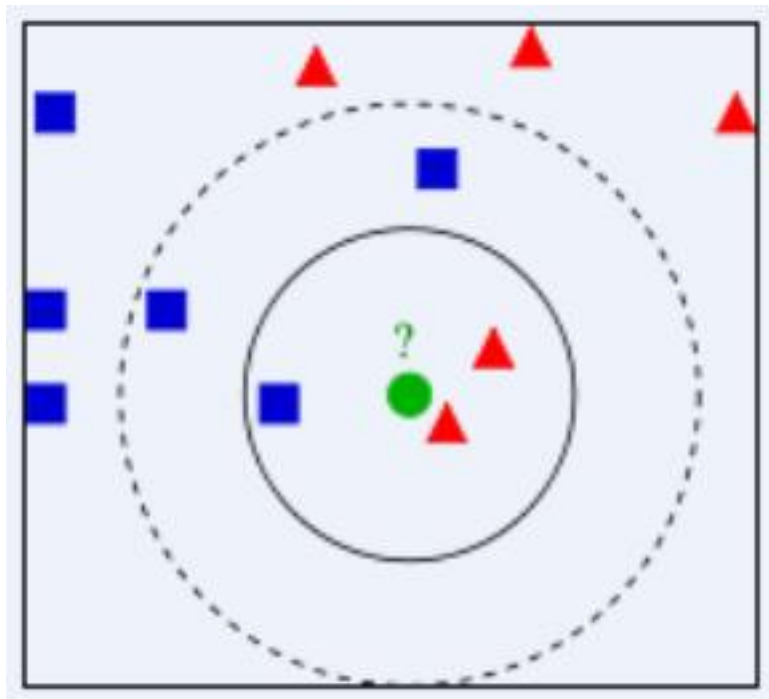


机器学习模型类型

- 有监督学习KNN（重点）
- 无监督学习K均值聚类（重点）
- 半监督学习
- 强化学习
- 自监督学习

有监督学习KNN

- K最近邻(k-Nearest Neighbor, KNN), 是一种常用于分类的算法, 是有成熟理论支撑的、较为简单的经典机器学习算法之一。
- **基本思路:** 如果一个待分类样本在特征空间中的k个最相似(即特征空间中K近邻)的样本中的大多数属于某一个类别, 则该样本也属于这个类别。即近朱者赤, 近墨者黑。
- 需要样本标签的支持, 是一种有监督学习算法。



- 所有样本可以使用一个二维向量表征。图中，蓝色方形样本和红色三角形样本为已知分类样本。绿色圆形为待分类测试样本。
- 当 $K=3$ 时，其3近邻中有2个红色三角形样本和1个蓝色方形样本，因此预测该待分类样本为红色三角形样本
- 当 $K=5$ 时，其5近邻中有3个红色三角形样本和2个蓝色方形样本，因此预测该待分类样本为蓝色方形样本

- **数据集：** 即必须存在一个样本数据集，也称作训练集，样本数据集中每个样本是有标签的，即我们知道样本数据集中每一个样本的标签。
- **样本的向量表示：** 即不管是当前已知的样本数据集，还是将来可能出现的待分类样本，都必须可以用向量的形式加以表征。向量的每一个维度，刻画样本的一个特征，必须是量化的，可比较的。
- **样本间距离的计算方法：** 欧氏距离、余弦距离、海明距离、曼哈顿距离等等。

■K值的选取：K值的选取会影响待分类样本的分类结果

- ✓K值较小：K值的减小就意味着更复杂的决策边界，每个训练样本都会形成一个决策模型，容易发生过拟合。
- ✓K值较大：K值的增大就意味着整体的模型变得简单，会导致偏差变大。比如k为总的训练样本数量，那么每次投票肯定都是训练数据中多的类别获胜。

■KNN的优点

- ✓简单，易于理解，易于实现，无需参数估计，无需训练
- ✓对异常值不敏感（个别噪音数据对结果的影响不是很大）
- ✓适合对稀有事件进行分类
- ✓适合于多分类问题

