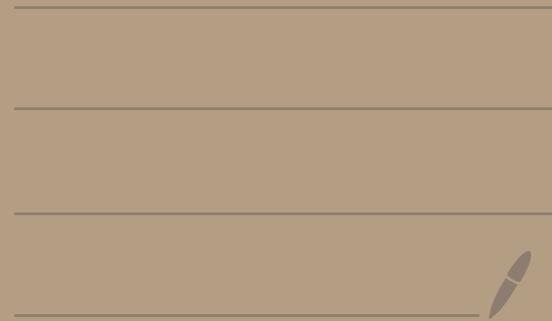
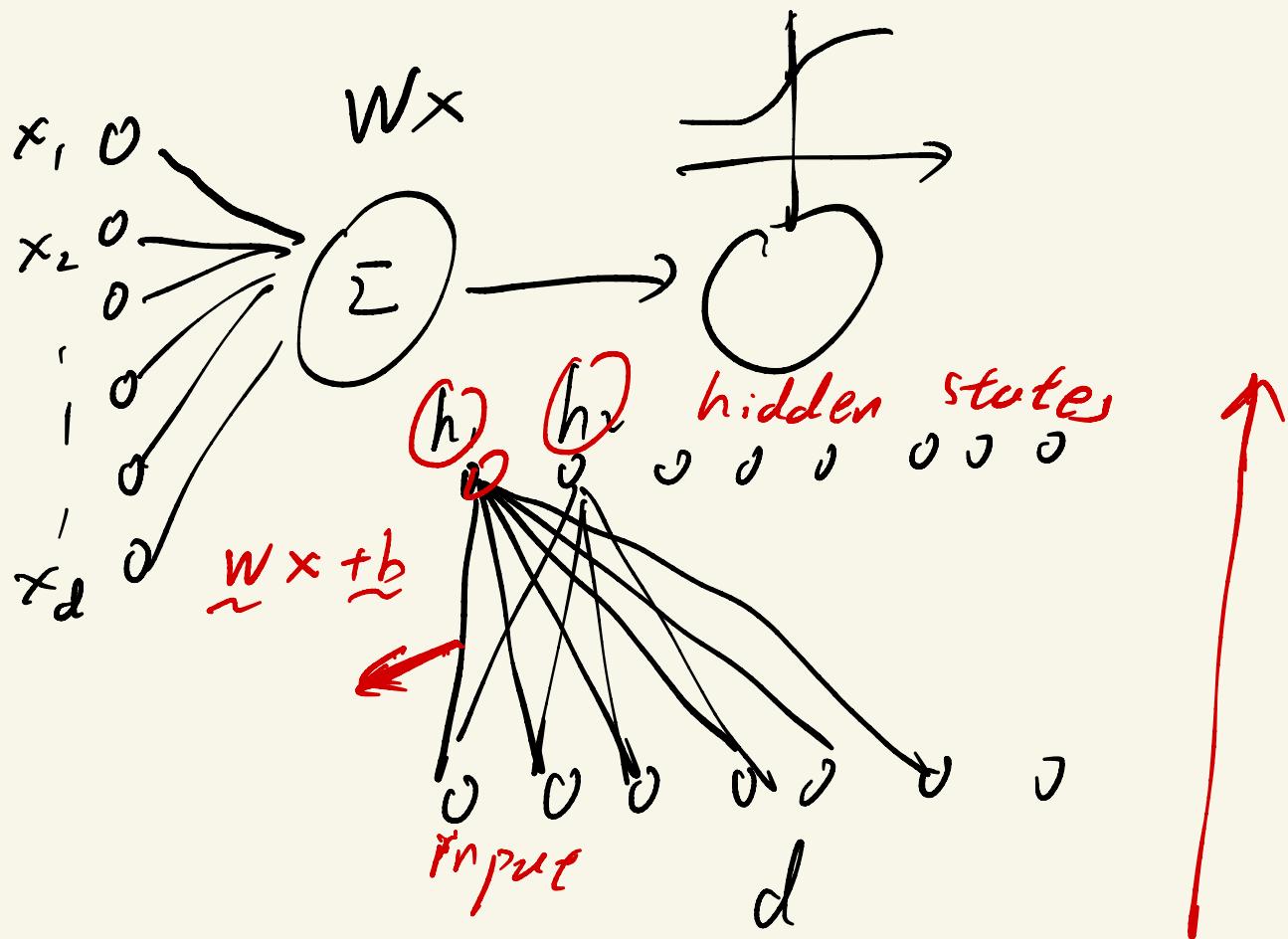


