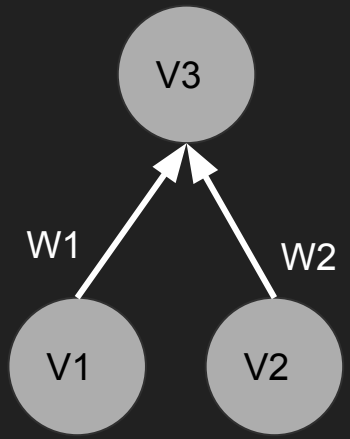
# Deep Learning

An Extremely Short Intro

# The baby neural net

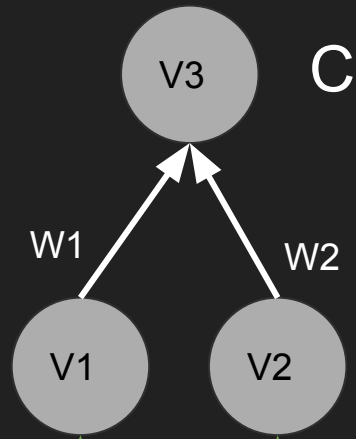



$$V3 = \underline{W1} \times V1 + \underline{W2} \times V2$$

$$V3 = \text{Tanh}(\underline{W1} \times V1 + \underline{W2} \times V2)$$

# Supervised Training: Stochastic Gradient Descent

| V1 | V2 | T3 |
|---|---|---|
| 0 | 0 | 0 |
| 0 | 1 | 0 |
| 1 | 0 | 0 |
| 1 | 1 | 1 |



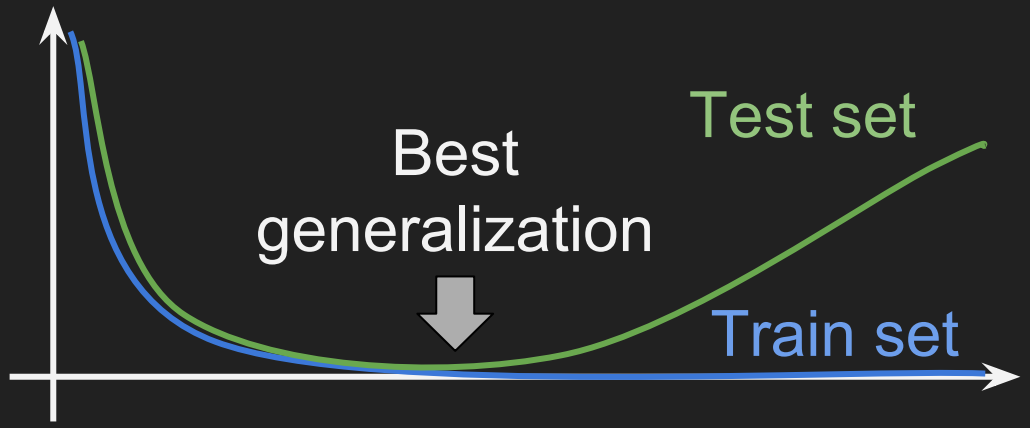

C = fancy_distance(T3, V3)

