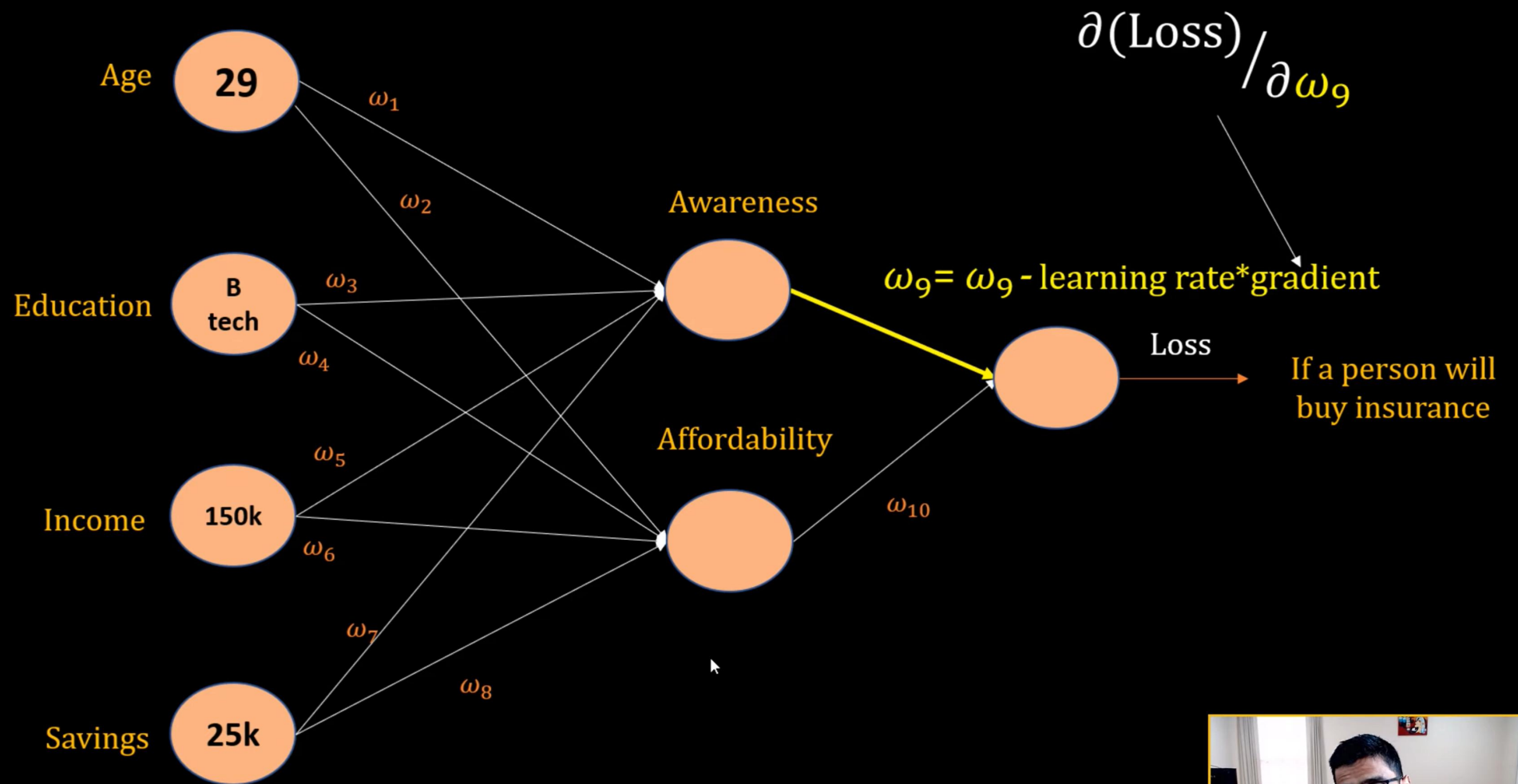


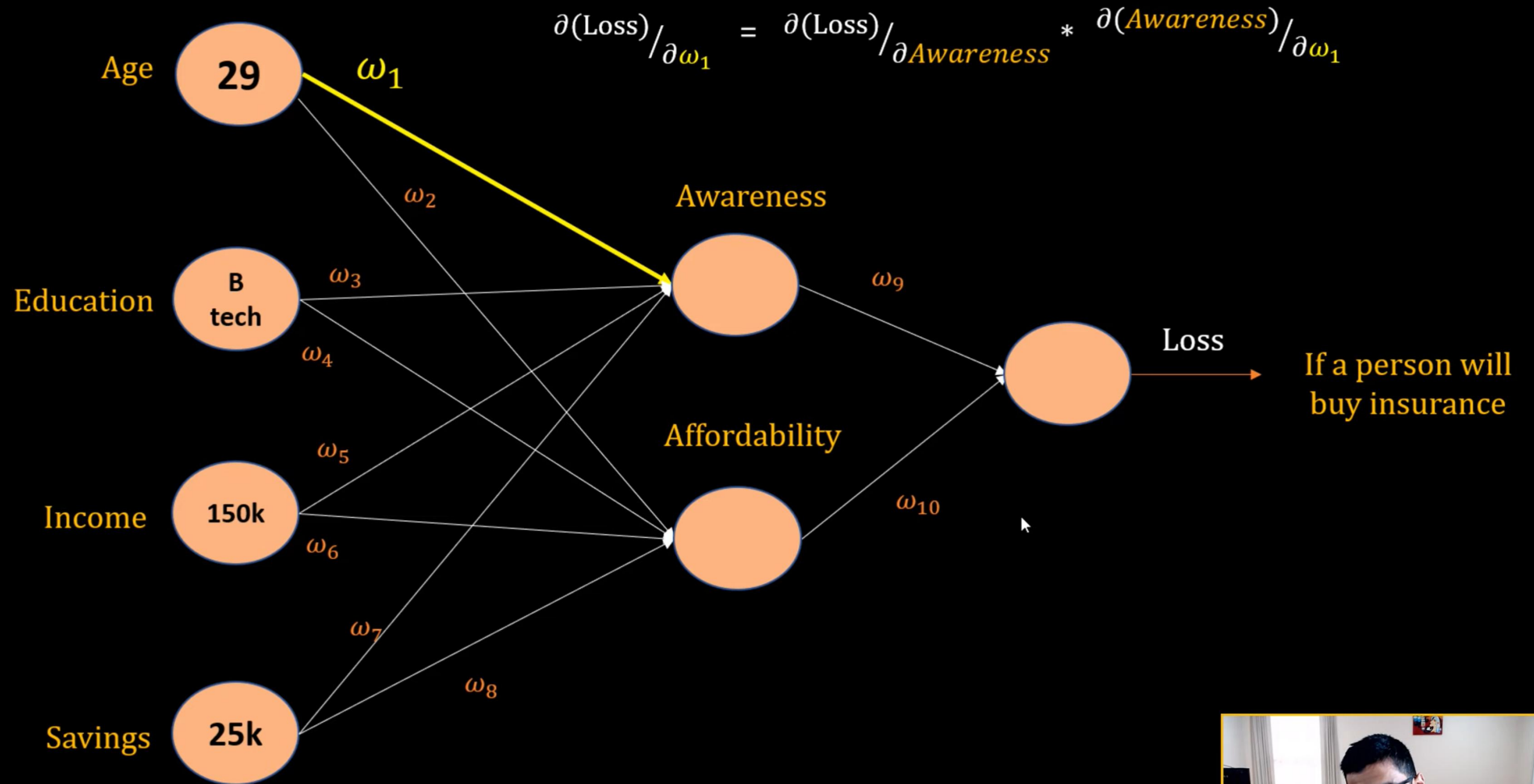
Train





Train





Train



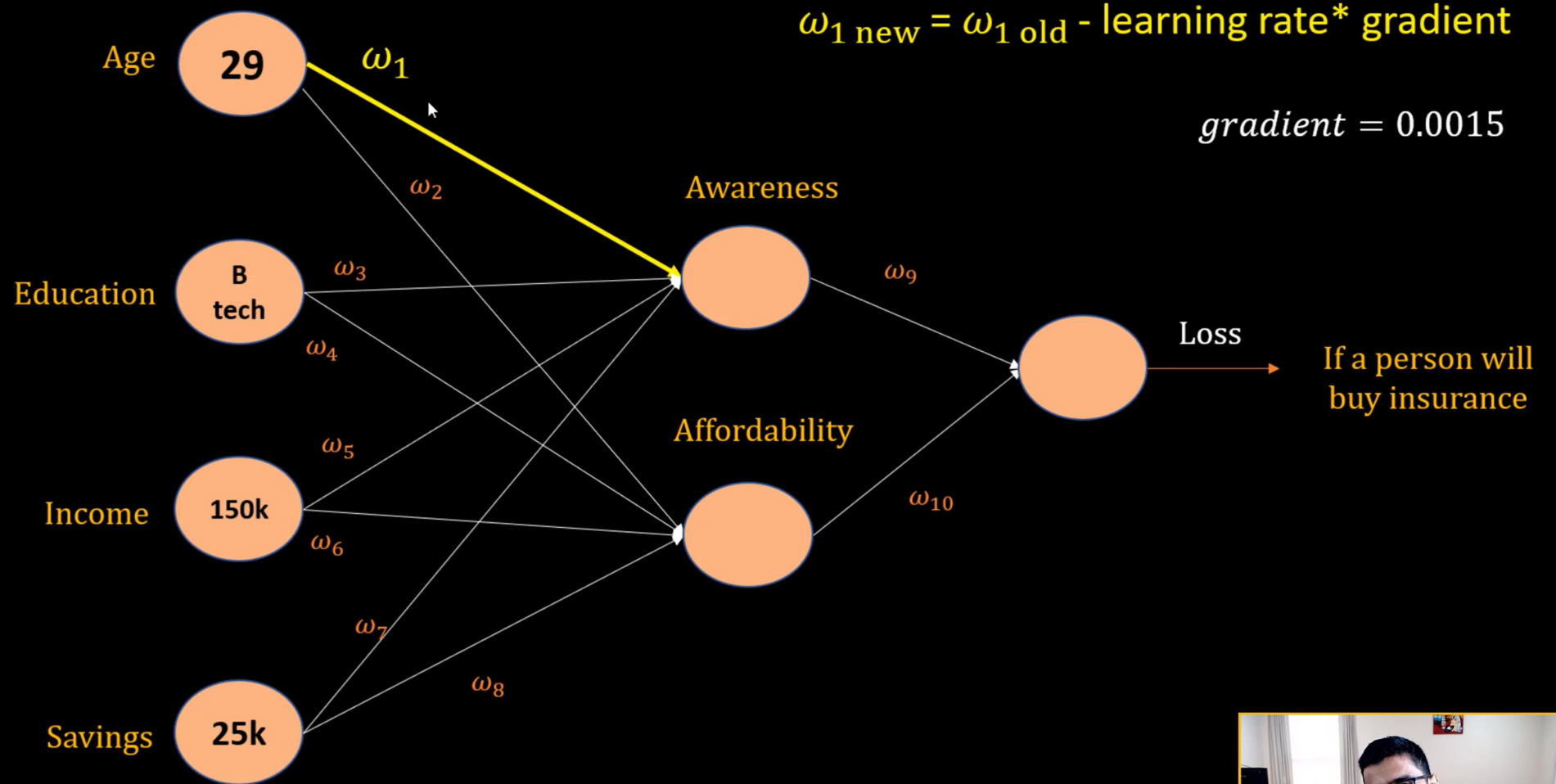
$$\frac{\partial(\text{Loss})}{\partial \omega_1} = \frac{\partial(\text{Loss})}{\partial \text{Awareness}} * \frac{\partial(\text{Awareness})}{\partial \omega_1}$$

$$\text{gradient} = d1 * d2$$

$$\text{gradient} = 0.03 * 0.05$$

$$\text{gradient} = 0.0015$$





Train



As number of hidden layers grow, gradient becomes very small and weights will hardly change . This will hamper the learning process.

