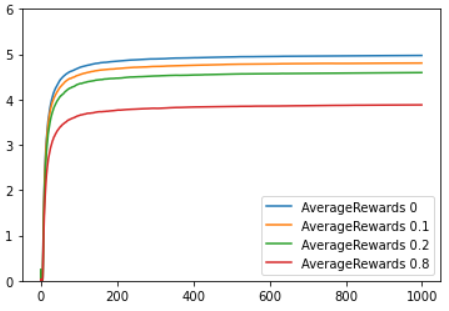
108061217鍾永桓

Q1:因為本題action的reward間差距較大，因此基本上只要每個action都嘗試後就找到了最佳的action，幾乎不用探索，若設為更大的epsilon 0.2或0.8，反而花太多次於探索上，選最佳action的次數很少，導致最終的aceragereward減少。



Q2:如果要在Epsilon為0時可以得到最佳答案，也就表示過程中完全不會進行exploration，必須要能確保所選的action是最佳的，所以可行的方法應該是在開始選擇前要先純粹進行exploration，也就是不斷隨機選擇，評估每個actionvalue，等到從每個action取得足夠的資訊能確保其actionvalue正確後，再開始進行選擇，此時每次都會選擇actionvalue最大者，所以不用在exploration，epsilon設為0即可，不過這個方法主要就是前面預先花時間exploration，開始選擇後便不再exploration。

Q3:因為本題當中的所有action的reward皆為normal distribution，可以預期只要執行次數夠多，actionvalue最終將會非常接近於它們各自的mean，而隨機選取的機率為epsilon， 最終應會收斂至max(mean)\*(1-epsilon)+(mean的總和/action的數量)\*epsilon，因為有epsilon的機率會選mean最大者，而1-epsilon則隨機選取。