Me: Michael Gelbart

(PhD Student
Visting, Mark this year)

mgelbart@seas.harvard.edu

Terminology / Notation

σ : nonlinear element-wise function, "activation function"

$\underline{W}$: "weights" - free parameters

b : "biases" - " "

x : "units", "neurons", "hidden units"

train / learn / optimize → set W, b

Supervised learning: learn $y = f(x)$ from labeled data

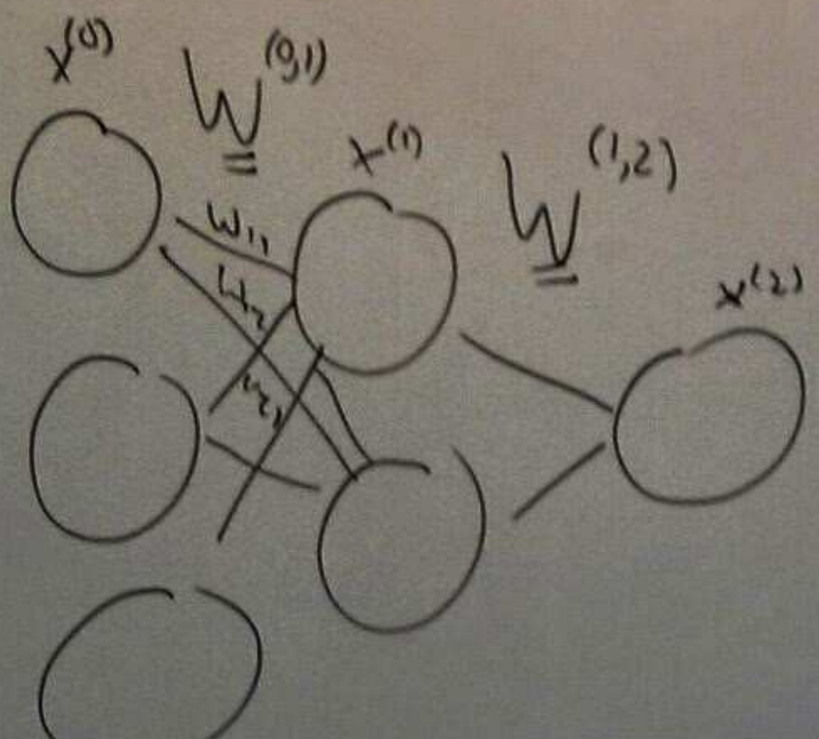
$f(x) = 2x$ function

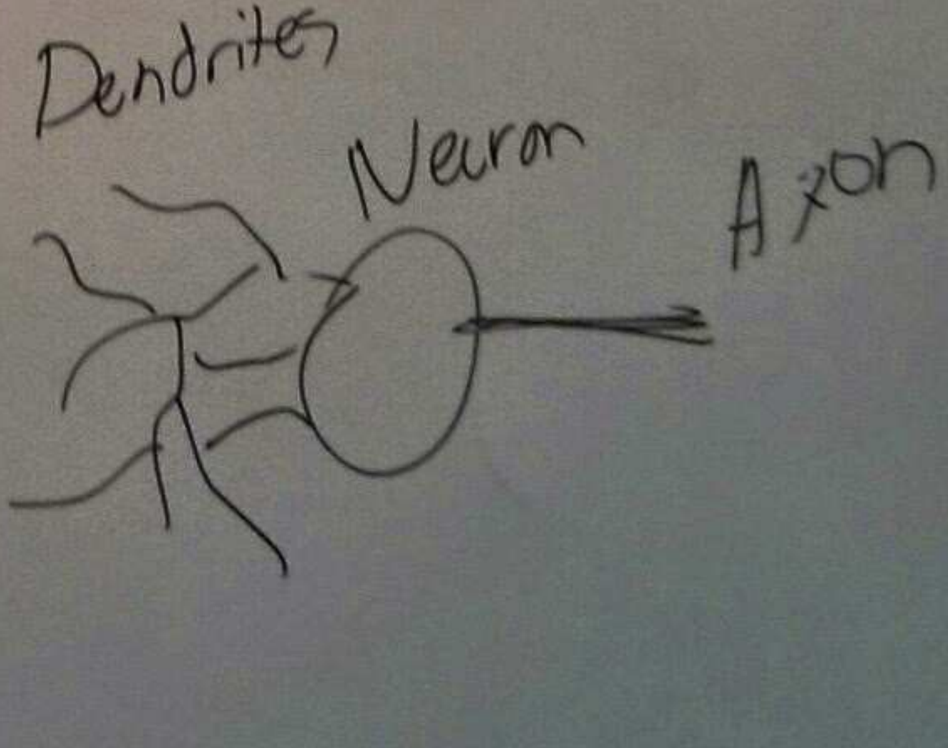
$f(x) = ax^b$ class of function

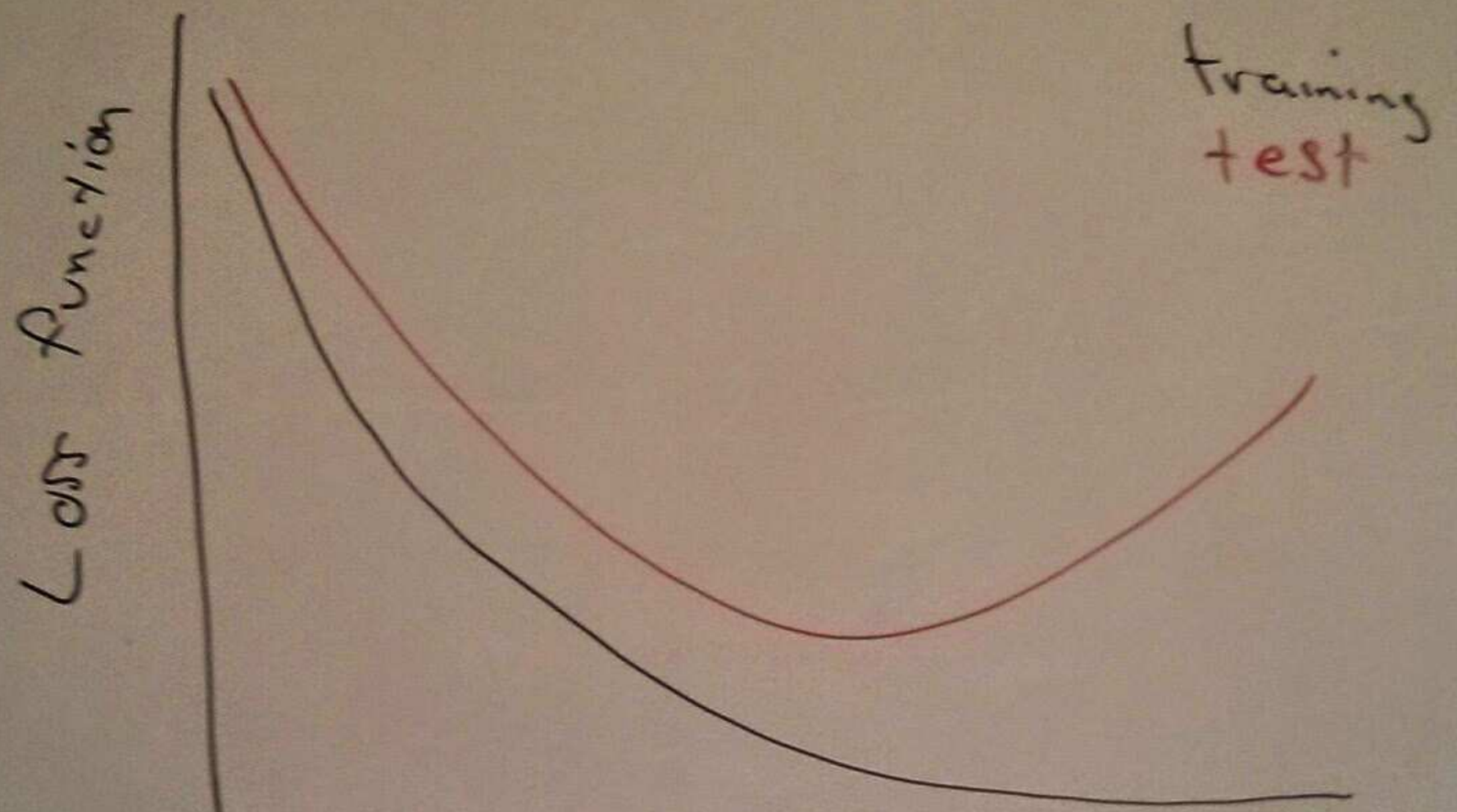
polynomials

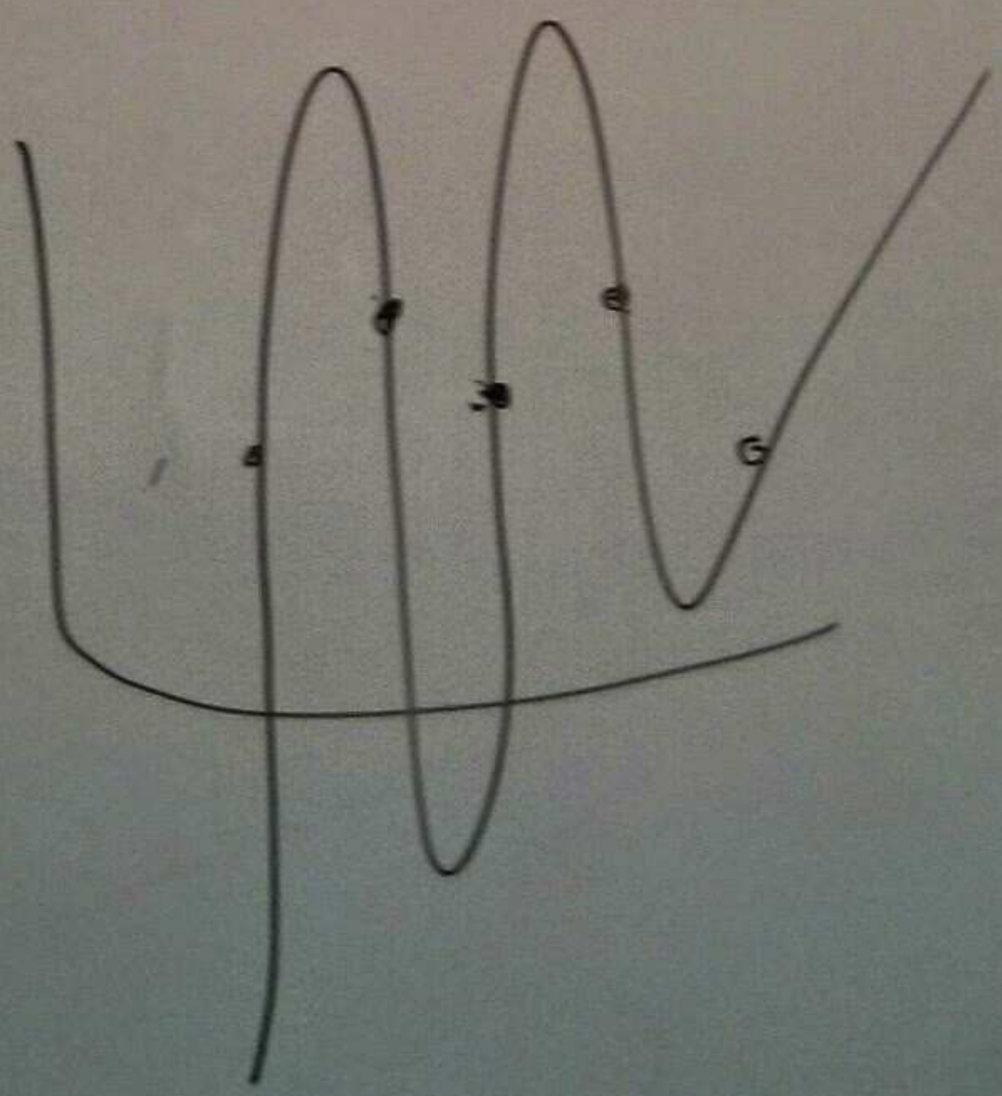
Neural networks

$$x^{(j+1)} = \sigma\left(\underline{\underline{W}}^{(j,j+1)} \underline{x}^{(j)} + \underline{b}^{(j)}\right)$$






Loss function
training