■KNN的缺点

- ✓计算量大，内存开销大。
- ✓可解释性差。无法告诉你哪个样本更重要
- ✓K值的选择。当样本不平衡时会导致错误
- ✓KNN是一种消极学习方法、懒惰算法

无监督学习K均值聚类

- “类”指的是具有相似性的集合。聚类是指将数据集划分为若干类，使得类内之间的数据最为相似，各类之间的数据相似度差别尽可能大。
- k-means算法是一种简单的迭代型聚类算法，采用距离作为相似性指标，从而发现给定数据集中的K个类，且每个类的中心是根据类中所有值的均值得到，每个类用聚类中心来描述。聚类优化目标函数：

$$J = \sum_{k=1}^K \sum_{i=1}^{n_k} \|x_i - u_k\|^2$$

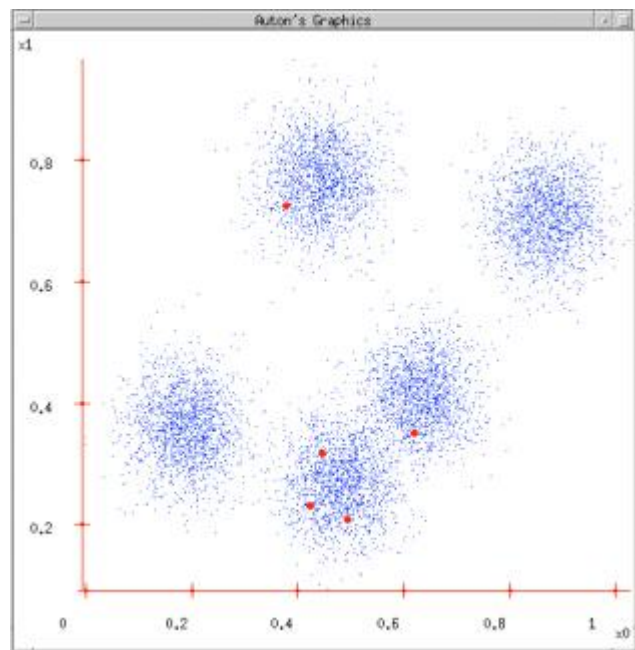
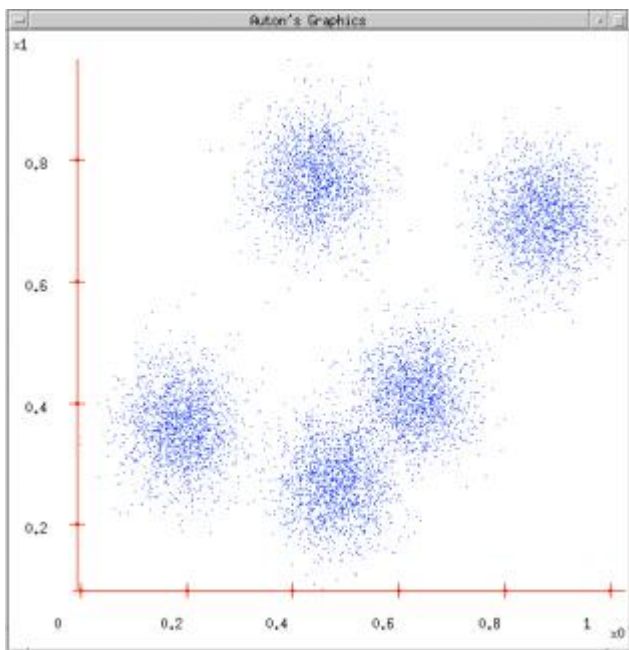
其中K为类别总数量， x_i 为当前类别中的第i个样本， u_k 为第k个类的均值， n_k 为第k个类内的样本数量

■算法分为四个步骤：

- ① 选取数据空间中的K个对象作为初始中心，每个对象代表一个聚类中心；
- ② 对于样本中的数据对象，根据它们与这些聚类中心的欧氏距离，按距离最近的准则将它们分到距离它们最近的聚类中心（最相似）所对应的类；
- ③ 更新聚类中心：将每个类别中所有对象所对应的均值作为该类别的聚类中心，计算目标函数的值；
- ④ 判断聚类中心和目标函数的值是否发生改变，若不变，则输出结果，若改变，则返回2) 。

