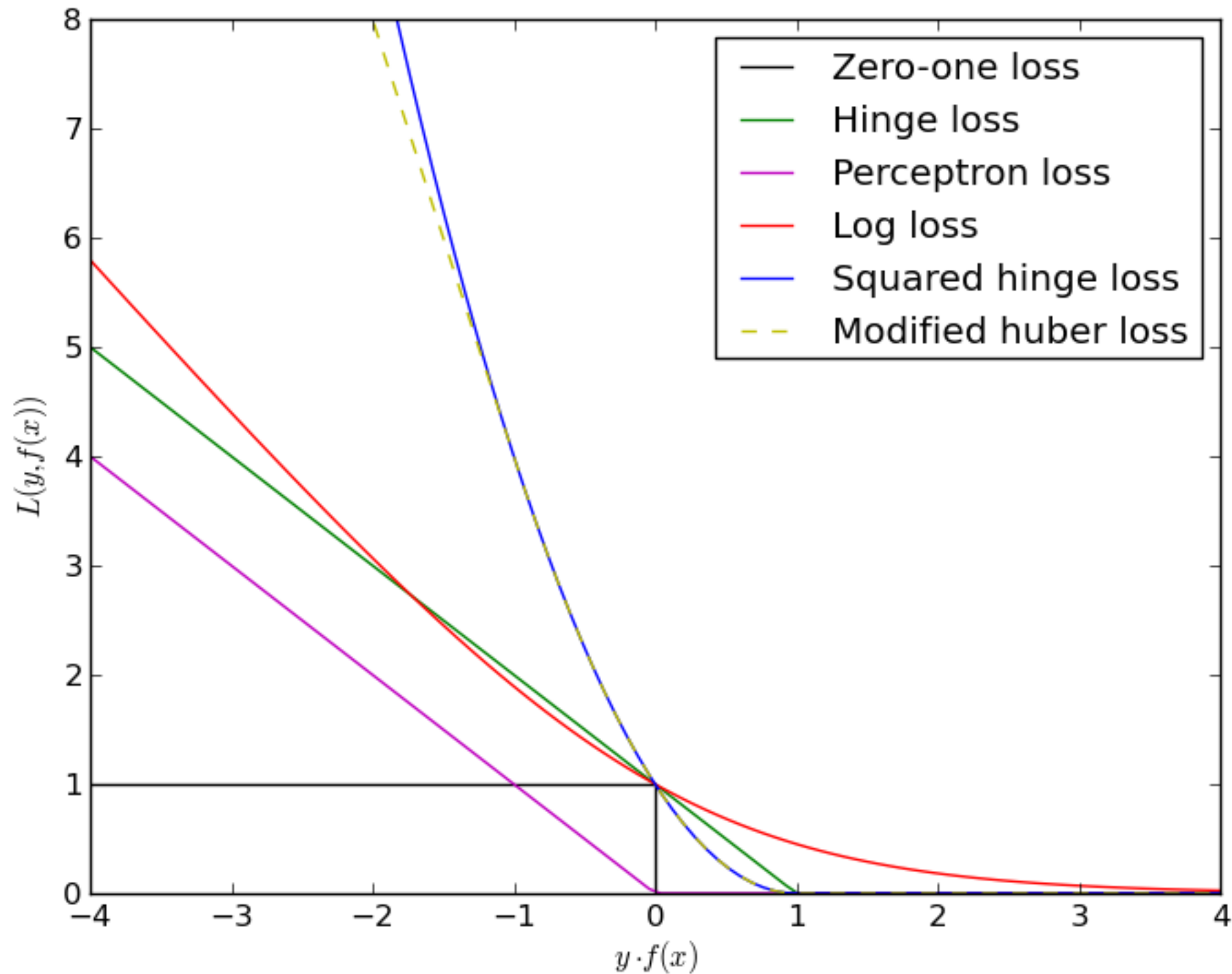
Batch Gradient Descent

- Compute the gradient over all training instances
- Deterministic
- Very slow for large datasets

