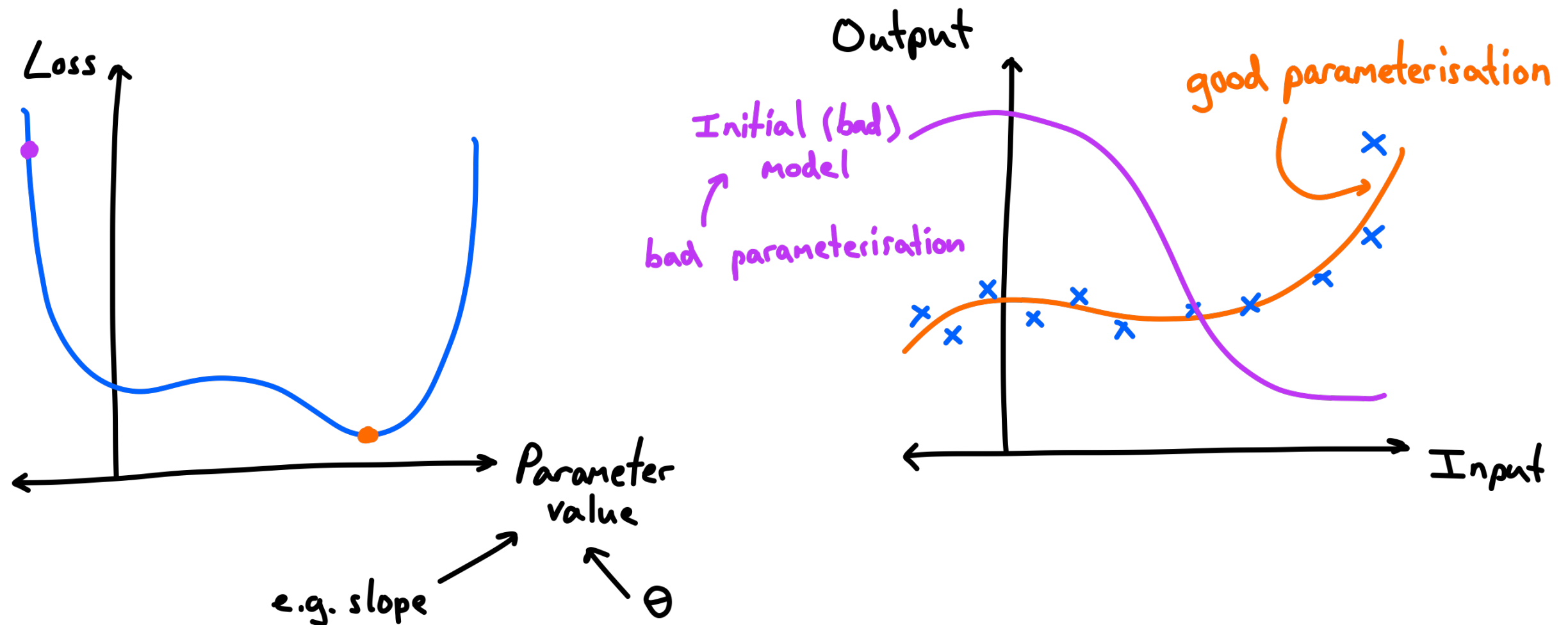


Loss functions

Remember gradient descent



Gradient descent: $\theta \leftarrow \theta - \alpha \frac{\partial L}{\partial \theta}$

We are using gradient descent to find the minima of some function. But what function?

Mean squared error - continuous output

examples in mini-batch

hypothesis = model output

$$L_{MSE} = \frac{1}{M} \sum_{i=1}^M (h - y)^2$$

Binary cross entropy - classification

$$L_{BCE} = - \left[y \log(\hat{y}) + (1-y) \log(1-\hat{y}) \right]$$

