Let's call the cost function$F(x)$ and assume it is defined and differentiable for all values.

Let ∇F be the derivative of F.

Assume point a_n is our current location on the cost function, representing the weights and biases of the network.

Υbeing our learning rate such that: $\Upsilon \in \mathbb{R}_+$ resolves to the next point a_{n+1}being:

$a_{n+1} = a_n - \Upsilon \nabla F(a_n)$

and therefore we have a monotonic sequence such that:

$F(a_1) \geq F(a_2) \geq F(a_3) \geq ... \geq F(a_n)$

meaning that$F(a_n)$ will be a local minimum of the cost function,

and as such, our weights and biases array - which is represented by a_n- will be semi-idealized to the problem

$\Longrightarrow$The neural network has learnt to "solve" the given problem.