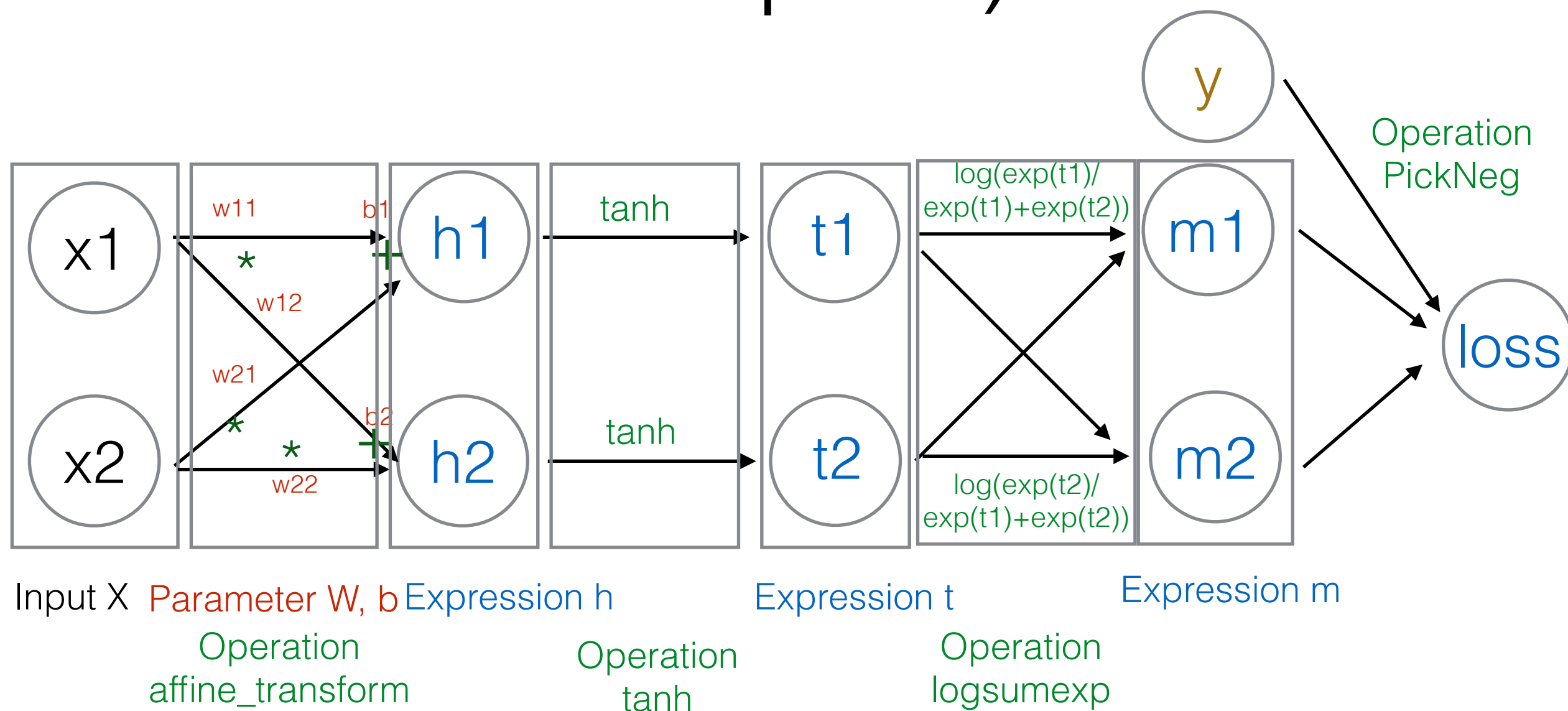
Stochastic Gradient Descent

- Compute the gradient and update parameters at each instance
- Stochastic
- Suitable for very large datasets
- Tends lead to better results than batch GD (not sure why!)

Other Loss Function?



Be More Neural (Be Deeper?)