test
iteration



Backpropagation

is an (efficient) algorithm to compute the gradient of the Loss w.r.t. weights (making extensive use of the chain rule)

$$z^{(0)} \to a^{(1)} \xrightarrow{\sigma} z^{(1)} \longrightarrow a^{(2)} \xrightarrow{\sigma} z^{(2)} \longrightarrow L$$

$W^{(0,1)}$, $W^{(1,2)}$

$$L = \frac{1}{2} \sum_j \left(z_j^{(2)} - \hat{z}_j \right)^2$$

$$z_j^{(n)} = \sigma\left(a_j^{(n)}\right)$$

$$a_j^{(n)} = \sum_i W_{ji}^{(n-1,n)} \, z_i^{(n-1)}$$

$$\frac{\partial L}{\partial W_{ji}^{(1,2)}} = \left(z_j^{(2)} - \hat{z}_j\right) \sigma'\left(a_j^{(2)}\right) z_i^{(1)}$$

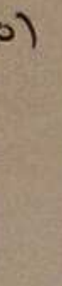
$$z_j^{(2)} - \hat{z}_j \qquad \sigma'\left(a_j^{(2)}\right)$$

$$\frac{\partial L}{\partial a_j^{(2)}} = \frac{\partial L}{\partial z_j^{(2)}} \frac{\partial z_j^{(2)}}{\partial a_j^{(2)}}$$

$$\frac{\partial L}{\partial W_{ji}^{(1,2)}} = \frac{\partial L}{\partial a_j^{(2)}} \frac{\partial a_j^{(2)}}{\partial W_{ji}^{(1,2)}}$$

$$z_i^{(1)}$$

$$\frac{\partial L}{\partial W_{ji}^{(0,1)}} = \frac{\partial L}{\partial a_j^{(1)}} \frac{\partial a_j^{(1)}}{\partial W_{ji}^{(0,1)}} = z_i^{(0)} \frac{\partial L}{\partial z_j^{(1)}} \frac{\partial z_j^{(1)}}{\partial a_j^{(1)}}$$



$$= z_i^{(0)} \sigma'(a_j^{(1)}) \frac{\partial L}{\partial z_j^{(1)}}$$

$$= z_i^{(0)} \sigma'(a_j^{(1)}) \sum_k \frac{\partial a_k}{\partial z_j} \frac{\partial L}{\partial a_k}$$

$$= z_i^{(0)} \sigma'(a_j^{(1)}) \sum_k w_{kj}^{(1,2)} \left(z_k^{(2)} - \hat{z}_k\right) \sigma'\left(a_k^{(2)}\right)$$