## 問3　★★★　➡解答　p.271

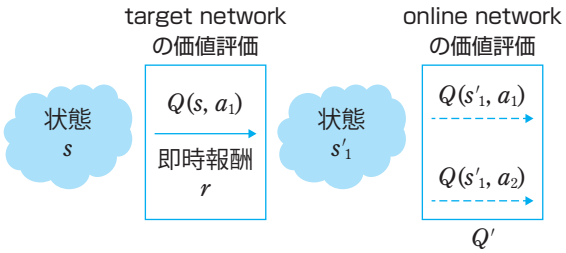
次の文章のうち、Deep Q Networkに関して最も適切な選択肢を選べ。

1. ニューラルネットワークは、入力に状態を表現するベクトルを受け取り、Q関数を近似する。行動選択などの制御はあらかじめ設定した方策によって行う。
2. ニューラルネットワークの入力は状態、出力は報酬となっており報酬関数を近似する。この報酬を最大化するようにパラメータを学習する。
3. ニューラルネットワークは、状態sを入力とし、入力に対する行動aを出力する方策を近似する。
4. ニューラルネットワークを用いたQ関数Q(s,a)の学習には、行動aの結果得られた報酬と次の状態s'において方策によって選択した行動a'に対応するQ関数Q(s',a')を用いる。

## 問4　★★★　➡解答　p.272

DQNを改良したDouble DQN手法について、次の文書を読み、設問に答えよ。

強化学習において、多くの手法はデータのサンプリングを前提として学習が行われる。以下の図は通常のDQNにおいて現状態$s$で行動$a_1$を選択したときに得られた即時報酬$r$と、現状態$s$と次状態$s_1$'の状態行動価値($Q$値)からDQNの損失関数を計算する過程を示している。




$$Loss = (\underbrace{r + \max(Q')}_{\text{教師データ}} - \underbrace{Q(s, a_1)}_{\text{学習データ}})^2$$

※$Q(s, a)$は状態$s$における行動$a$の価値





DQNはニューラルネットワークを用いてQ値を推論することで価値評価を行っており、target networkのパラメータは学習を進めるのにしたがって変化する。ここで通常のDQNではtarget networkとonline networkは同じ重みを利用している。

一方で通常のDQNを改良したDouble DQNと呼ばれる手法は、価値評価に用いる2つのネットワークで違う重みを利用するようにした手法である。

(設問)

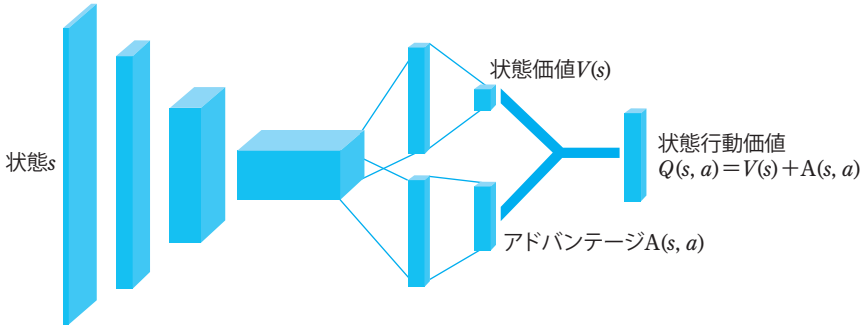
Double DQN手法において、2つの価値評価を異なるネットワークで行うことよるメリットを次の選択肢から1つ選べ。

1. 偏ったQ値の過大評価を改善することができる。
2. 学習中のメモリの使用量を減らせる。
3. 実運用時も2つのネットワークで推論するため高精度な価値評価ができる。
4. データの時間変化が持つ特徴を効率的に学習することができる。

## 問5　★★★　➡解答　p.273

DQNを改良したDueling Network手法について、次の文章を読み、設問に答えよ。

通常のDQNは状態を入力としてQ値を推論するニューラルネットワークである。ここで通常のDQNを改良したDueling Networkと呼ばれる手法では、DQN同様に状態$s$を入力としてQ値を出力するが、以下のようにネットワーク内部でQ値(状態行動価値)を状態価値$V(s)$とアドバンテージ$A(s, a)$に分解している。




ここで状態価値$V(s)$とはその状態にいることがどれだけ良いかを測るもので