■ 举例

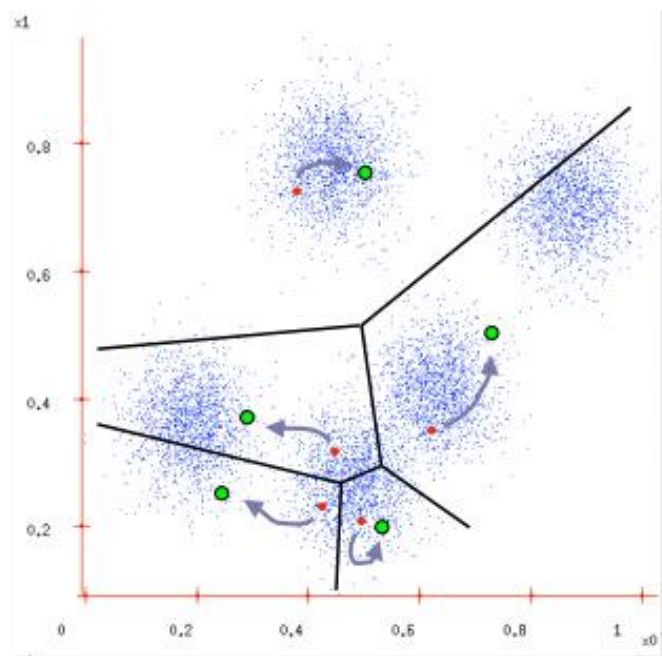
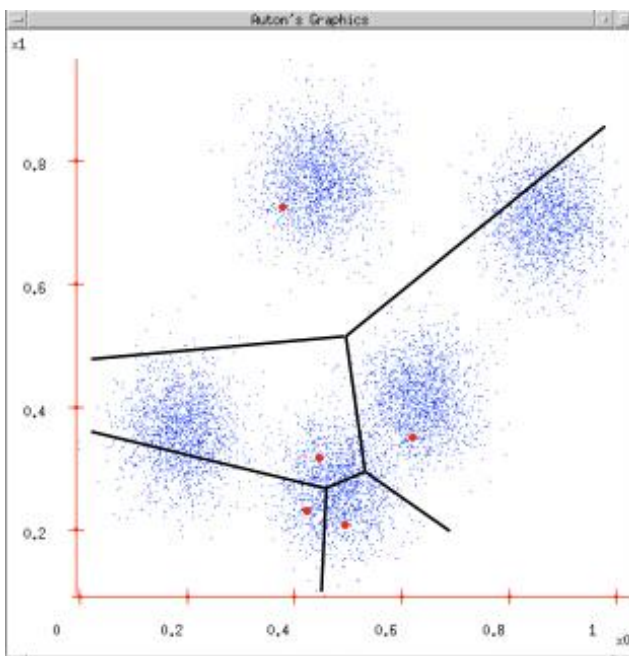
给定数据集



