

问题: 在你的报告中观察并记录智能车在采取随机动作时的行为。它最终到达目标位置了吗？还有什么其他有趣的现象值得记录下来？

Answer: 当 `enforce_deadline = False` 的时候，在 100 次试验中大约有 20 次，智能车能够在置顶时间内到达为位置，即约有 20% 的概率到达目标位置。当 `enforce_deadline = True` 的时候，约有 15% 的概率在 `deadline` 前到达目标位置。两种情况相比较，说明采取随机动作有可能做出最优选择，也有可能做出的是较差的选择。所以当陷入 `local optimal` 的时候可以采取 `random pick` 的方式来摆脱这种困境，这是某种意义上的随机游走。

在 `random pick` 的过程中，可以发现大多数情况下智能车都是无视交通规则的，即不分红绿灯，允许两车相撞。

问题: 你认为哪些状态适合对智能出租车和环境建模？为什么你认为在这个题目中这些状态是合适的？

Answer: 在建模过程中，字典 `inputs` 中的 `key` 和 `next_waypoint` 对建模都很重要。在 `Inputs` 中，`light` 的状态告诉了智能车如何在遵守交通规则的情况下免收惩罚。`Oncoming`, `left` 和 `right` 的状态则告诉了智能车其他智能车相对于其的位置，从而避免因相撞而受到惩罚。`next_waypoint` 则最为重要，因为它告诉了智能车相对于目的地的方向，从而在学习的过程中做出逼近目标位置的选择。

在实际建模过程中，一开始仅仅使用了 `inputs` 中的四个状态而并没有使用 `next_waypoint`，智能车会陷入在某个位置不断重复一种动作的行为。这是因为并没不知道目标位置相对于智能车的位置，使得其认为当前选择的动作始终是最优动作从而陷入不断重复。

共有可能的状态为：48(一共 48 个点) 2(`light` 对应的两种状态) 3(`left` 对应的三种状态) 3(`right` 对应的三种状态) 3(`next_waypoint` 对应的三种状态) = 2592

问题: 与一直选择随机动作相比，你发现智能车的行为有了什么样的变化？为什么会发生这种变化？

Answer: 与随机动作相比，智能车在相同的实验中到达目标位值得次数更多，根据 `Q_Learning` 学习公式可知: $Q'(s,a) = \text{reward} + \max(Q'(s,a)) * \text{gama}$ ，通过学习可以获得一个较优状态，并根据其进行选择一个最优的路线达到目标位置。

问题：把你实现基本 **Q-Learning** 时的参数调节过程记下来。哪个参数组合智能车表现最好？它最终的表现有多好？

在这部分我试验了不同的参数： $\text{gama} = 0$ ，调节 α ； $\alpha = 0$ ，调节 gama ；从上述组合中挑选较好的 α 和 gama 与较差组合进行比较。

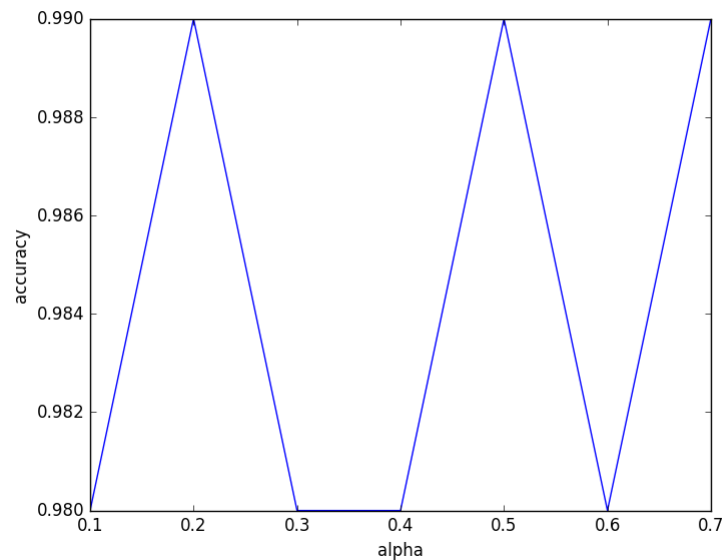


Fig1 gama=0，调节 alpha

从图中可以看出，当 α 从 0.1 到 0.7 变化时，accuracy 仅仅在 0.98 和 0.99 中变化，最好的表现为 0.99。

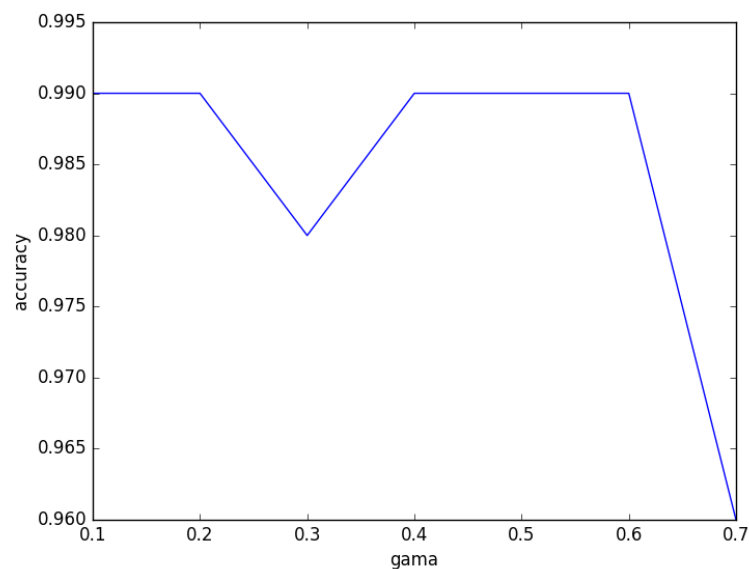


Fig2 alpha=0，调节 gama

从上图中可以看出， gama 从 0.1 到 0.7 变化时，最大的 accuracy 为 0.99，最小的为 0.96. 其中当 gama 从 0.6 到 0.7 变化时，accuracy 下降的比例最大。

针对上述图，选择 α 和 gama 的参数调整组合分别为 (0.1,0.1), (0.1,0.2), (0.2,0.2), (0.3,0.3), (0.4,0.4), (0.5,0.5)

Alpha-Gamma	Accuracy
(0.1,0.1,)	0.99
(0.1,0.2)	0.95
(0.2,0.2)	0.96
(0.3,0.3)	0.99
(0.4,0.4)	0.98
(0.5,0.5)	0.97

从上表可以看出，最大的 accuracy 为 0.99，最小的为 0.96，且 alpha 和 gama 的最优组合为(0.1,0.1)和(0.3,0.3)，从较少的组合中并不能明显发现 alpha 和 gamma 间的关系，如果想要进一步探索，需要更多参数组合。所以综上所述，在有限的参数组合中智能车能够达到的最好表现是 99%的准确率。

问题：你觉得你的智能车已经几乎找到了最佳策略吗？例如，能够在最短时间内到达目的地，不遇到任何惩罚。在这个问题中，你觉得应该怎样定义最佳策略？

我的智能车已经找到了最佳策略，能以 99%准确率在 deadline 前到达目的地。我觉得最佳策略的应该能够在某种程度上达到准时到达目的地，遵守交通规则，避免行车事故和考虑特殊情况的权衡。比如说，在距离 deadline 只有 1 天了，然后在可以避免行车事故的情况下是否可以考虑接受违反交通规则的惩罚，换取及时到达目的地的激励。因此，我认为最佳的策略应该是尽快，遵守规则，避免事故，考虑意外等几种因素的权衡。