## 具有特征数据的联合补货问题

对于任何的时间，假设事件发生的顺序如下：

1. 在每个阶段的开始，决策者先看到特征数据例如历史收获量、商品被点击的数量，同时看到初始的库存量，其中
2. 接着决策者根据特征数据以及现有的库存数据，做出库存的补货量，其中
3. 真实的需求到来，由此可以得到下一周期的初始库存量为=
4. 在经历了完整的周期之后，会出现本周期的成本之和，其中K为固定成本，v为每单位的进货成本，h为持有成本，b为惩罚成本。

由上述问题描述，根据MDP问题，可以得到：

1. State =[]
2. Decision=

其中

1. Exogenous information variables=
2. Transition function =
3. Objective function

## 对于（s,c,S）这类策略的考虑

按照老师上次开会的说法，也是一个policy gradient的方法，其实也是一种ADP或者说的强化学习的方法。首先random出（s，c，S）的值，每迭代K个周期对参数进行梯度下降，一直下降到梯度下降不带来总支出的减少的时候，完成迭代，是这样的思路吗？这样就可以避开类似demand distribution的预测这种