K=5

选择种子点

样本分类



更新聚类中心

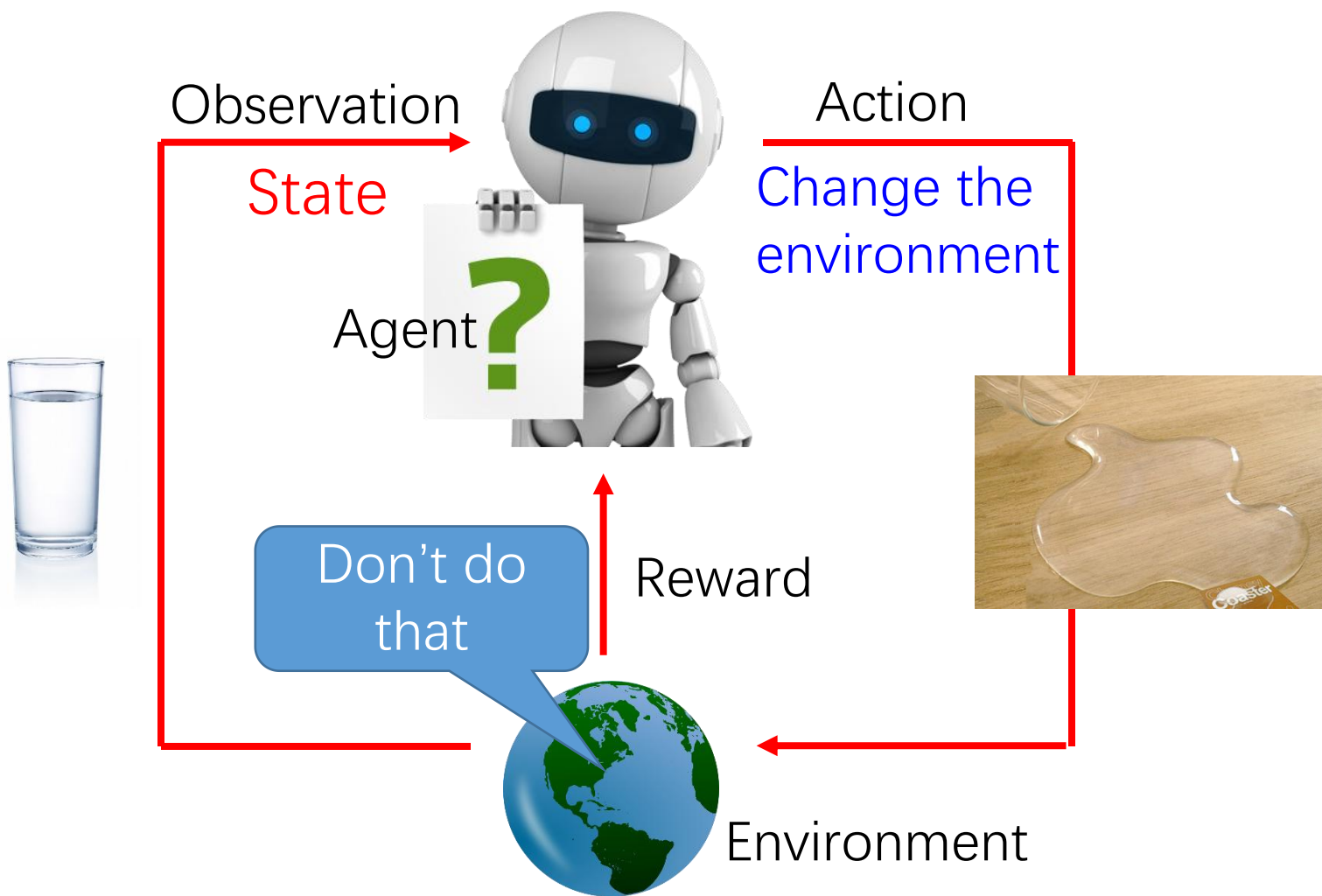
半监督学习

- 只有少量样本带标签
- 更符合实际需求

强化学习

- 单步没有标签
- 流程走完时有标签
