# W 6

1. (13 points) MDPs and RL

A new golfer, Mr. Roboto, is playing the Masters tournament. Mr. Roboto's game can be represented as a MDP with the following information:

State Space: { Tee-T, Fairway-F, Sand-S, Green-G }
Actions: { Conservative-C, Risky-R } Shot
Initial State: Tee
Terminal State: Green

With a reward function (* is a wildcard, or don't care):

| $s$ | $R(*,*,s)$ |
|---|---|
| Fairway | 0 |
| Sand | -2 |
| Green | 3 |

Transition Model

| $s$ | $a$ | $s'$ | $T(s,a,s')$ |
|---|---|---|---|
| Tee | Conservative | Fairway | 0.9 |
| Tee | Conservative | Sand | 0.1 |
| Tee | Risky | Green | 0.3 |
| Tee | Risky | Sand | 0.7 |
| Fairway | Conservative | Green | 0.8 |
| Fairway | Conservative | Sand | 0.2 |
| Fairway | Risky | Green | 0.6 |
| Fairway | Risky | Sand | 0.4 |
| Sand | Conservative | Sand | 0.1 |
| Sand | Conservative | Fairway | 0.9 |
| Sand | Risky | Fairway | 0.7 |
| Sand | Risky | Green | 0.3 |

(a) (3 points) Consider the policy of always taking the conservative shot. Assume $\gamma = 1$. Perform two Bellman updates to compute the values of this policy. Use the formula for a value of a state $s$ under a policy $\pi$,

$$V_{i+1}^{\pi}(s) = \sum_{s'} T(s,\pi(s),s') \left[R(s,\pi(s),s') + \gamma V_i(s')\right].$$

| $i$ | $V_i^{\pi_C}(\mathtt{T})$ | $V_i^{\pi_C}(\mathtt{F})$ | $V_i^{\pi_C}(\mathtt{S})$ |
|---|---|---|---|
| 0 | 0 | 0 | 0 |
| 1 | -0.2 | 2 | -0.2 |
| 2 | 1.58 | 1.46 | 1.58 |

(b) (9 points) Assume $\gamma = 1$ and initial estimates are 0. Perform two rounds of Q-value iteration. Also, report the values for each iteration using the calculated Q-values.

| $i$ | $Q_i^*(\mathtt{T},\mathtt{C})$ | $Q_i^*(\mathtt{T},\mathtt{R})$ | $Q_i^*(\mathtt{F},\mathtt{C})$ | $Q_i^*(\mathtt{F},\mathtt{R})$ | $Q_i^*(\mathtt{S},\mathtt{C})$ | $Q_i^*(\mathtt{S},\mathtt{R})$ |
|---|---|---|---|---|---|---|
| 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 1 | -0.2 | -0.5 | 2 | 1 | -0.2 | 0.9 |
| 2 | 1.69 | 0.13 | 2.18 | 1.36 | 1.69 | 2.3 |

| $i$ | $V_i^*(\mathtt{T})$ | $V_i^*(\mathtt{F})$ | $V_i^*(\mathtt{S})$ |
|---|---|---|---|
| 0 | 0 | 0 | 0 |
| 1 | -0.2 | 2 | 0.9 |
| 2 | 1.69 | 2.18 | 2.3 |

(c) (1 point) Report out the optimal policy, using your Q-values from iteration 2 in part (b).

| $\pi_2^*(\mathtt{T})$ | $\pi_2^*(\mathtt{F})$ | $\pi_2^*(\mathtt{S})$ |
|---|---|---|
| C | C | R |

2. (7 points) Naïve Bayes
Consider the following data set of three input variables $A$, $B$, $C$ and a target classification $Z$.

| $A$ | $B$ | $C$ | $Z$ |
|---|---|---|---|
| T | T | T | T |
| T | F | T | F |
| T | F | F | F |
| F | T | T | F |
| F | F | F | T |

(a) (6 points) Suppose we have a new sample with $A = F$, $B = T$ and $C = F$. Using Naïve Bayes, what is the probability assigned to the two target values ($Z = T$ or $Z = F$)? That is, what is the estimate of $P(Z = T \mid A = F, B = T, C = F)$ and $P(Z = F \mid A = F, B = T, C = F)$? Use the maximum likelihood estimate of the probabilities. (Note, the probabilities assessed are simple, therefore please carry out the calculations)
Which classification is preferred?

