Week 8 BI Thinking

郭想

为什么股票预测问题容易出现过拟合，也就是在训练时结果很好，在真实环境中结果没那么好

因为在股票预测的时候数据集中噪音会比较多，可能占了信息量的90%，所以虽然模型仍然可以训练，但是会把这些噪音也学习进去，导致在后续的真实环境预测中结果不好。

Prophet与ARMA/ARIMA相比，优势在哪些地方

首先，我们可以先了解一下ARIMA和ARMA模型的不足。对于ARMA模型来说，它要求时序数据是稳定的（平稳序列），而现实中的数据一般都很难符合这一点因为可能会有一些黑天鹅事件导致某个时刻的数值偏高或者偏低。ARMA还容易在训练的时候过拟合导致真实环境下的预测没那么好。

对于ARIMA模型来说，虽然可以把非平稳数据转变为平稳序列，但它是一个线性模型，无法处理非线性关系，同时要求数据点的间隔等长。对于这两个模型来说，如果数据缺失，则需要使用插值等方法来预估缺失值，然后再使用预估值来进行参数拟合，这样可能会引入噪音，因为预估值不是实际的一个真实值。

Prophet是一个基于相加模型的时间预测，**可以精准的拟合非线性的周期趋势**，这一点是ARIMA模型无法做到的。**然后时序数据可以是不稳定的**，这一点又比ARMA要好。在周期性上能够对yearly，weekly和daily的周期性使用非线性拟合。最后Prophet模型还添加了holidays（影响因子），**可以很好的对节日带来的活跃数据进行预测**。