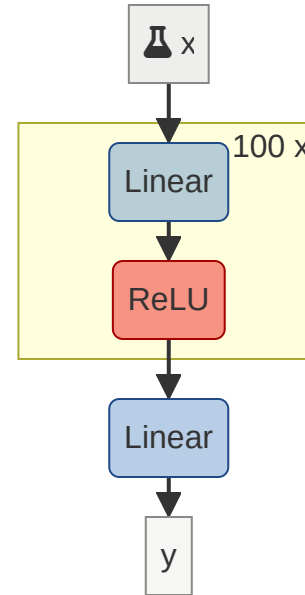


Residuals and Normalizations

Let's train really deep networks

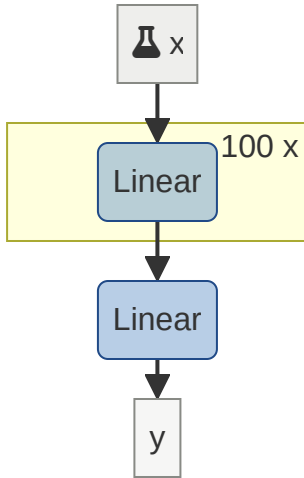
What happens if we train 100-layer networks?

- It wont train 😞

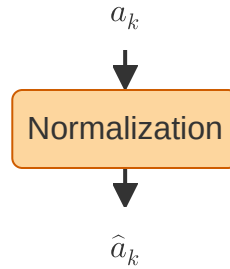


Vanishing Gradients, Normalizations, Residuals

Vanishing Activations and Gradients



Normalizations



Residuals