- 常用于游戏等人工智能应用中

强化学习的基本模式

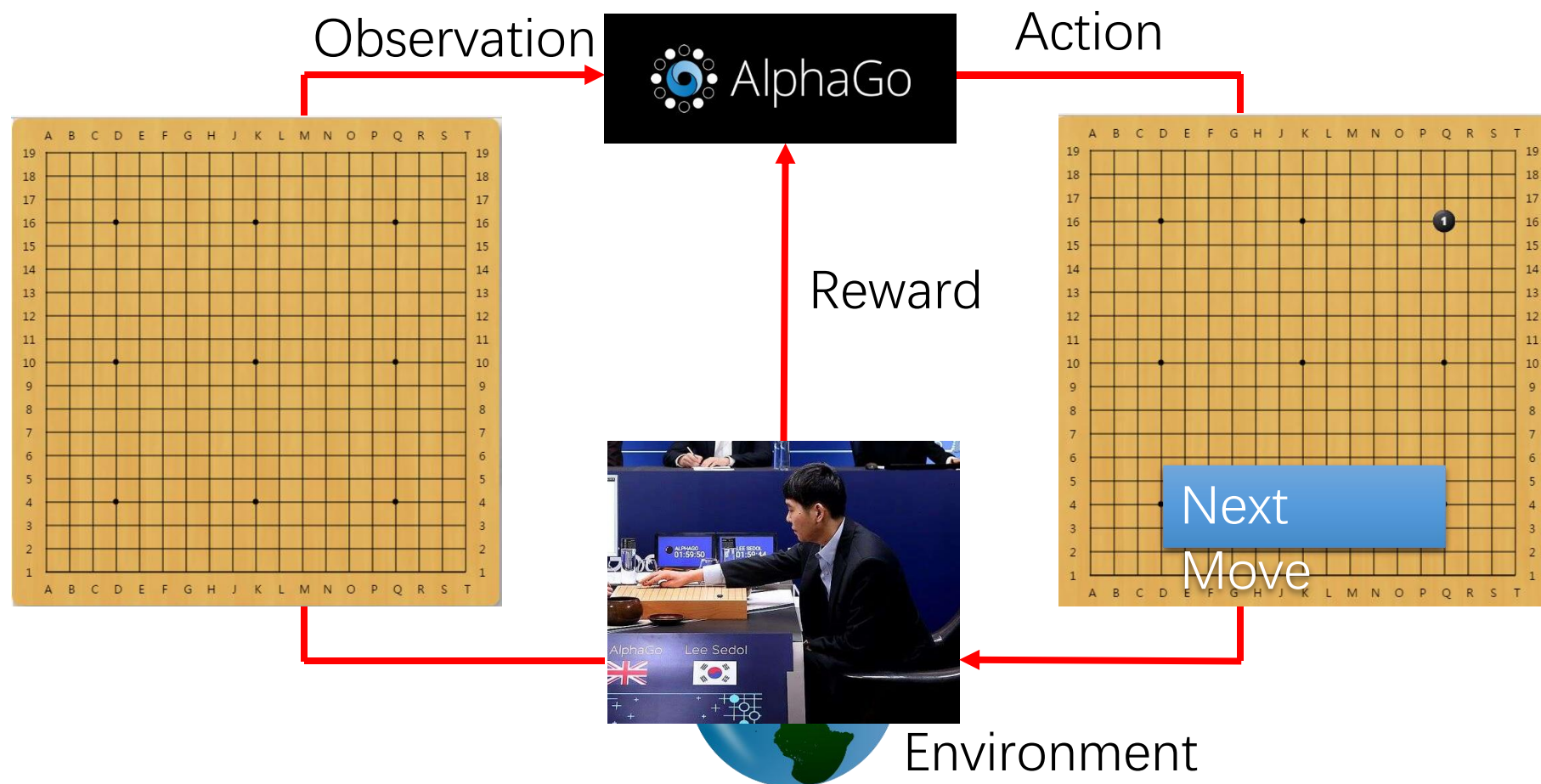


强化学习的基本模式

Agent learns to take actions to maximize expected reward.