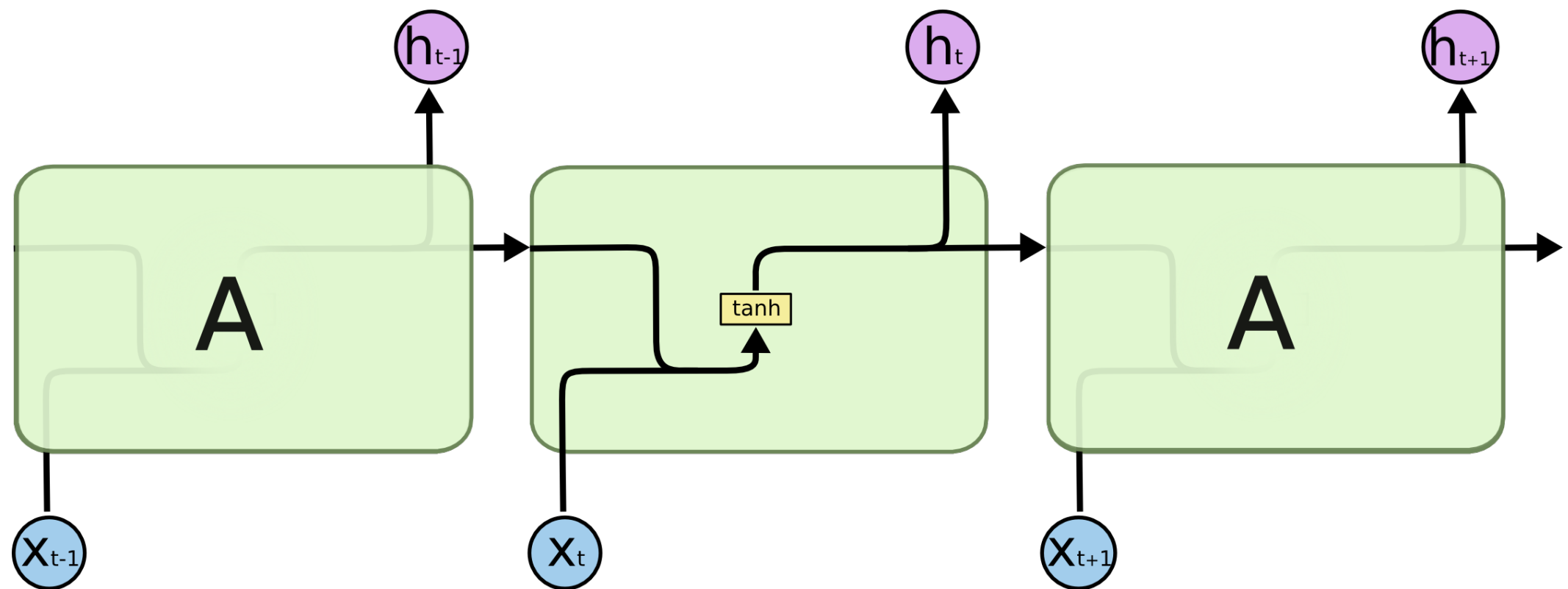


CNN code:

```
Expression loss = pickneglogsoftmax(tanh(affine_transform({b, W, x})), gold_y);
```

