INITIALIZATION

OF NEURAL NET PARAMETERS

- Parameters often initialized as small values randomly drawn from uniform or normal distributions.

- Scale of parameter initialization matters. Both too big or too small are bad.

- One option for scale is normalized Initiation: $U\left(-\sqrt{\frac{6}{m+n}}, \sqrt{\frac{6}{m+n}}\right)$

Uniform distribution ↘

\# of layer inputs
\# of layer outputs

Chris Albon