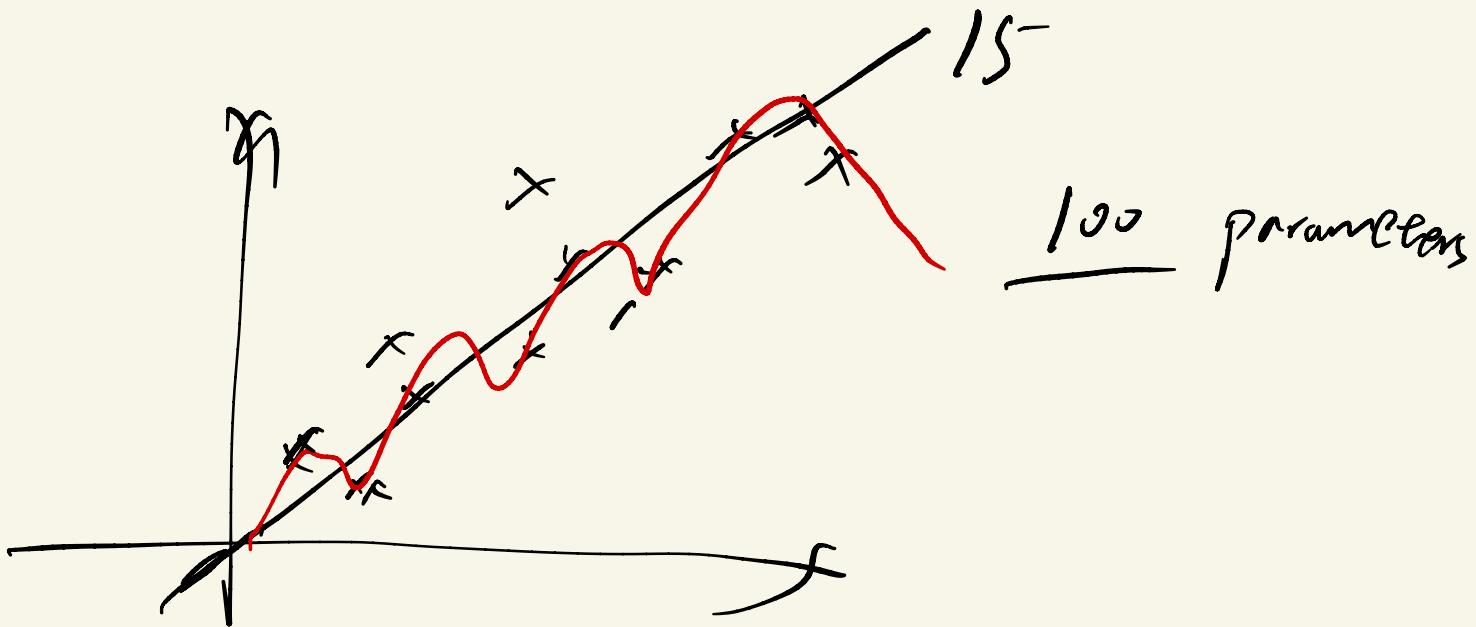
Lecture 18 NN, back propagation



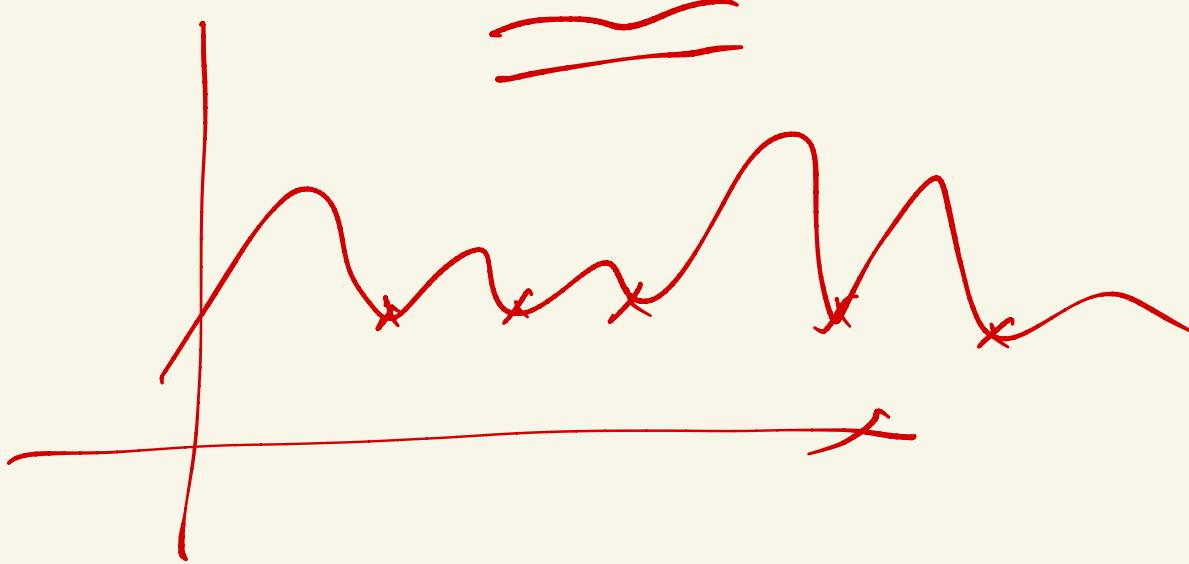


non-linear \rightarrow non-convex

Over parameterized



HMM \rightarrow non-convex

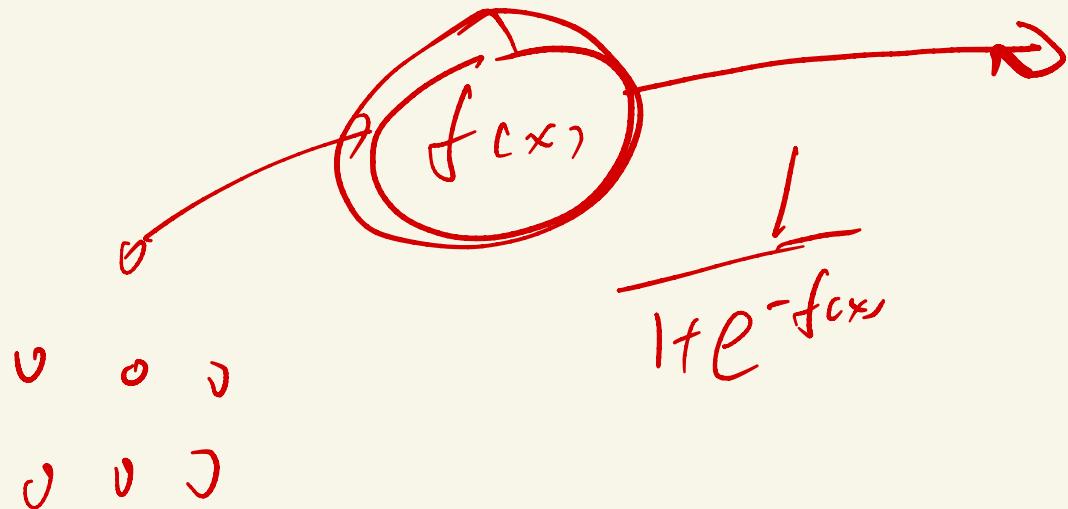


$P(c_{S_t} | s_{t-1})$

$P(o_t | s_t)$

MSE

$$\frac{1}{n} \sum (y - f(x))^2$$



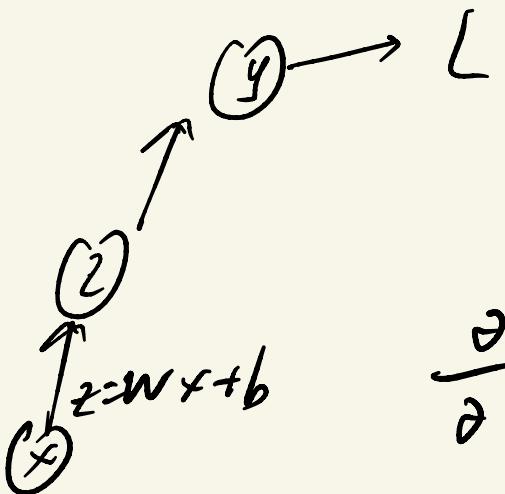
$L \rightarrow$ loss

$w \rightarrow$ weight

$\alpha \rightarrow$ learning rate

$$\text{net} = \sum_{i=0}^n w_i x_i$$

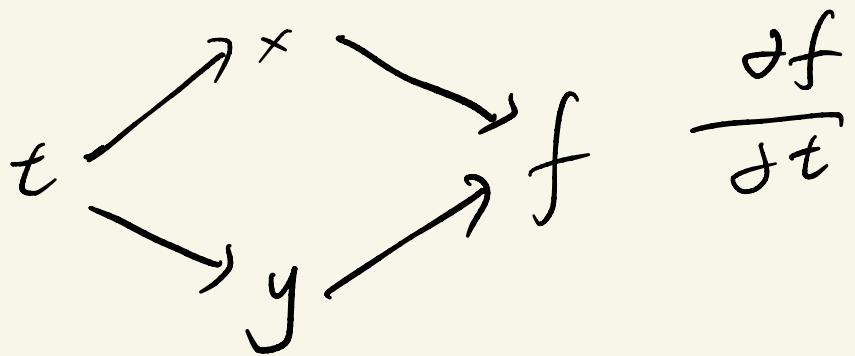
$$\frac{\partial o}{\partial w_i} = \frac{\partial o}{\partial \text{net}} \cdot \underbrace{\alpha}_{\frac{\partial \text{net}}{\partial w_i}}$$



$$\frac{\partial L}{\partial w} = \frac{\partial L}{\partial y} \cdot \frac{\partial y}{\partial z} \cdot \frac{\partial z}{\partial w}$$

A hand-drawn diagram showing the backpropagation formula for the gradient of the loss function with respect to the weight w . The formula is written as a product of three terms, each enclosed in a red oval. The first term is the derivative of the loss with respect to the output y . The second term is the derivative of y with respect to z . The third term is the derivative of z with respect to w . A black curved arrow points from the rightmost term towards the leftmost term, indicating the flow of the gradient.

fan-out



$$\frac{\partial f}{\partial x} \cdot \frac{dx}{dt} + \frac{\partial f}{\partial y} \cdot \frac{dy}{dt}$$

