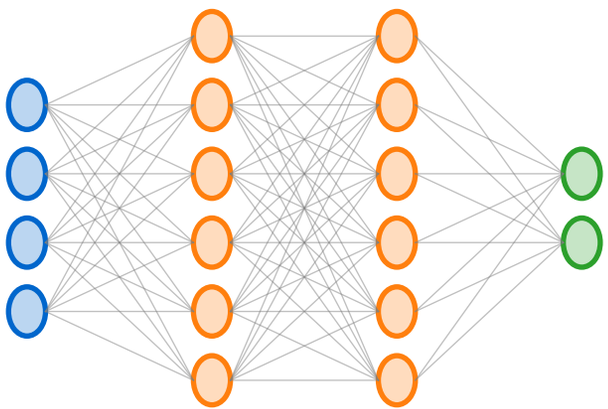
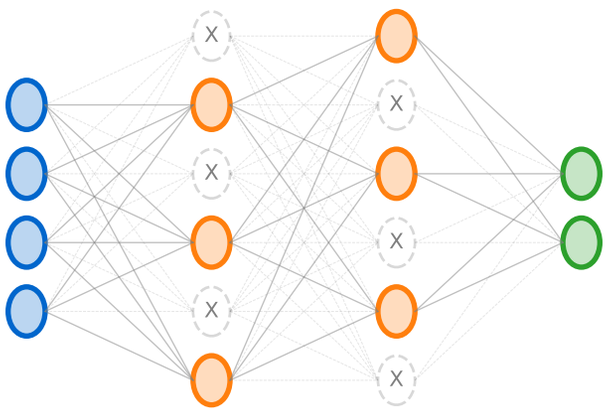
Dropout: A Simple Regularization Technique

Full Network (Inference)



All neurons active
during inference

Training with Dropout (p=0.5)



50% neurons randomly
dropped each batch

Dropout Explained

Training: Randomly drop neurons with probability p

Forces redundant representations

Prevents co-adaptation

Inference: Use ALL neurons

Scale outputs by (1-p)

Effect: Ensemble of thin networks

Typical p: 0.2-0.5 for hidden layers

During training: $\tilde{h} = h \cdot m$, where $m \sim \text{Bernoulli}(1-p)$

Introduced by Hinton et al. (2012) - Simple yet highly effective!