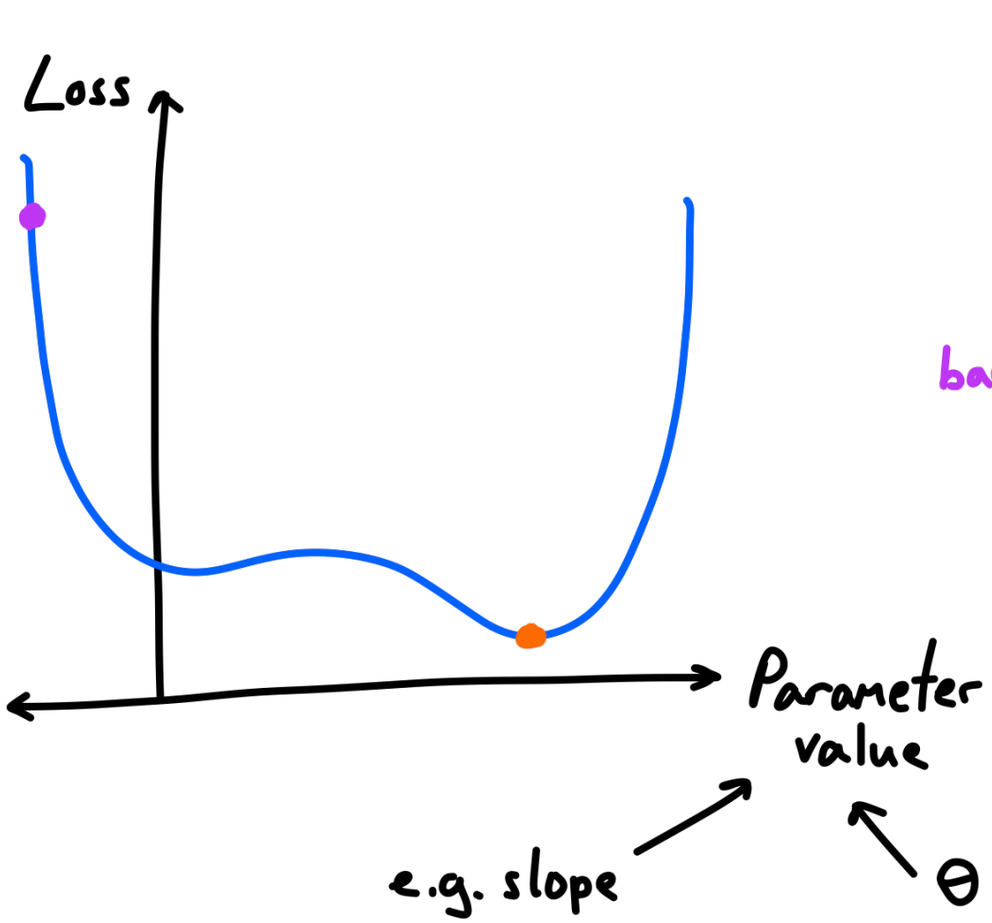
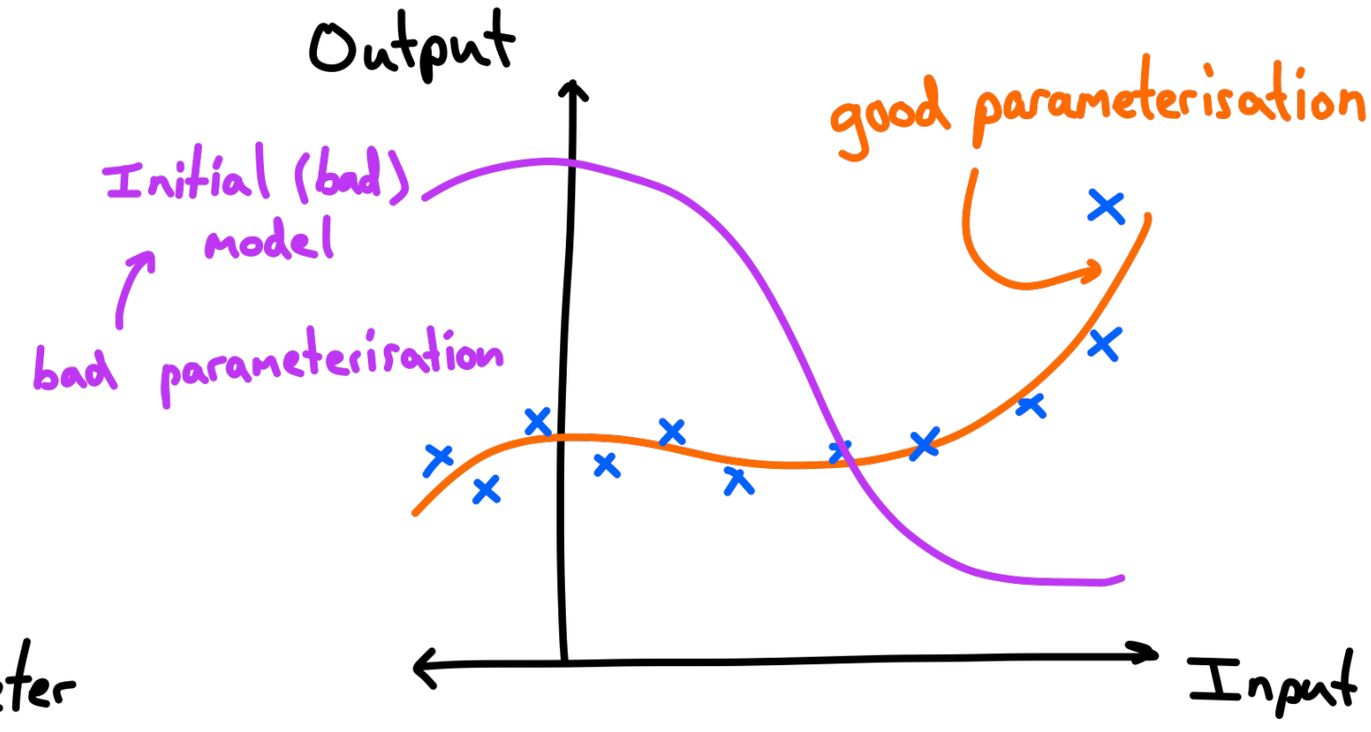
Loss functions