$$\frac{\partial L}{\partial w_{ii}}$$

$$\frac{\partial L}{\partial y_1}$$

$$\frac{\partial L}{\partial y_2}$$

$$\frac{\partial L}{\partial y_i}$$

$$-\frac{t_i}{y_i}$$

$$\frac{\partial L}{\partial y_j}$$

$$-\frac{t_j}{y_j}$$

$$\frac{\partial L}{\partial z_i}$$

$$\frac{\partial y_i}{\partial z_i} = y_i \cdot c_i - y_i$$

$$y_i \cdot c_i$$

$$-y_i$$

$$\frac{\partial L}{\partial w_{ii}}$$

$$\frac{\partial L}{\partial z_i}$$

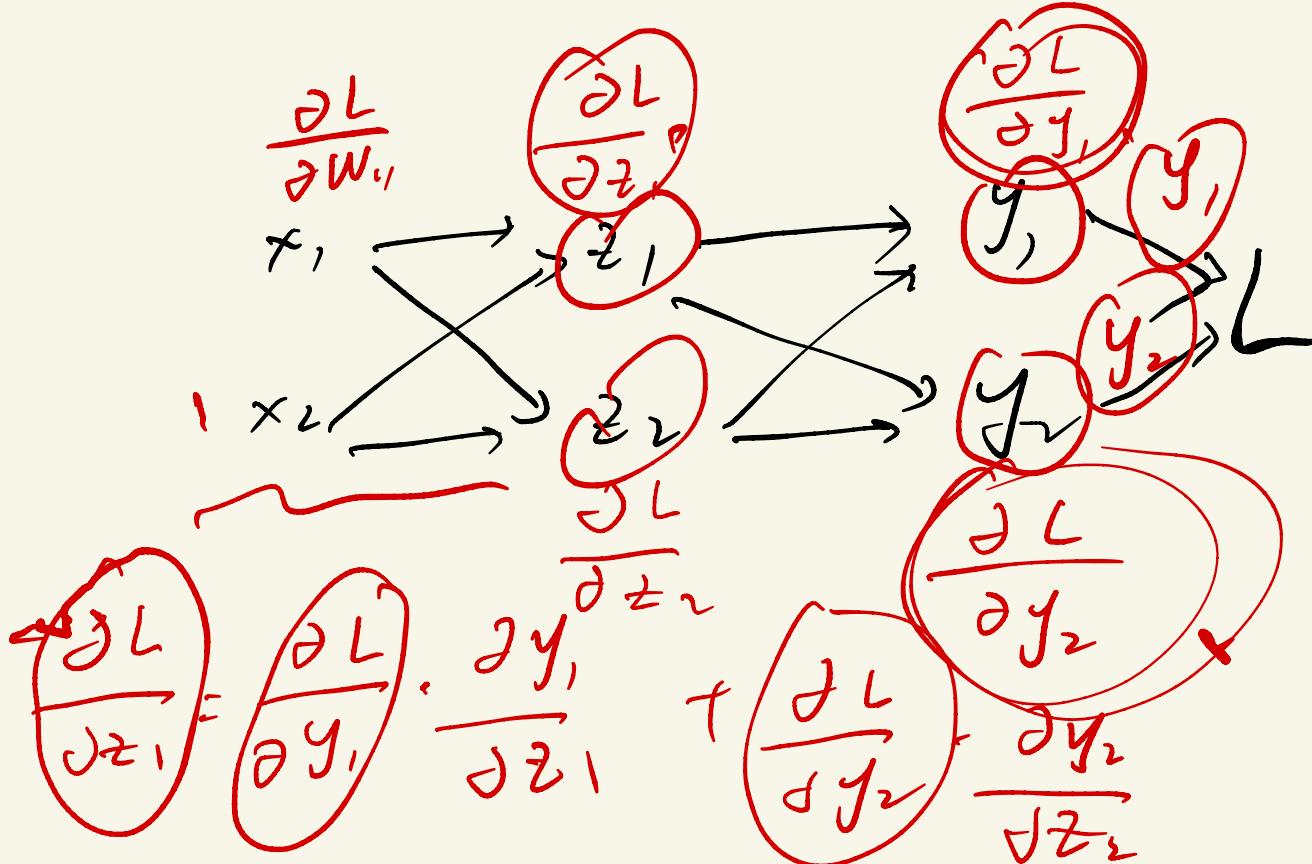
$$\frac{\partial L}{\partial z_i}$$

$$\frac{\partial z_i}{\partial w_{ii}}$$

$$\rightarrow x_i$$

$$\frac{\partial L}{\partial z_i} = \frac{\partial L}{\partial y_i} \cdot \frac{\partial y_i}{\partial z_i} + \frac{\partial L}{\partial y_j} \cdot \frac{\partial y_j}{\partial z_i}$$

$$\frac{\partial y_i}{\partial z_i}$$



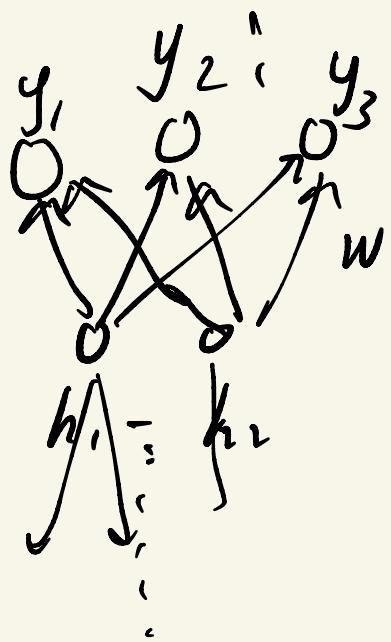
$$\frac{\partial L}{\partial w_{ii}} = \frac{\partial L}{\partial z_i} \cdot \frac{\partial z}{\partial w_{ii}}$$