<http://www.cs.cmu.edu/~lingpenk/arkmeeting/code2.html>

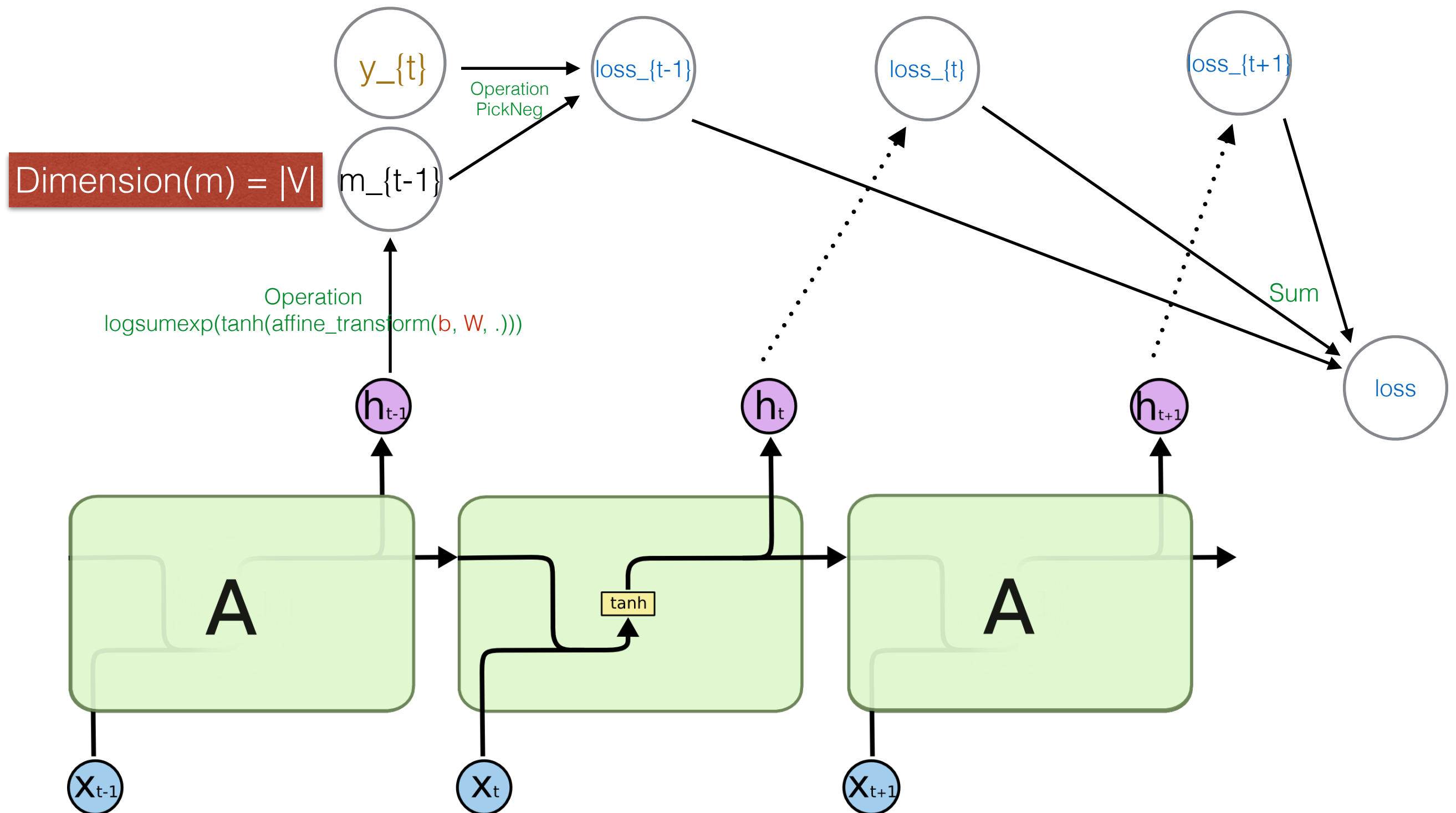
What is recurrent? (Be larger)



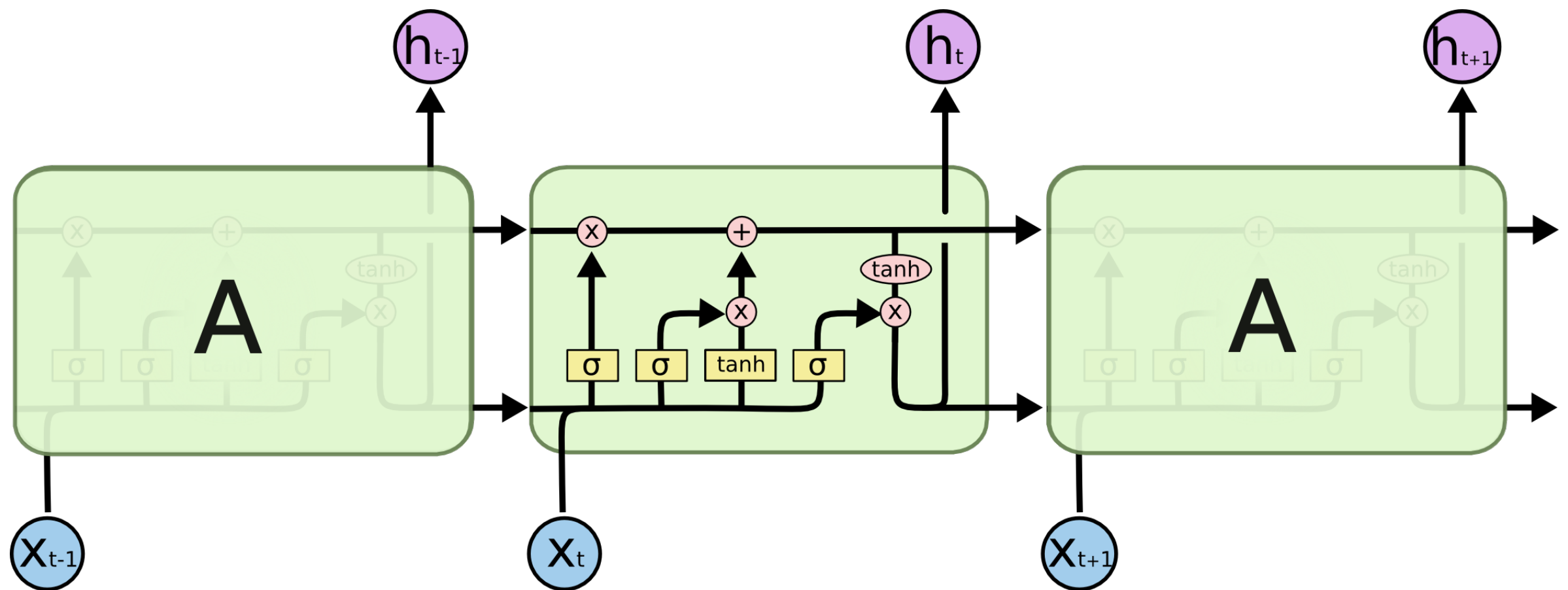
Wait! Where is the loss?

