1. Define Loss function $J^i(w) = \frac{1}{2}(y_i - \hat{y}_i)^2$

~~Note that~~ $\hat{y}_i = w_0 \cdot 1 + w_1 x_{i1} + w_2 x_{i2}$

$$\frac{\partial J^i(w)}{\partial w_j} = -(y_i - \hat{y}_i)\frac{\partial \hat{y}_i}{\partial w_j}$$
$$= -(y_i - \hat{y}_i)\, x_{ij}$$

Due to SGD, we need to randomly choose i.

For instance, we choose $i=1$.

$$w_{new} \leftarrow w_{old} - \alpha \frac{\partial J^1(w)}{\partial w}$$

$w_{old} = (0, -0.017, -0.048)$

$\alpha = 0.05$

$$\frac{\partial J^1(w)}{\partial w} = -(y_1 - \hat{y}_1)(x_{10}, x_{11}, x_{12})$$

Since $y_1 = 2$, $\hat{y}_1 = (0, -0.017, -0.048)\begin{pmatrix} 1 \\ 4 \\ 1 \end{pmatrix} = -0.116$

$x_{10} = 1$, $x_{11} = 4$, $x_{12} = 1$

$$\therefore \frac{\partial J^1(w)}{\partial w} = -(2-(-0.116))(1, 4, 1) = (-2.116, -8.464, -2.116)$$

So $w_{new} = (0, -0.017, -0.048) - 0.05(-2.116, -8.464, -2.116)$
$= (0.1058, 0.4062, 0.0578)$

The next estimate of w is $(0.1058, 0.4062, 0.0578)$ if $i=1$

2. (a) States: possible total scores and terminal state $\{0, 2, 3, 4, 5, \text{Done}\}$

Actions: $\{\text{Draw}, \text{Stop}\}$.

(b) Transition Function:

$T(s, \text{Stop}, \text{Done}) = 1$

$$T(s, \text{Draw}, s') = \begin{cases} 1/3 & \text{if } s'-s \in \{2,3,4\} \\ 1/3 & \text{if } s=2 \text{ and } s' = \text{Done} \\ 2/3 & \text{if } s=3 \text{ and } s' = \text{Done} \\ 1 & \text{if } s \in \{4,5\} \text{ and } s' = \text{Done} \\ 0 & \text{otherwise} \end{cases}$$

Reward Function:

$R(s, \text{Stop}, \text{Done}) = s$, $s \leq 5$

$R(s, a, s') = 0$ otherwise

(c) In this case, the optimal policy would be "Draw" if $s \leq 2$; "Stop" otherwise.

The optimal policy is given by taking the argmax instead of max in the final iteration.

V	0	2	3	4	5	Done
V_0	0	0	0	0	0	0
V_1	0	2	3	4	5	0
V_2	3	3	3	4	5	0
V_3	10/3	3	3	4	5	0
Policy Extraction	10/3 Draw	3 Draw	3 stop	4 stop	5 stop	0 stop

3. (a) For (B, East, C, 2), update $V^{\pi}(B)$:

$$V^{\pi}(B) \leftarrow V^{\pi}(B) + \alpha(R(s,a,s') + \gamma V^{\pi}(C) - V^{\pi}(B))$$

$$= 0 + 0.5(2 + 1\times 0 - 0) = 1$$

doing the same computations

Transitions	A	B	C	D	E
initial	0	0	0	0	0
(B, East, C, 2)	0	1	0	0	0
(C, South, E, 4)	0	1	2	0	0
(C, East, A, 6)	0	1	4	0	0
(B, East, C, 2)	0	3.5	4	0	0

So $V(B) = 3.5$ $V(C) = 4$ $V(s) = 0$ $\forall s \in \{A, D, E\}$

(b) $Q(C, South) = 2$ $Q(C, East) = 3$ $Q(B, East) = 2$

$Q(s, a) = 0$ for all other Q-states (s, a)

Transitions	(B, East)	(C, South)	(C, East)
initial	0	0	0
(B, East, C, 2)	1	0	0
(C, South, E, 4)	1	2	0
(C, East, A, 6)	1	2	3
(B, East, C, 2)	3	2	3