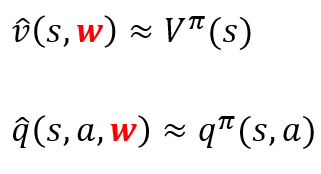
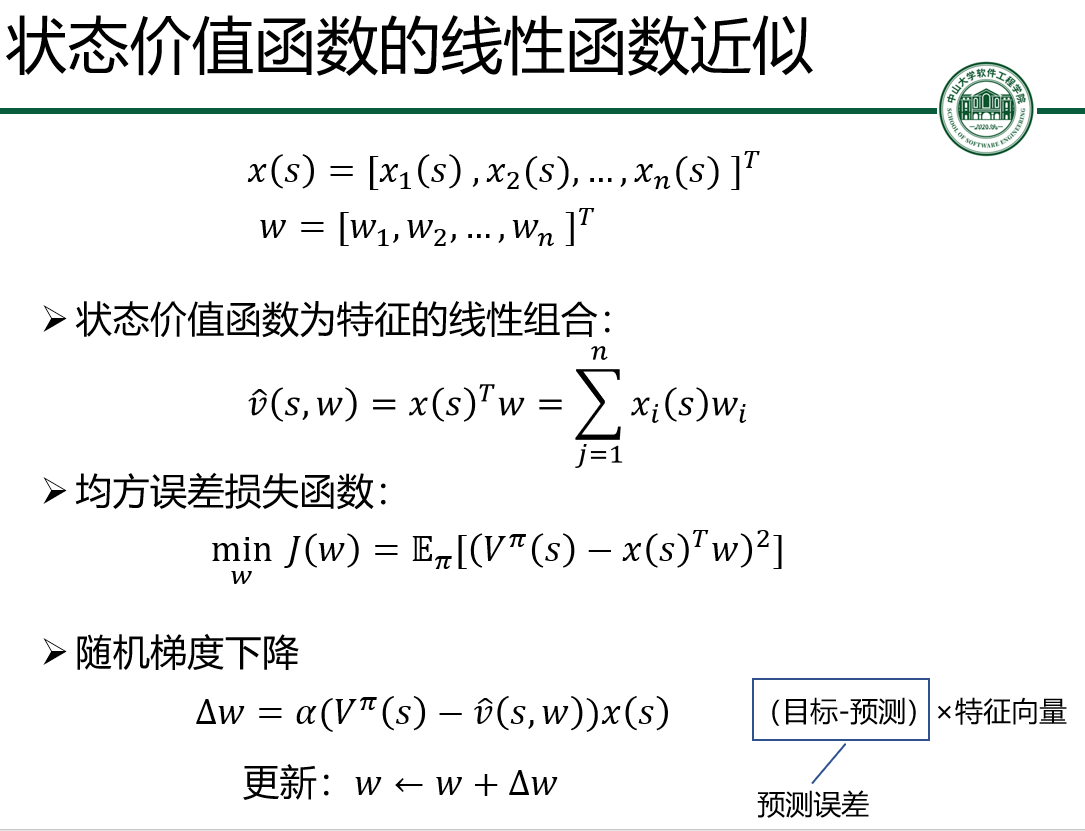
状态价值函数和动作价值函数可有表储存：

函数近似：解决大规模MDP存储空间大、学习效率低的问题。

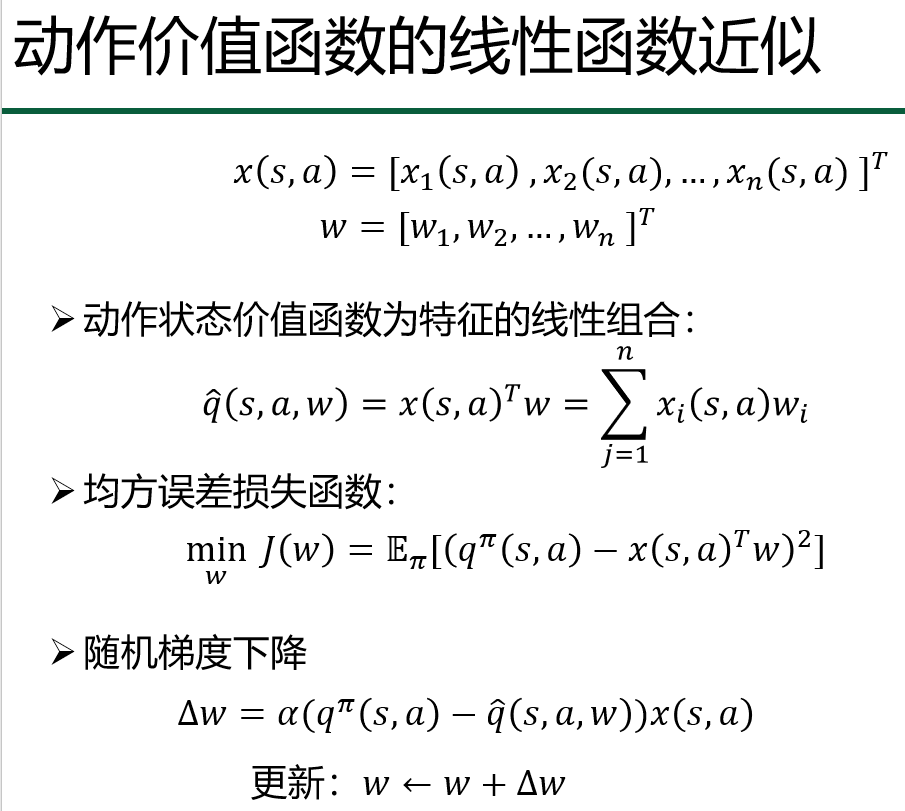
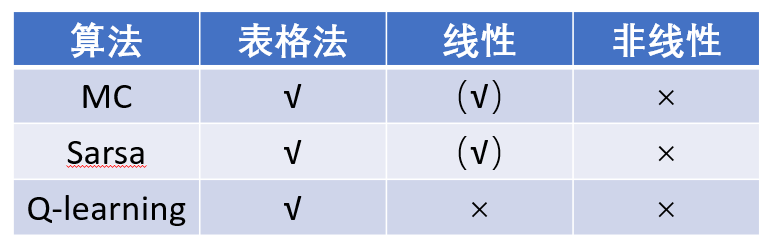
近似函数比实际函数多一个待定参数w用来学习拟合。