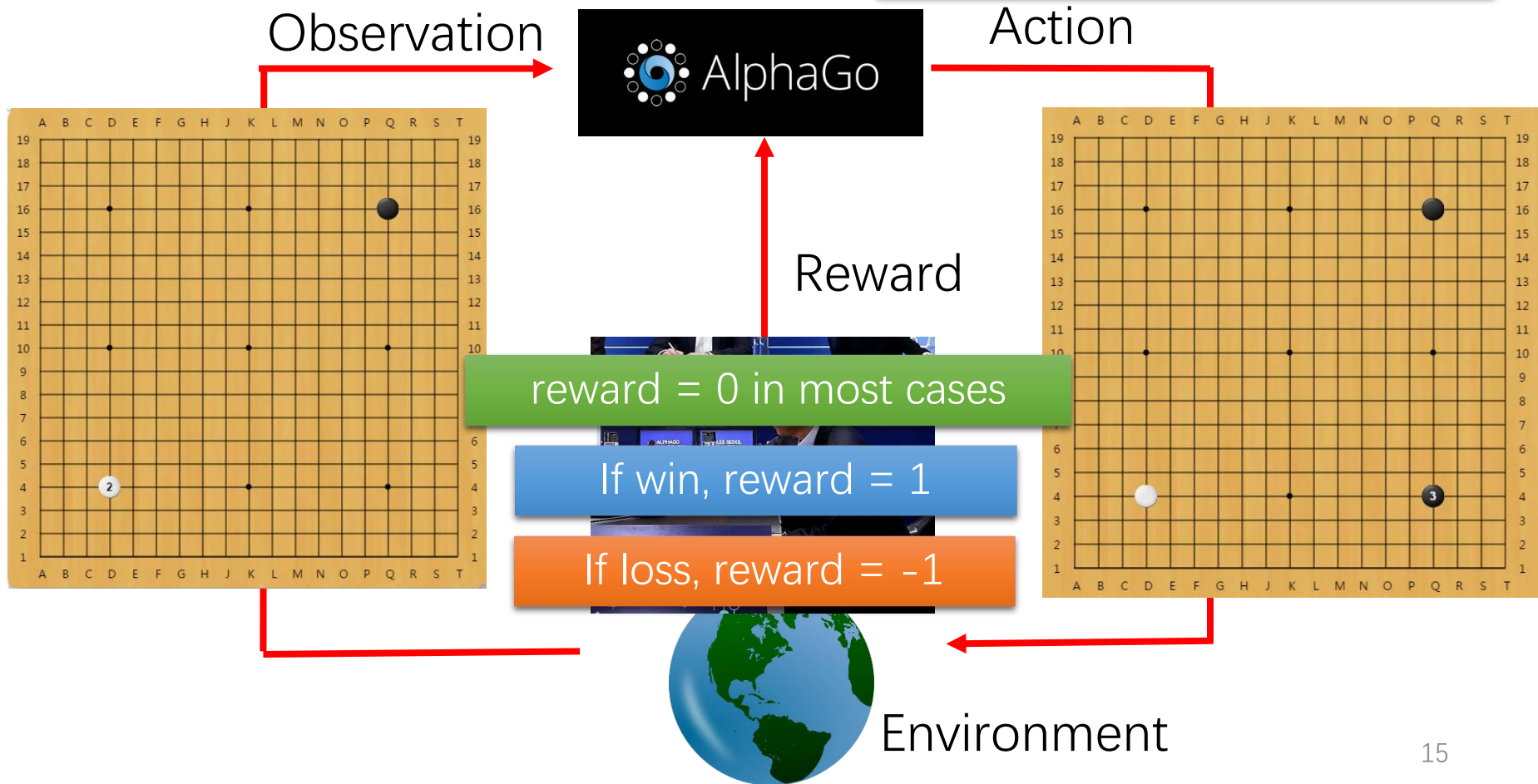


强化学习的应用场景



强化学习的应用场景

Agent learns to take actions to maximize expected reward.



- Supervised:

Learning from teacher



Next move:
"5-5"



Next
move:
"3-3"

- Reinforcement Learning

Learning from experience

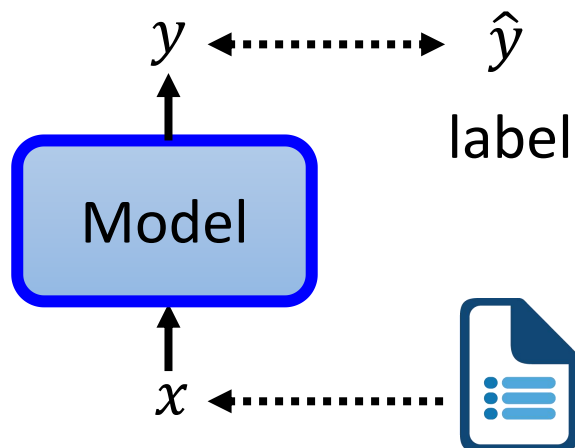
First move ➡ many moves ➡ Win!

(Two agents play with each other.)

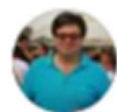
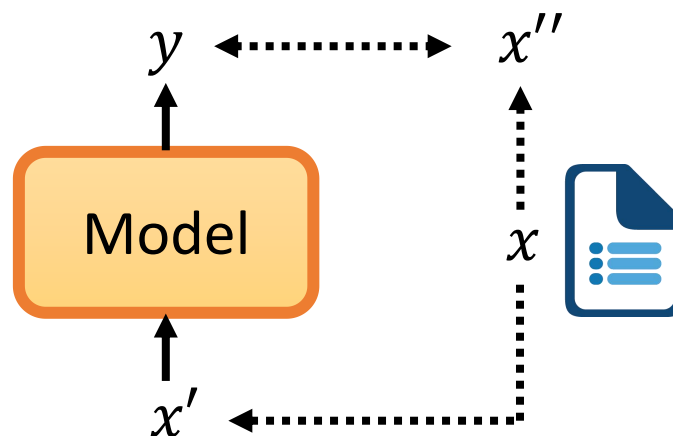
Alpha Go: supervised learning +
reinforcement learning.

自监督学习

Supervised



Self-supervised



Yann LeCun

2019年4月30日 · 🌐

I now call it "self-supervised learning" because "unsupervised" is both a loaded and confusing term.

In self-supervised learning, the system learns to predict part of its input from other parts of its input. In other words a portion of the input is used as a supervisory signal to a predictor fed with the remaining portion of the input.