ADAM Algorithm

Gradient Descent

$$w_j = w_j - \alpha \frac{\partial}{\partial w_j} J(\vec{w}, b)$$

Adam: Adaptive Movement Estimation

model.compile
(optimizer = tf.keras.optimizers.Adam
(learning_rate = 1e-3),

$\alpha = 10^{-3}$

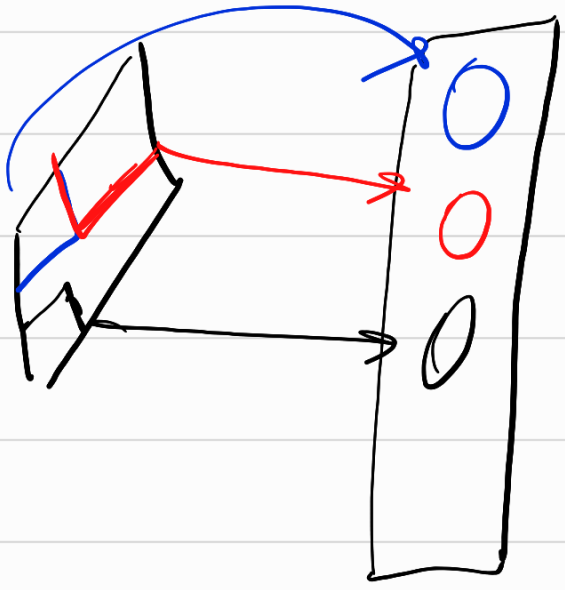
0.001

Additional Layer Types

Dense layer

→ neuron output is function of all activation output of previous layer

Convolutional Layer:



faster computation
less prone to overfitting

multiple → CNN layers