**NP难问题？**

NP(Nondeterministic Polynomially，非确定性多项式)类问题是指一个复杂问题不能确定是否在[多项式](https://baike.baidu.com/item/%E5%A4%9A%E9%A1%B9%E5%BC%8F/10660961)时间内找到答案，但是可以在多项式时间内验证答案是否正确。NP类问题数量很大，如完全子图问题、[图着色问题](https://baike.baidu.com/item/%E5%9B%BE%E7%9D%80%E8%89%B2%E9%97%AE%E9%A2%98)、旅行商([TSP](https://baike.baidu.com/item/TSP/2905216))问题等。在P和NP问题中，P的难度最低，NP由于只对验证答案的时间作了限定，从而有可能包含某些无法在多项式时间内找到答案的问题，即NP是比P更困难的问题。

**如果数据量太大，比方1T或更大，单机无法全部加载，你怎么训练模型**

内存不够存取整个数据集，我们就得通过额外的IO开销来保持系统的均衡性。pytorch的pytorch.utils.data子库，就提供了直接创建loader的API，可以定义批次数据的batch大小，读取线程数，是否打乱等能够满足基本要求的参数。当然也提供了API来对数据做hold-out。

**数据量如果太少，你有什么办法解决。**

Data Augmentation[ɔːgmɛnˈteɪʃ(ə)n]：数据增强是根据已有的数据生成新的数据。与上述技术相反，数据增强从问题的根源（训练数据集）着手解决问题。使用数据增强方法扩展数据集不仅有助于利用有限的数据，还可以增加训练集的多样性，减少过度拟合并提高模型的泛化能力。

**二叉搜索树的优缺点**

二叉搜索树查找速度快、查找方便（logn时间）。但是其维护困难，且数据的插入和删除十分的复杂。