

RANDOM INITIALIZATION

$\text{optTheta} = \text{fminunc}(@\text{costFunction}, \text{initialTheta},$
 $\text{options})$

Symmetry will mean redundancy!

$\text{initialTheta} = (\text{i}) \text{zeros}(\text{M}, \text{L})$

So break it symmetry! Averp

$(\text{rand}(25 \times 3) \backslash ((\text{zeros}(\text{M}, \text{L}))^\top)$

$\text{Theta1} = \text{rand}(10, 11) * 2 * \text{INIT_EPSILON} - \text{INIT_EPSILON};$

$\text{Theta2} = \text{rand}(1, 11) * 2 * \text{INIT_EPSILON} - \text{INIT_EPSILON};$

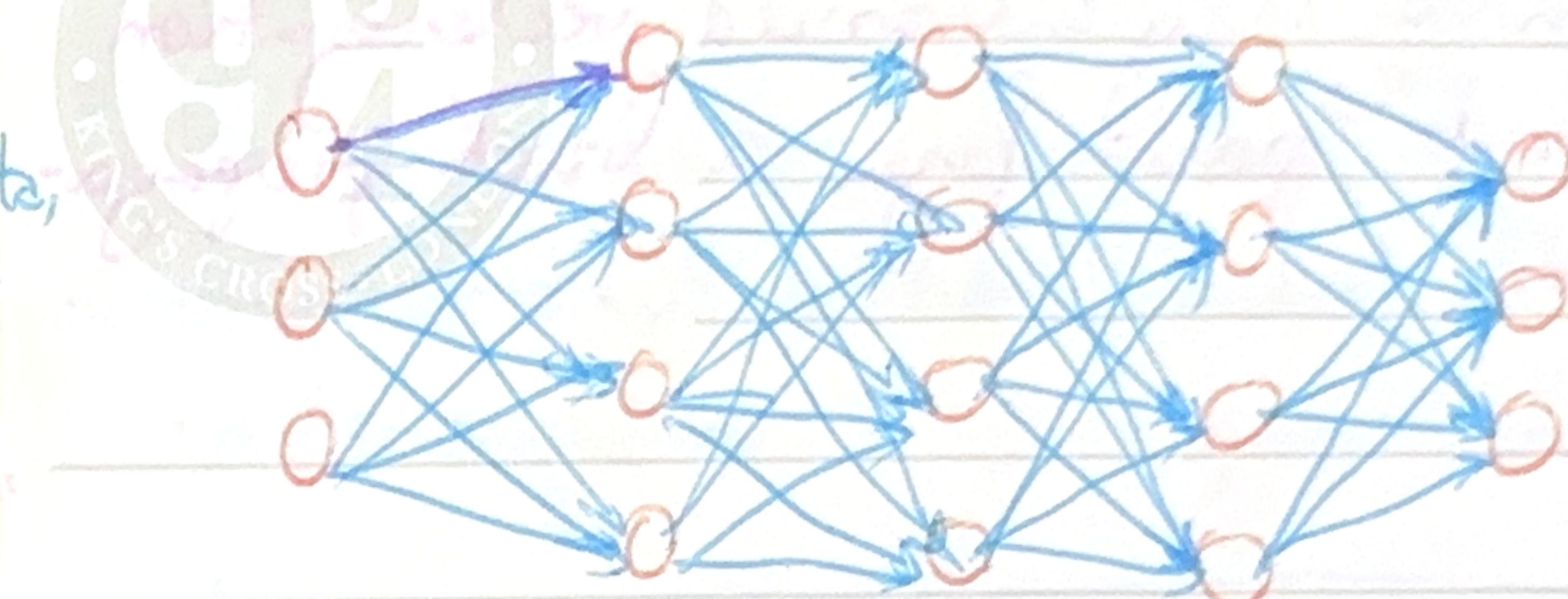
Doing this will initialize all parameters with values between $[-E, E]$

SVT & good hope can do it -

and, and the SOFTMAX -

softmax will give us the

WRAPPING UP



No. of input units : dimension of features

No. of output units : number of classes

Hidden layers: 1 is reasonable
but usually the more
the better

Quando tiver mais de uma hidden layer, Prof. Andrew sugere manter o mesmo número de neurônios em cada uma.

Além disso, sugere manter $s(2:L-1)$ comparável a $s(1)$, talvez no máximo 2, 3 ou 4x o número de features.

→ Andrew fala que ele mostra em outros vídeos como escolher a arquitetura NN