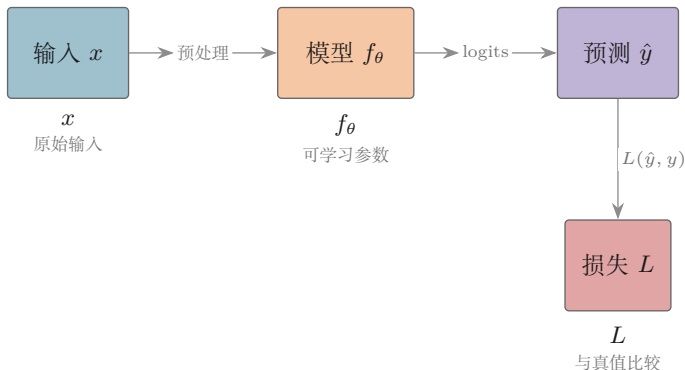


$$x \xrightarrow{f_{\theta}} \hat{y}, \quad \min_{\theta} L(f_{\theta}(x), y)$$

推理流程：一次前向



概念	一次从输入到预测的前向；损失在预测之后单独计算
对象	数据、模型参数、预测、损失
机制	预处理后送入模型；模型产生 logits 得到预测；损失比较预测与真值，梯度再回到参数