$$\sum_j w_{ij} x_j$$

2017

torch dy net caffe

Sigmoid

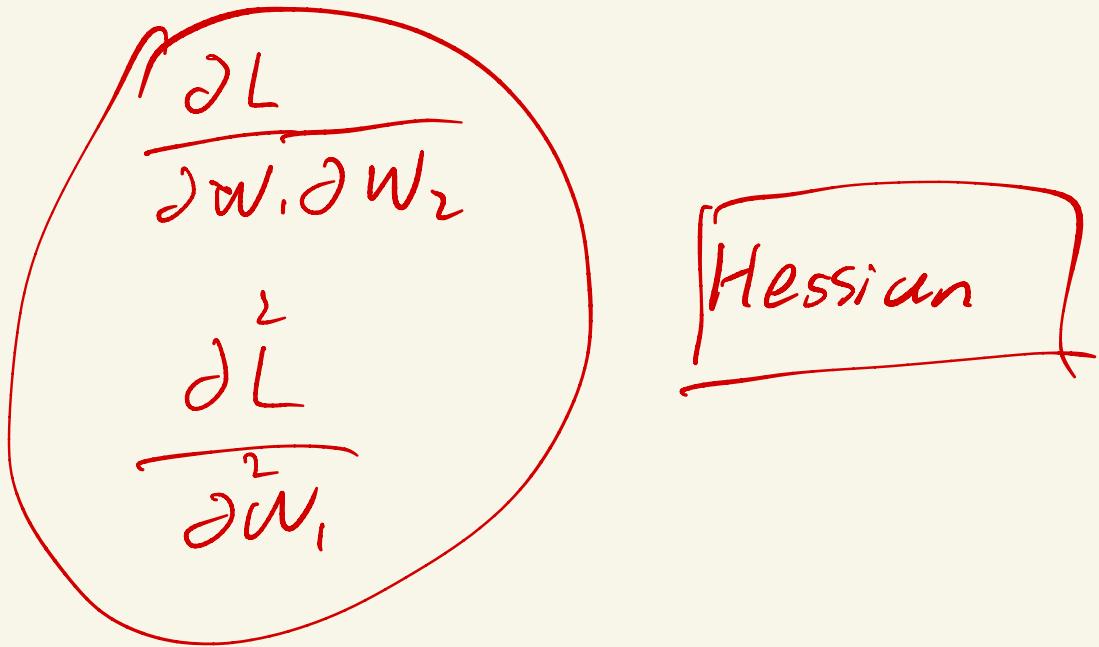


$$\vec{y} = \vec{w} \vec{h}$$

$$\frac{\partial L}{\partial y_1} \quad \frac{\partial L}{\partial y_2} \quad \frac{\partial L}{\partial y_k}$$

$$\frac{\partial L}{\partial h_i} = \underbrace{\sum_k \bar{y}_k w_{ki}}$$

O(k^2)



