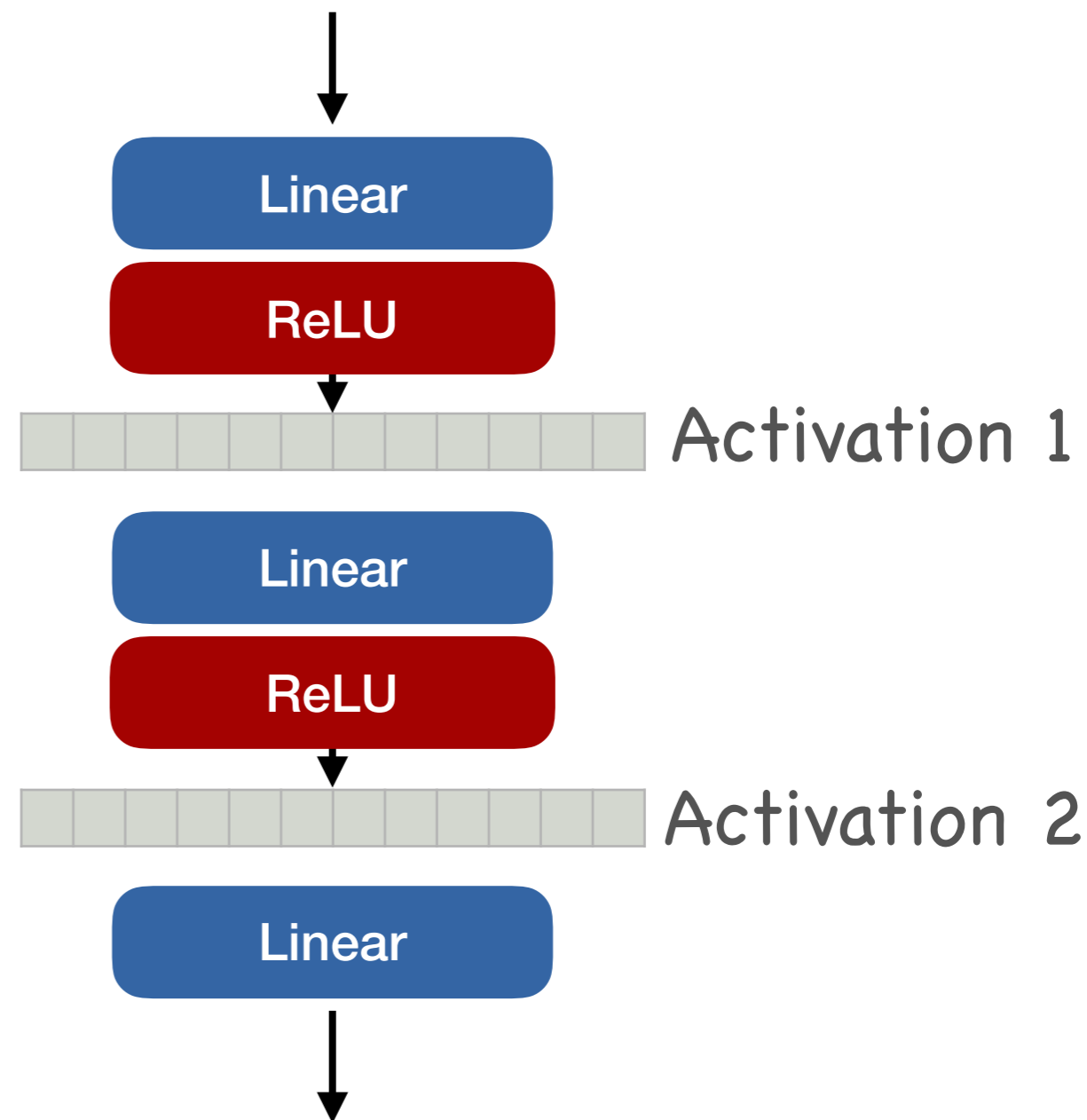


Dropout

© 2019 Philipp Krähenbühl and Chao-Yuan Wu

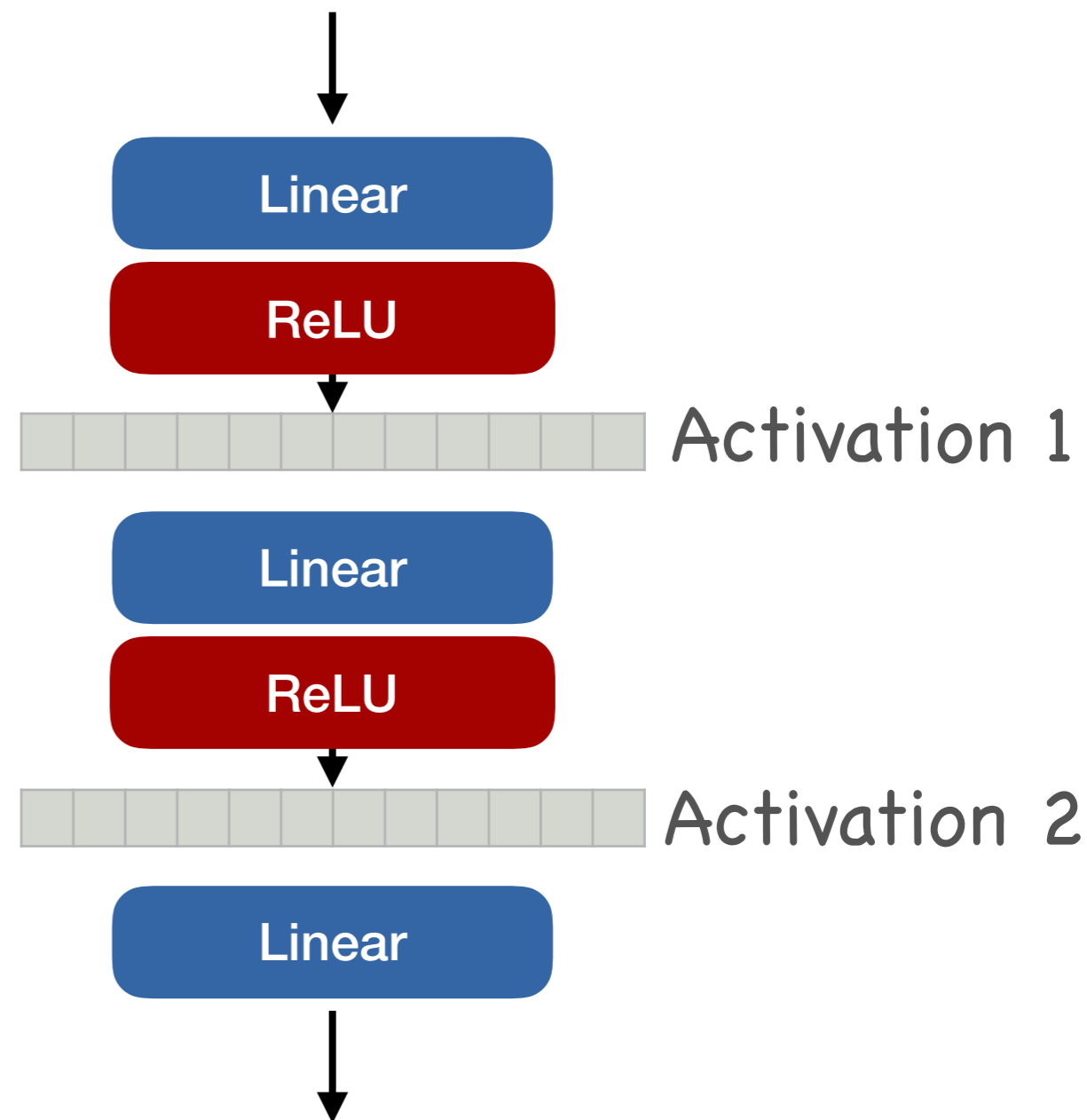
Overfitting in deep networks

- Overfitting
 - Exploit patterns that exist in training data, but not in the validation / test data
- Not all activations overfit