<http://colah.github.io/posts/2015-08-Understanding-LSTMs/>

RNNs — Language Model



LSTM — Be Fancy



<http://demo.clab.cs.cmu.edu/cdyer/lstms.pdf>

Computation Graph

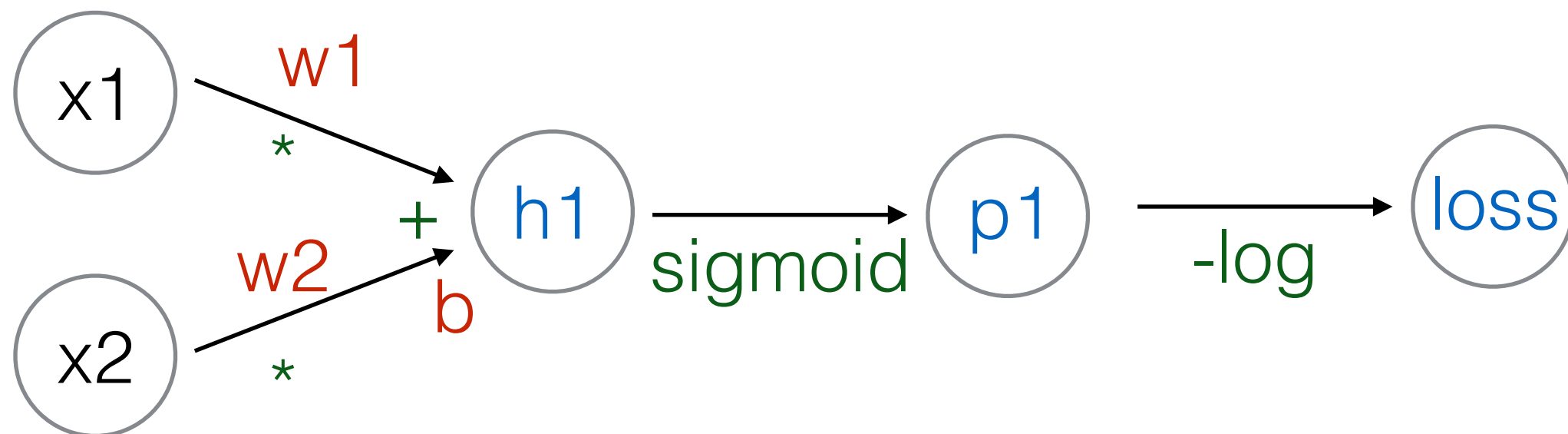
Input x1 x2

Parameter w1 w2 b

Expression h1 p1 loss

Special

Operation * + sigmoid -log



Thanks!