W1 = W1 - 0.01 x dW1/dC

W2 = W2 - 0.01 x dC/dW2

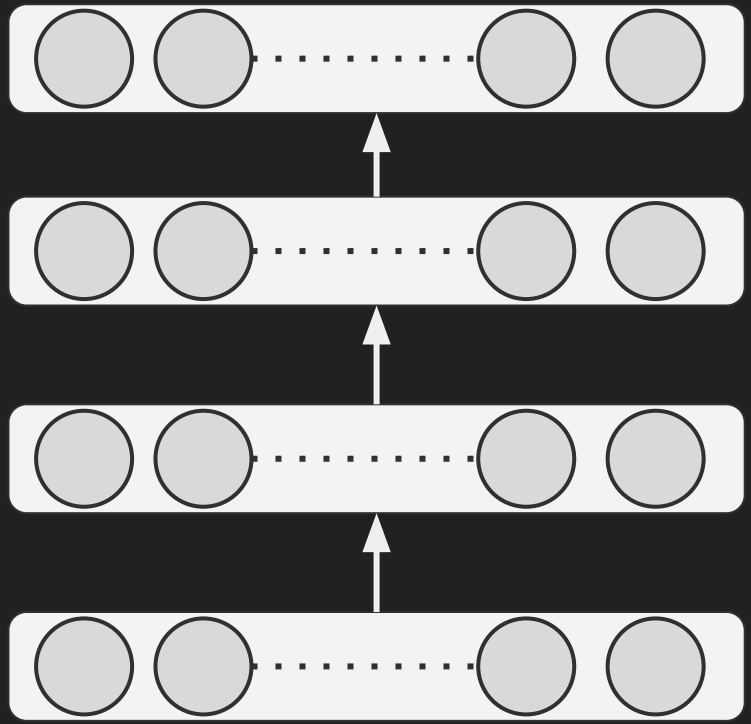
Learning rate

# Supervised Training #2

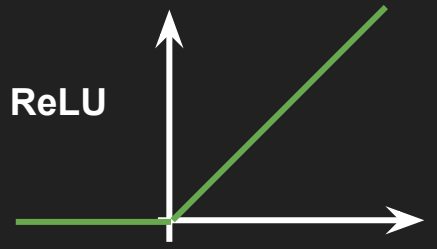



**Model has too much capacity ie:
too many weights / latitude**

# What is a Deep Neural Network?

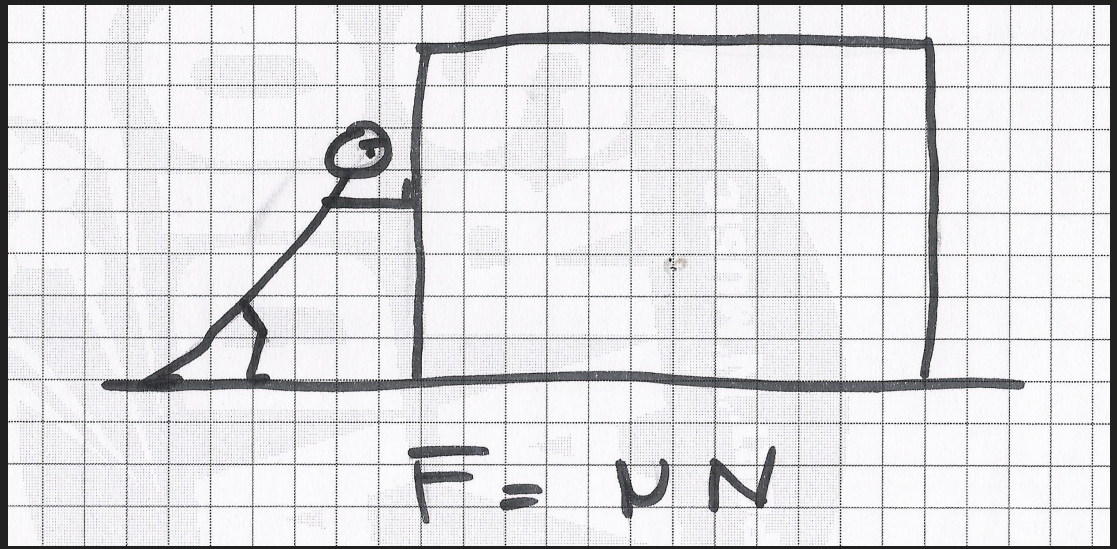


- Use ReLU instead of Tanh
- Give it a lot of data
- Be patient, learn tricks and optimize



# Overfitting: Regularization L1

$$C = \text{fancy\_distance}(\text{Target}, \text{Output}) + L||\mathbf{W}||$$



# Overfitting: Regularization L2

$$C = \text{fancy\_distance}(\text{Target}, \text{Output}) + L||\mathbf{W}||^2$$