

## Session 2)

1. No, weights need to be initialized independent of each other
2. Yes
4. SELU>ELU>leaky RELU>RELU>tanh>logistic
  - a. Network architecture prevents self-normalization -> ELU (Eg. RNNs)
  - b. Runtime latency limitations ->leaky RELU
  - c. Need for speed -> RELU
6.
  - a. Set small weights to 0
  - b. Apply strong l1 regularization
  - c. TF-MOT pruning API
7. DO slows down convergence, not inference on new data
  - a. MC slows down convergence and inference compared to no DO