# imitation learning

## behavior cloning（lecture2 part2）

定义：直接使用监督学习的方法，将专家数据看作样本输入

### 目标

我们在下训练，但是我们用测试，但

所以不考虑以下目标函数

首先，定义损失函数为

我们关心的是情况，我们需要做的是

$$\mathrm{minimize~}E\_{\mathbf{s}\_t\thicksim p\_{\pi\_\theta}(\mathbf{s}\_t)}[c(\mathbf{s}\_t,\mathbf{a}\_t)]\qquad\text{(3)}$$

假设：，对于

因此有

这个式子表明BC的cost会随着决策步数的增加而呈现平方次增加

更加泛化的说，对于，我们假设

对于，我们有

式子前一项表明不犯错（即train data）的概率，后一项表明其他分布

然后有

其中，所以小于

对于我们的目标函数来说