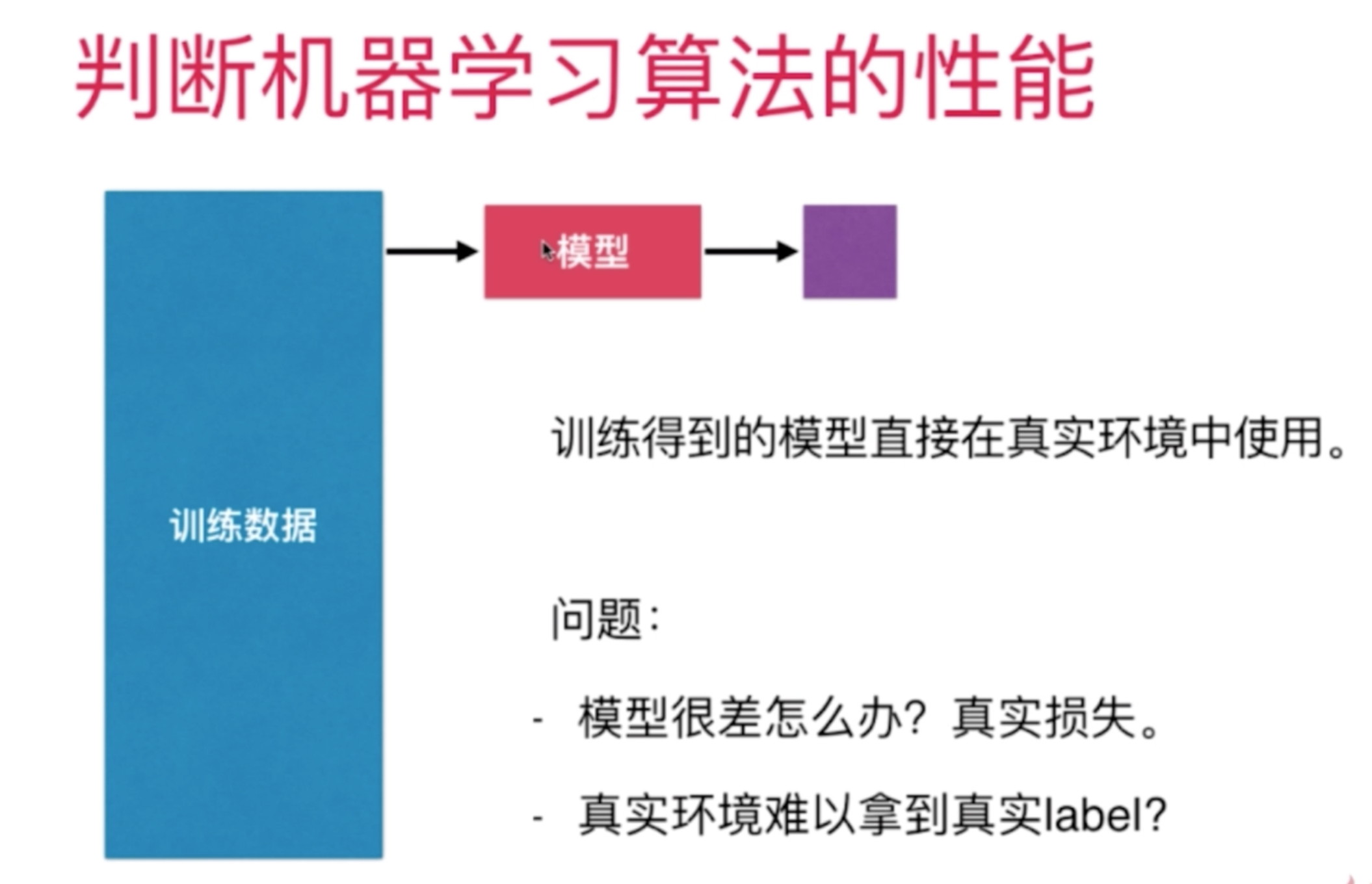


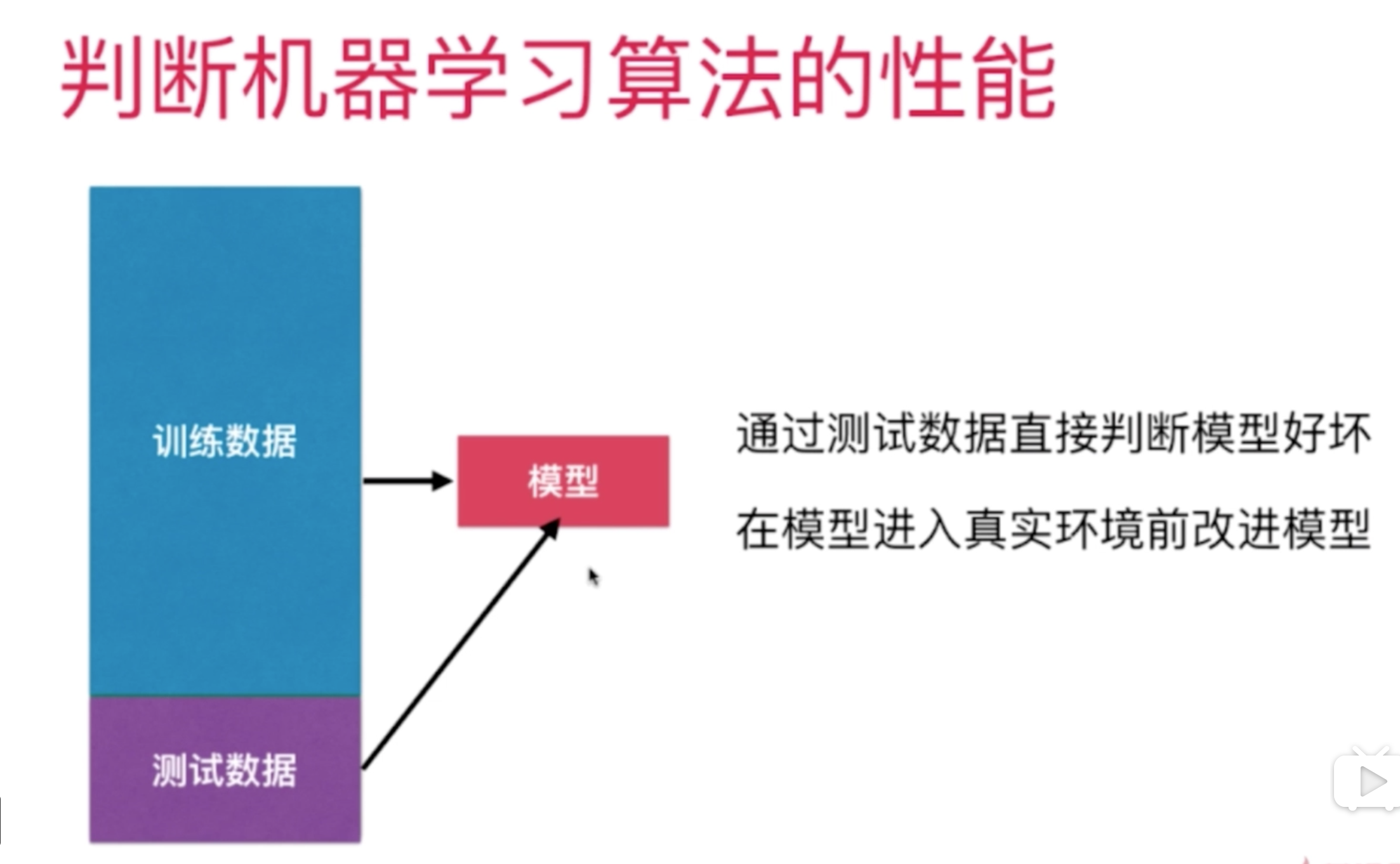


这是手里有的全部原始数据



**所以，把所有的原始数据都当做训练数据集，是不恰当的。**

解决的办法是把原始数据分为：训练数据集和测试数据集：



我们把这种把原始数据拆分做训练数据集和测试数据集的方式叫做：

train test split

接下来，我们采用 train test split 的方式来测试我们的算法：

[04-03 Train-Test-Split.ipynb]

还可以把刚才 ipynb 中实现的 train test split 逻辑包装成一个自己的函数：

[model\_selection.py]

End