Softmax: only occurs at the end of the neural network. Spits out a vector whose elements sum to 1, and all elements are between 0 and 1.

For all outputs, exponentiate them (i.e. do e^y for all y outputs), add them all together, and then divide each exponentiated output by the total.

Softmax: wants to pick something, and only one thing. Don't use for multi-label classification!

*** unfreeze vs bn-freeze: bn stands for batch normalization. What's that?