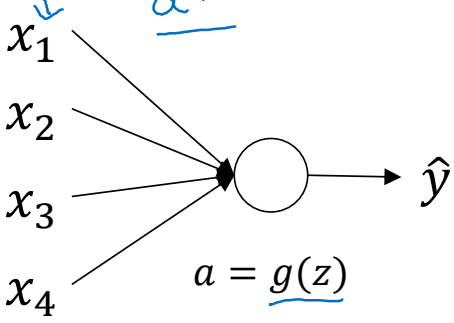
deeplearning.ai

Setting up your optimization problem

Weight initialization for deep networks

~~Vanishing/exploding gradients~~

Single neuron example



$a^{[l]}$

$w^{[l]}$

$z = w_1 x_1 + w_2 x_2 + \cdots + w_n x_n$ ~~+b~~

large n → smaller w_i

$Var(w_i) = \frac{1}{n} \quad \frac{2}{n}$

$w^{[l]} = np.random.randn(shape) * np.sqrt\left(\frac{2}{n^{[l-1]}}\right)$

ReLU

$g^{[l]}(z) = ReLU(z)$

Other variants:

tanh

$\frac{1}{n^{[l-1]}}$

Xavier initialization

$\frac{2}{n^{[l-1]} + n^{[l]}}$