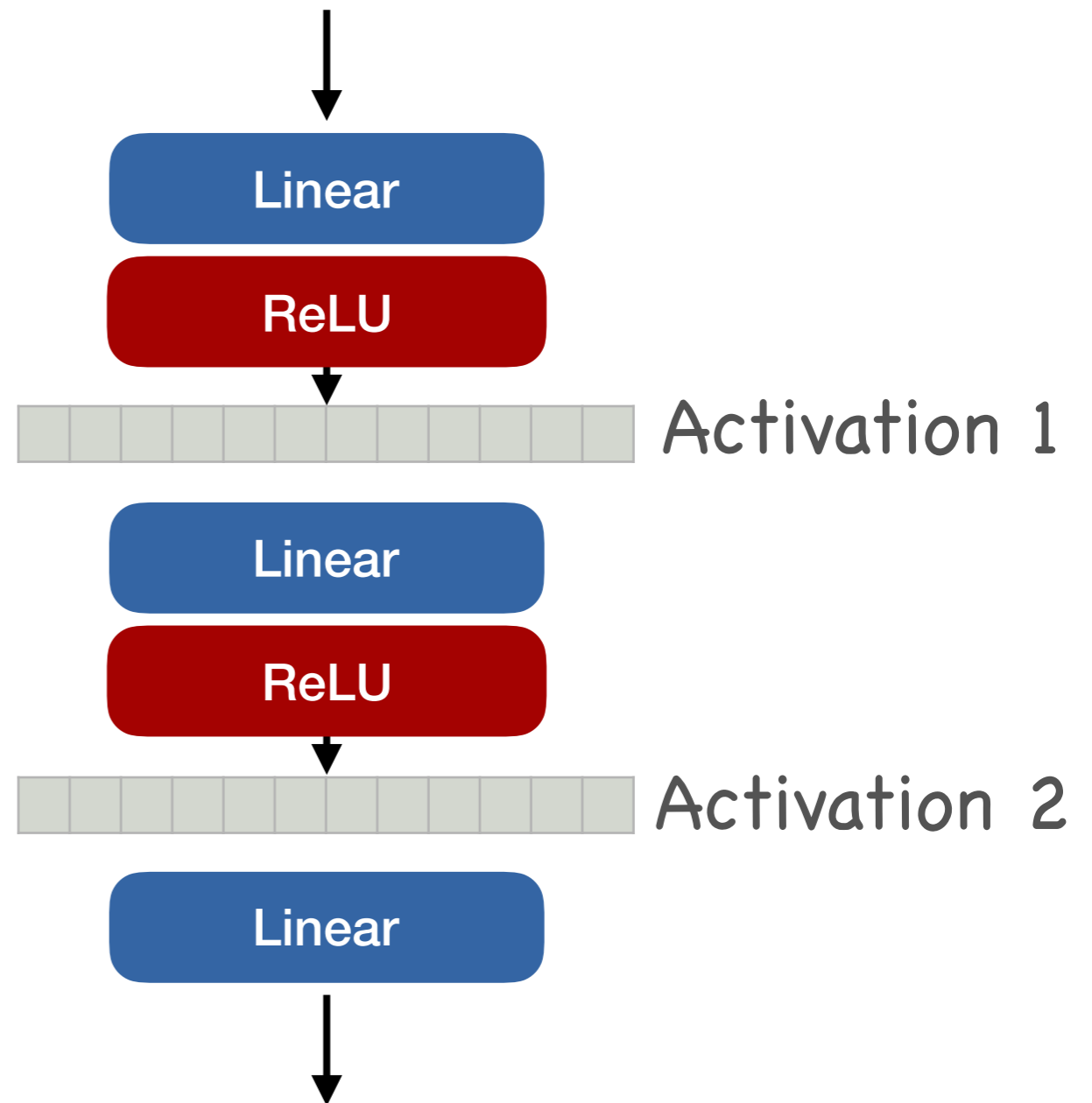
Overfitting in deep networks

- Deeper layers overfit more
- Rely on overfit activations from previous layers



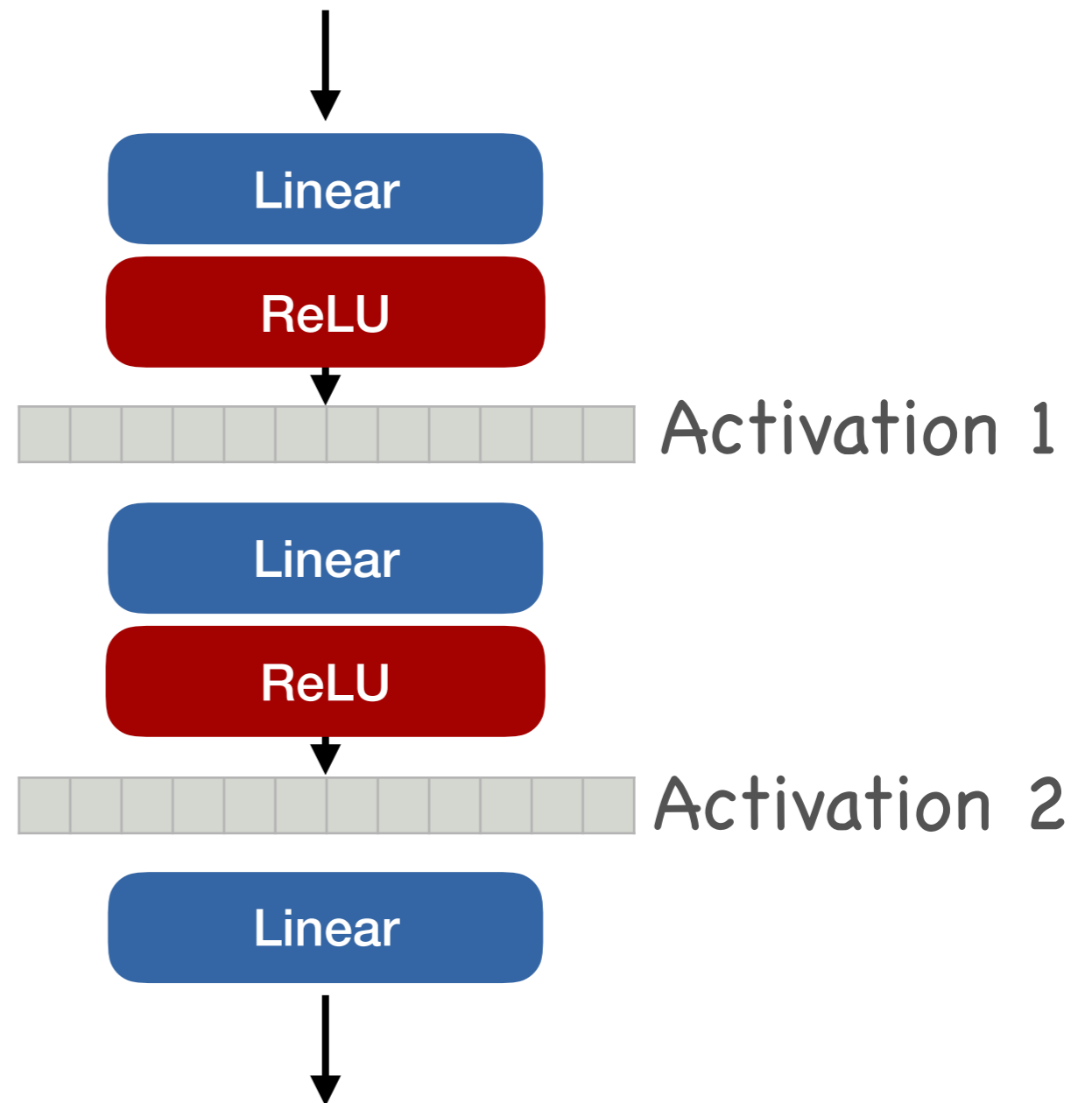
Preventing overfitting in deep networks

- Reduct reliance on specific activations in previous layer
- Randomly remove activations



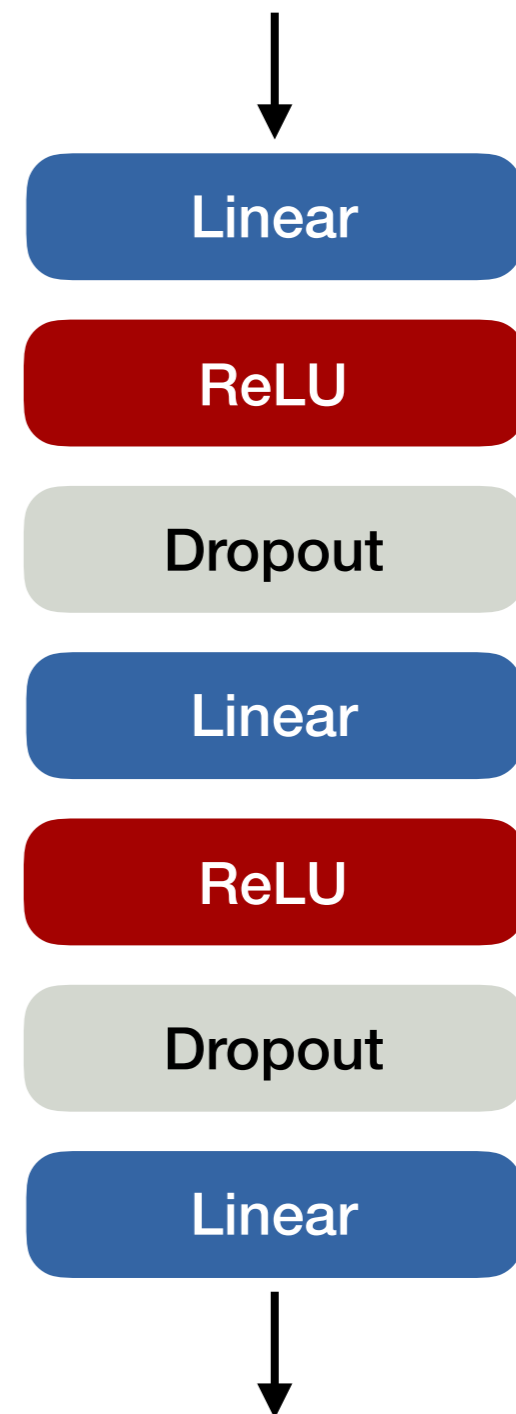
Dropout

- During training
 - With probability α set activation $a_l(i)$ to zero
- During evaluation
 - Use all activations, but scale by $1 - \alpha$



Dropout in practice

- A separate layer `torch.nn.Dropout`
- During training
 - With probability α set activation $a_l(i)$ to zero
 - Scale activations by $\frac{1}{1 - \alpha}$
- During evaluation identity



Where to add dropout?

- Before any large fully connected layer
- Before some 1x1 convolutions
- Not before general convolutions