$$P(Z=T/A=F,B=T,C=F) = \frac{P(Z=T)\cdot P(A=F/Z=T)\cdot P(B=T/Z=T)P(C=F/Z=T)}{P(A=F,B=T,C=F/Z)}$$

$$= \frac{\frac{2}{5}\times\frac{1}{2}\times\frac{1}{2}\times\frac{1}{2}}{\frac{1}{20}+\frac{1}{45}} = \frac{9}{13} \approx 0.6923$$

$$P(Z=F|A=F,B=T,C=F) = \frac{P(Z=F)\cdot P(A=F|Z=F)\cdot P(B=T|Z=F)P(C=F|Z=F)}{P(A=F,B=F,C=F|Z)}$$

$$= \frac{\frac{3}{5}\times\frac{1}{3}\times\frac{1}{3}\times\frac{1}{3}}{\frac{1}{20}+\frac{1}{45}} = \frac{4}{13} \approx 0.3077$$

, $Z=T$ is preferred.

(b) (1 point) How can we improve this Naïve Bayes model?

Use more training data to make the probabilities converge better.

3. (7 points) Perceptrons

(a) (6 points) Consider a binary perceptron that takes as input a bias term plus three features. The features and weight vectors are specified as $f(x) = \langle 1, f(x)_1, f(x)_2, f(x)_3 \rangle$ and $w = \langle w_1, w_2, w_3, w_4 \rangle$.

The perceptron will classify according to the following:

$$y = \begin{cases} +1 & \text{if } w \cdot f(x) \geq 0 \\ -1 & \text{if } w \cdot f(x) < 0 \end{cases}$$

Perform three updates of the Perceptron algorithm, processing the training data.

| | $w_1$ | $w_2$ | $w_3$ | $w_4$ |
|---|---|---|---|---|
| Initial weights | -1 | 1 | 1 | -1 |
| $f(x) = \langle 1, 0, 1, 1 \rangle, y = +1$ | 0 | 1 | 2 | 0 |
| $f(x) = \langle 1, 1, 0, 1 \rangle, y = -1$ | -1 | 0 | 2 | -1 |
| $f(x) = \langle 1, 1, 1, 0 \rangle, y = +1$ | -1 | 0 | 2 | -1 |

(b) (1 point) Look at a set of training samples below, where the first three samples were used in part(a). Will your perceptron classifier able to learn a rule that makes no mistakes on the training data?

| $f(x)_1$ | $f(x)_2$ | $f(x)_3$ | $y$ |
|---|---|---|---|
| 0 | 1 | 1 | +1 |
| 1 | 0 | 1 | -1 |
| 1 | 1 | 0 | +1 |
| 0 | 0 | 0 | +1 |
| 1 | 0 | 0 | +1 |
| 1 | 1 | 1 | -1 |

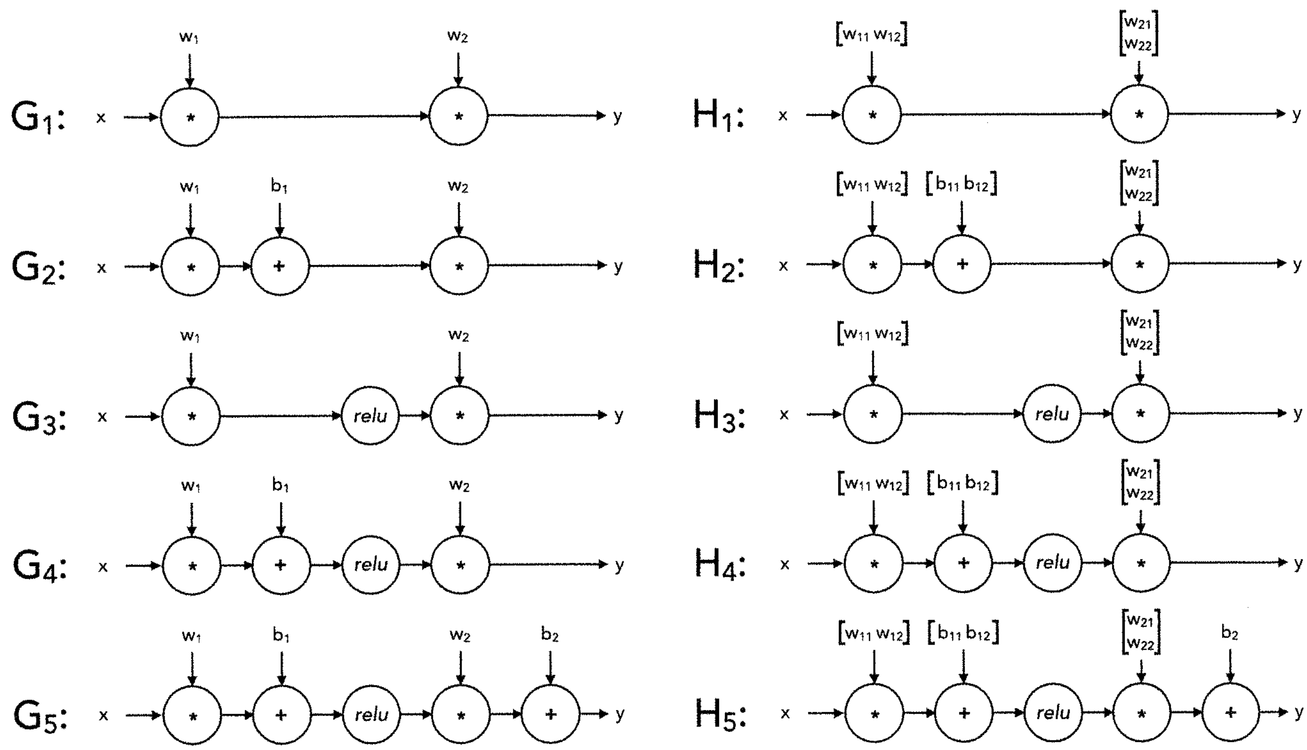
Yes there is a combination of $w_1, w_2, w_3, w_4$ to satisfy:

$$\begin{cases} w_1 + w_3 + w_4 \geq 0 \\ w_1 + w_2 + w_4 < 0 \\ w_1 + w_2 + w_3 \geq 0 \\ w_1 \geq 0 \\ w_1 + w_2 \geq 0 \\ w_1 + w_2 + w_3 + w_4 < 0 \end{cases}$$

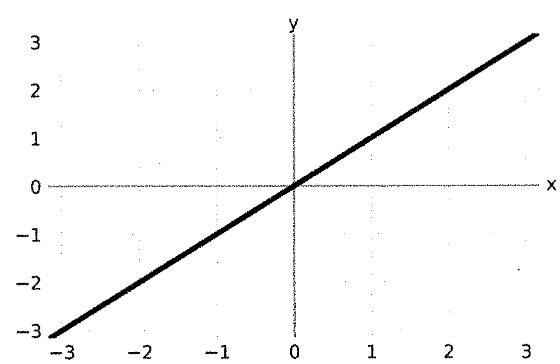
$w_1 = 2, w_2 = -1, w_3 = 1, w_4 = -3$

4. (0 points) Neural Networks: Representation

   For each of the piece-wise linear functions given, list all networks: $G_1, \ldots, G_5$, $H_1, \ldots, H5$, or none, that can represent the function exactly on the range $x \in (-\infty, \infty)$. In the networks, $relu$ denotes the element-wise $ReLU$ nonlinearity: $relu(z) = max(0, z)$. The networks, $G_i$ use 1-dimensional layers, while the networks $H_i$ have some 2 dimensional intermediate layers.

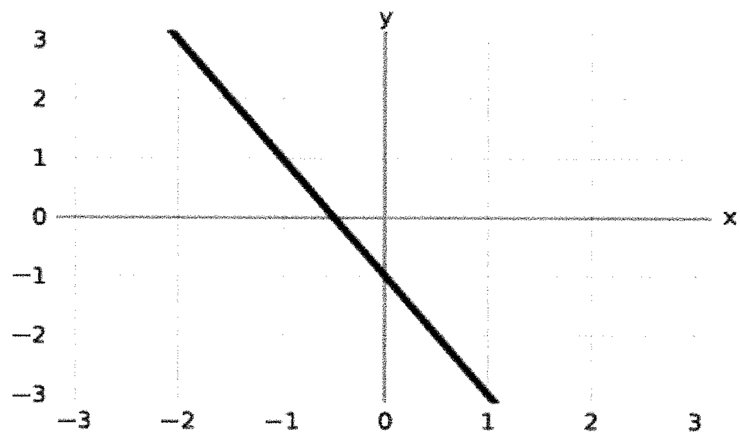



(a) (1 point (bonus)) List all networks or "none" that can represent the function.




G1, G2, H1, H2

(b) (1 point (bonus)) List all networks or "none" that can represent the function.




$G2, H2$

(c) (1 point (bonus)) List all networks or "none" that can represent the function.

$G3, H3,$
$G4, H4,$
$G5, H5$