Remember gradient descent

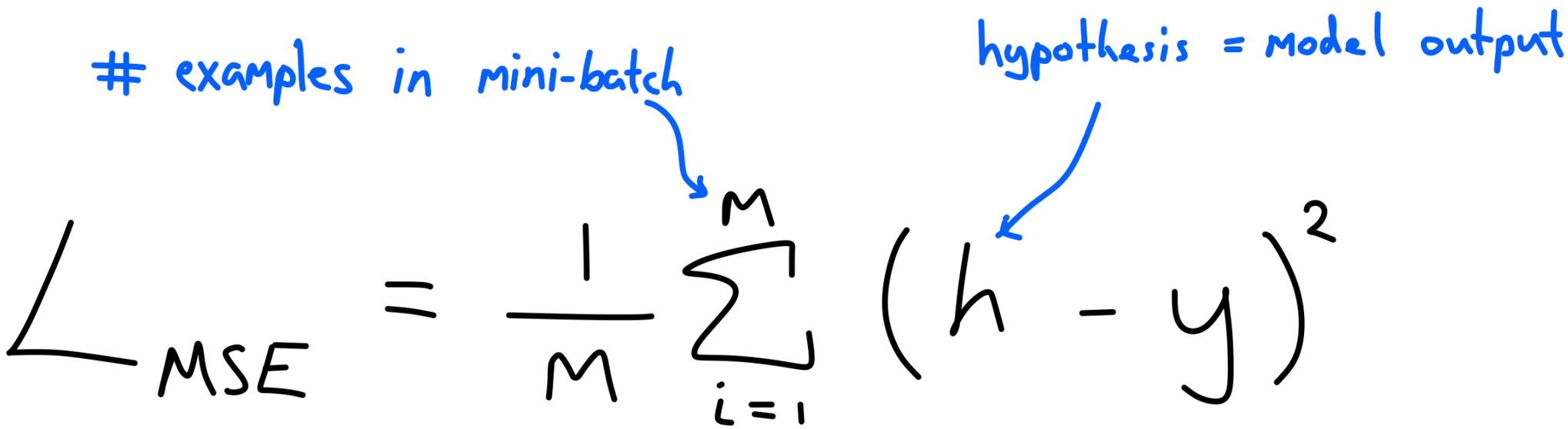




Gradient descent: $\theta \leftarrow \theta - \alpha \frac{\partial L}{\partial \theta}$

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

Mean squared error - continuous output



Binary cross entropy - classification

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