機械学習

10/15

1　強化学習

1.1　softmax 法

softmax 法は V(i) ¿ 0 にしか使えないから $e^{(\frac{V(i)}{T})}$ にすることで、絶対に正の値になる。
T の大小で選択確率は変化。T を大きくすると、今回の例だと 1/3 に近く、小さく (T → 0) すると、価値 V の大きい action を主に選択

$$P(2) = \frac{e^{(\frac{20}{T})}}{e^{(\frac{10}{T})} + e^{(\frac{30}{T})} + e^{(\frac{20}{T})}} = \frac{1}{\frac{e^{(\frac{10}{T})}}{e^{(\frac{30}{T})}} + 1 + \frac{e^{(\frac{20}{T})}}{e^{(\frac{30}{T})}}} = 1z \tag{1}$$

1.2　強化学習の基本用語

Environment(環境) ・・・ 人でいう腕、口などの脳以外のもの。ロボットだとモーターとか、もちろん道の起伏とかも含まれる。
state(状態) ・・・ 腕がどの位置
action(行動) ・・・ やりたい行動

1.3　状態、行動、報酬

state ・・・ 最初に→を選べば c ←を選べば a

1.4　強化学習における問題

そう報酬をどうするか

2　Q 学習

行動価値関数 Q

3 11/26

$$\Delta w_i = -\varepsilon \frac{\delta E}{\partial w_i} = -\varepsilon \frac{dE}{dy}\frac{dy}{ds}\frac{\partial s}{\partial w_i} \tag{2}$$