常用的函数近似器

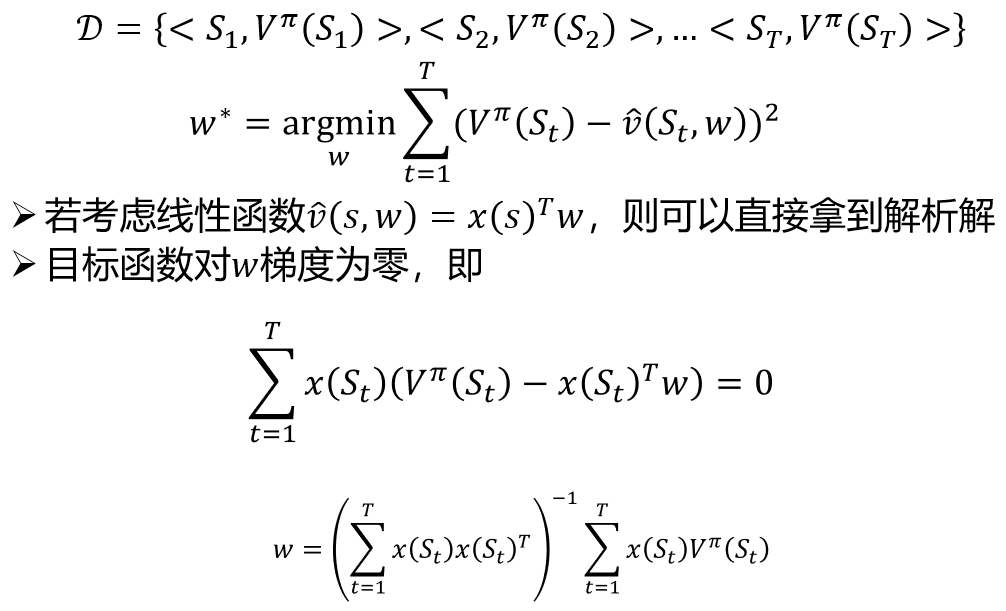
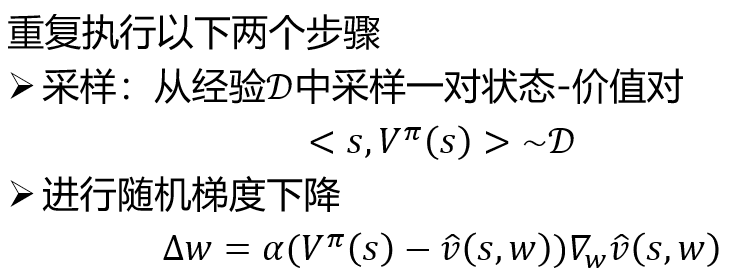
* 特征的线性组合（线性模型）：可用于MC、TD

拟合公式：

Ppt p15？

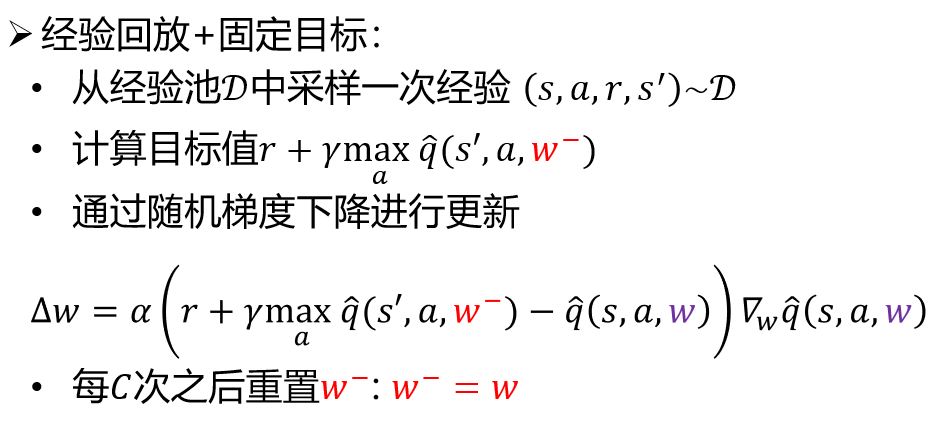
收敛性比较：

学习策略：

* + 最小二乘 (Least Square)预测（批量学习）
  + 迭代法：随机梯度下降（增量学习）
* 神经网络：Deep Q-Learning (DQN)

经验回放：在回放式的经验池中采样学习样本，以减少样本之间的关联性。

固定目标：每𝐶次迭代内固定目标保持不变，以增强学习的稳定性、减少震荡。

 具体实现：

DQN 的改进:

* + Double DQN
  + Dueling DQN
  + Prioritized Experience Replay
  + 其他…

整体理解？？