Vanishing Gradients



$$\frac{\partial(\text{Loss})}{\partial \omega_1} = \frac{\partial(\text{Loss})}{\partial \text{Awareness}} * \frac{\partial(\text{Awareness})}{\partial \omega_1}$$

$$\text{gradient} = d1 * d2$$

$$\text{gradient} = 100 * 500$$

$$\text{gradient} = 50000$$



When individual derivatives are large, the final derivative will also become huge and weights would change drastically.

Exploding Gradients



$$\text{gradient} = d1 * d2 * d3 * d4 * \dots * dn$$

Vanishing gradient problem is more prominent in very deep neural networks.



Vanishing gradient problem in **RNN**

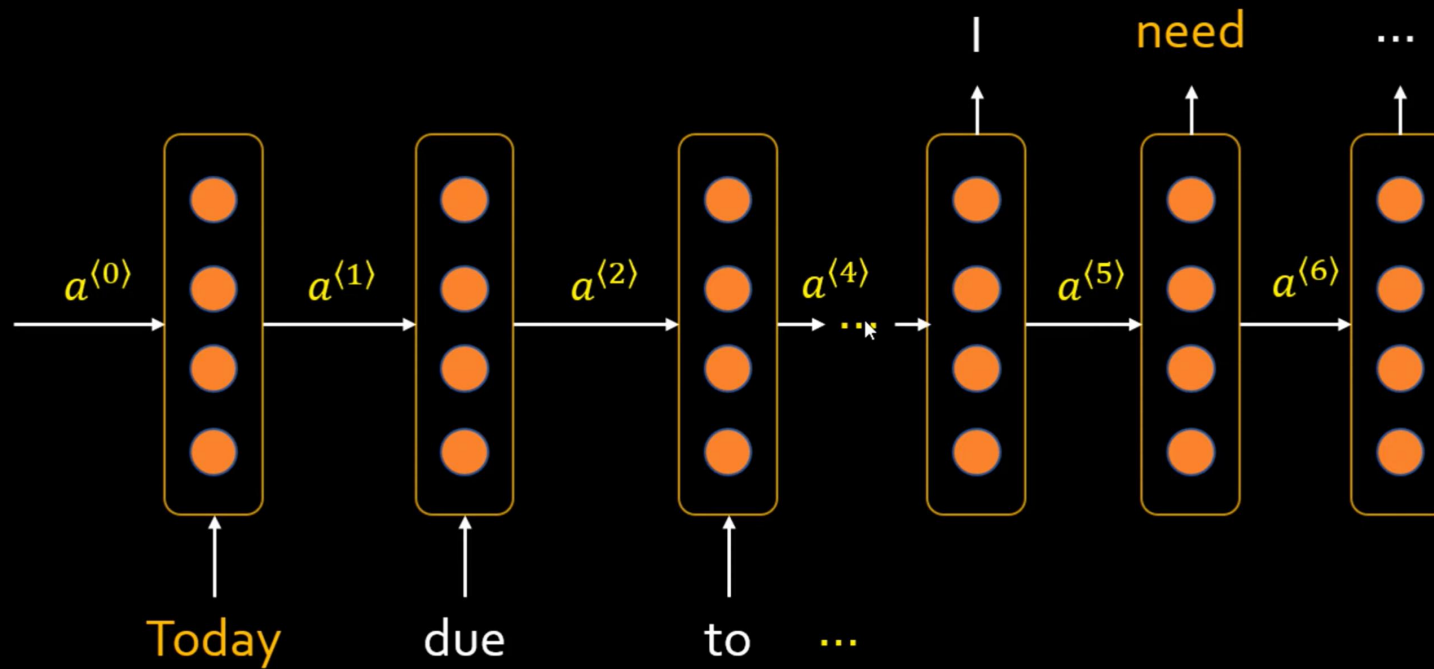


Today, due to my current job situation and family conditions, I need to take a loan.

Last year, due to my current job situation and family conditions, I had to take a loan.



Today, due to my current job situation and family conditions, I need to take a loan.



Solutions?

GRU

LSTM