$$\underbrace{E_{x \sim P_{\text{data}}}}$$
 i.i.d

training
 x_1, \dots, x_N unbiased samples from P_{data}

MC MC

$$\bar{E}_{x \sim p} f(x) \leftarrow \bar{N}$$

$$\approx \frac{1}{n} \sum_{i=1}^n f(x_i)$$

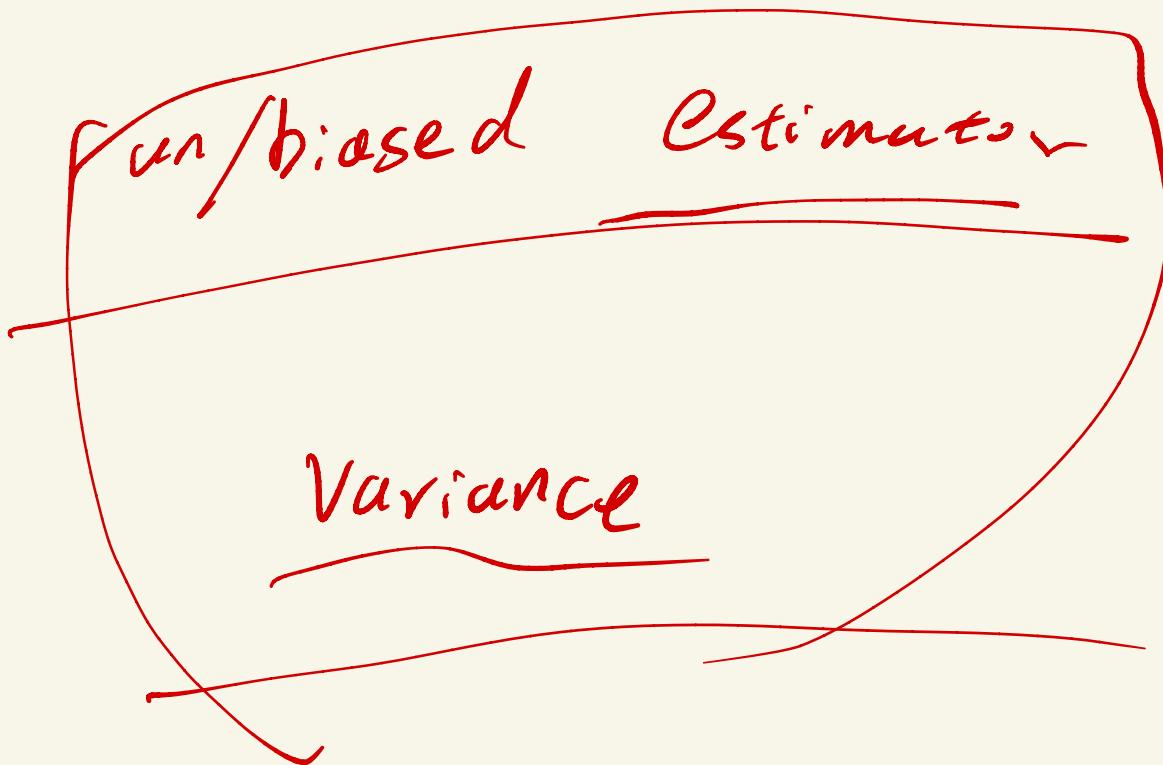


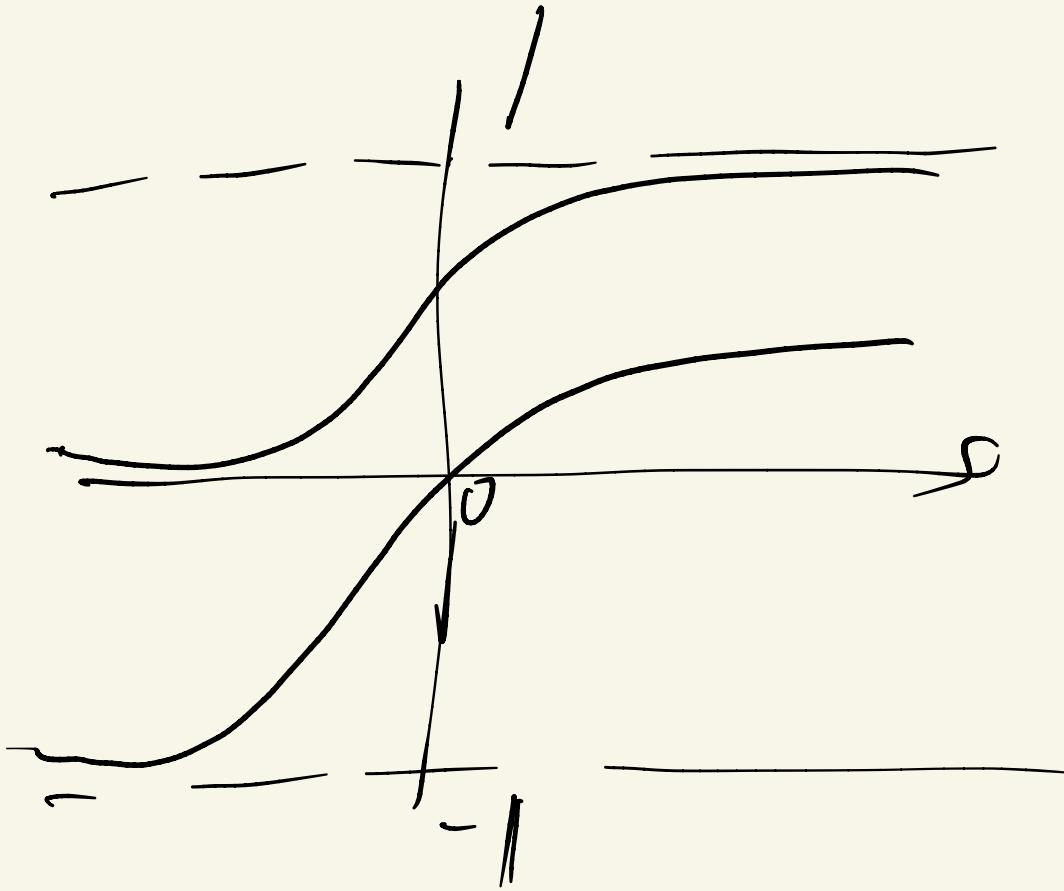
n = 32

n = 1

dropout

MCMC





$$y = \max(0, x)$$



fully connected

