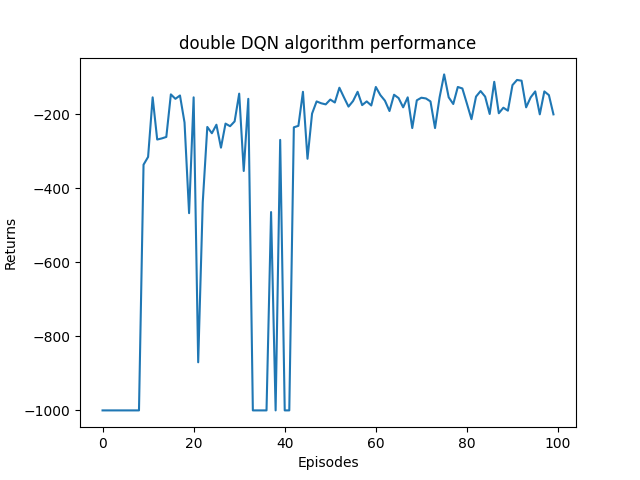
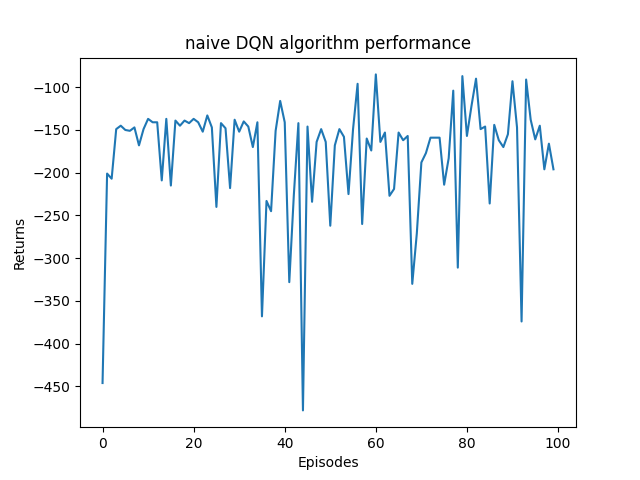
**强化学习作业A4 DQN**

叶增渝 123033910090

本次采用了传统的DQN网络与DoubleDQN网络。

由于MountainCar-v0下，除非达到终点可以得到reward为0，否则在每一个位置上得到的reward均为-1，不利于网络训练拟合，所以重新设计了reward，根据小车实际到达的高度，当高度大于0时，会给予一个与高度相关的reward，当达到终点时，会给予一个1000的高reward以保证训练的有效性，我们记此reward为virtual\_reward。

下图为2个DQN的训练的实际reward（在环境中，真实的reward代表了步数，我们设置训练时的maxstep为1000）：



由于任务简单容易学习，可以看到DQN网络可以在前几个episode就可以快速完成收敛，然后就能顺利地完成任务，完成任务的所需步数在100~200步之间。

而由于我们设计的不够完美reward不够完美，可能导致DQN在左右来回晃荡但不到达终点，以此来刷分，如DoubleDQN中间几个episode所示，但是最终还是能稳定在200步以内。

我们不难发现DoubleDQN的训练过程比普通的DQN更加曲折，这是由于任务足够简单，导致两者没有什么分别，但DoubleDQN的机制使得其收敛速度比较慢，所以需要更长时间收敛。

最后我们对两个记录下来的best\_model进行10次的测试（测试条件较为苛刻，需要在200步内完成任务），普通DQN的成绩为4次通过，平均实际reward为-182.4；而DoubleDQN的成绩为全部通过，平均实际reward为-124.5。

综上来看，DoubleDQN的最终效果较好，但两者均有完成任务的能力。

代码与对应的模型均放在文件中，可以直接使用ipynb的最后一个cell进行测试（但要记得将前面的环境与函数定义运行一遍，训练过程可以不用运行）