CARNEGIE MELLON UNIVERSITY 07-280
HOMEWORK 11

DUE: Thursday, April 16, 2026

https://www.cs.cmu.edu/~07280

INSTRUCTIONS

- **Format:** Use the provided LaTeX template to write your answers in the appropriate locations within the *.tex files and then compile a pdf for submission. We try to mark these areas with STUDENT SOLUTION HERE comments. Make sure that you don't change the size or location of any of the answer boxes and that your answers are within the dedicated regions for each question/part. If you do not follow this format, we may deduct points.

  You may also digitally annotate the pdf. Illegible handwriting will lead to lost points. However, we suggest that try to do at least some of your work directly in LaTeX.

- **How to submit written component:** Submit to Gradescope a pdf with your answers. Again, make sure your answer boxes are aligned with the original pdf template.
- **How to submit programming component:** See section Programming Submission for details on how to submit to the Gradescope autograder.
- **Policy:** See the course website for homework policies, including late policy, and academic integrity policies.

Name	
Andrew ID	
Hours to complete all components (nearest hour)	

1 [3 pts] Building GPT2

These written questions are for the corresponding programming homework problems in HW11.

1. **[2 pts]** Q6 Training Loss & Training Perplexity

   In Q6 of the programming HW, you produced 2 plots of the training loss and training perplexity. Insert your plots here:

   Plots

2. **[1 pts]** Q7 Evaluation & Generation

   In Q7 of the programming HW, you experimented with 4 different prompts and 4 different temperature settings. Note that you can run these cells multiple times, and see different outputs every time, because of the nondeterministic nature of LMs like our GPT2. Choose one of the prompts and paste the generation output into the box below.

   Hint: It might be helpful to paste the output as a screenshot, rather than text, since LaTeXwill distort the text.

   Output

2 [14 pts] MDPs: Racing (From Recitation)

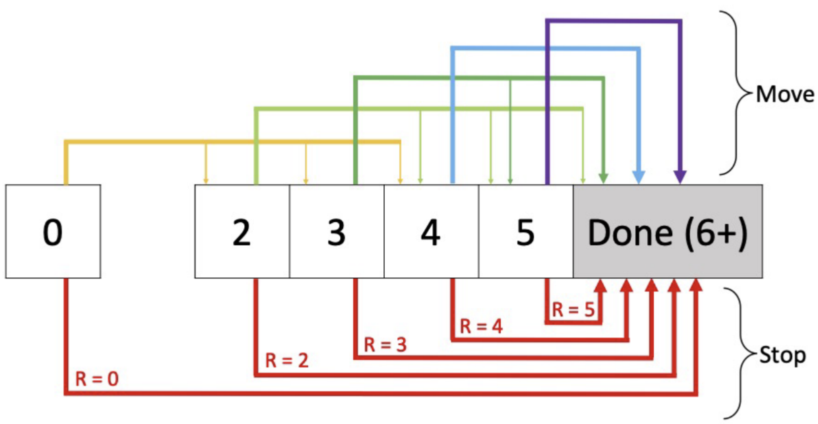
Note that this question was introduced as part of recitation (Recitation Q5), but must be turned in as part of this week's written homework.

Hint: Check your recitation resources!

Consider a modification of the racing robot car example seen in lecture. In this game, the car repeatedly moves a random number of spaces that is equally likely to be 2, 3, or 4. The car can either Move or Stop if the total number of spaces moved is less than 6.

If the total spaces moved is 6 or higher, the game automatically ends, and the car receives a reward of 0. When the car Stops, the reward is equal to the total spaces moved (up to 5), and the game ends. There is no reward for the Move action.

Let's formulate this problem as an MDP with the states $\{0, 2, 3, 4, 5, Done\}$.



1. **[3 pts]** What is the transition function for this MDP? (You should specify discrete values for specific state/action inputs.)

$T(s, Stop, Done) =$	1, for $s \neq Done$
$T(0, Move, s') =$	
$T(2, Move, s') =$	
$T(3, Move, 5) =$	
$T(3, Move, Done) =$	
$T(4, Move, Done) =$	
$T(5, Move, Done) =$	
$T(s, a, s') =$	0 otherwise

2. **[1 pts]** What is the reward function for this MDP?

Reward Function

3. **[3 pts]** Recall the value iteration update equation:

$$V_{k+1}(s) = \max_a \sum_{s'} T(s, a, s')[R(s, a, s') + \gamma V_k(s')]$$

Perform value iteration for 4 iterations with $\gamma = 1$.

States	**0**	**2**	**3**	**4**	**5**	**Done**
V_0						
V_1						
V_2						
V_3						
V_4						

4. **[2 pts]** You should have noticed that value iteration converged above. What is the optimal policy?

States	**0**	**2**	**3**	**4**	**5**
π^*					

5. **[2 pts]** How would our results change with $\gamma = 0.1$?

Answer

6. **[3 pts]** Now imagine we changed the rules so that moving to spaces after 5 does not immediately end the game, but rather wraps back around to the beginning states (this is the case where the states exist side by side in a loop).

   For example if you start in state 5, select the action Move, and move 2 spaces, you would end up in state 1 (this is a new state we would need introduce).

   To be clear, the states would now be states $\{0,1,2,3,4,5, Done\}$.

   How would this modification change our results? (Assume we again use γ=1) Would it add, remove, or have no change on the cycles in our state space graph? Explain.

   ◯ add cycles ◯ remove cycles ◯ no change

   Explain

   Would it change the converged value of the state?

   ◯ Yes ◯ No

   Would it change the optimal policy?

   ◯ Yes ◯ No

3 [5 pts] MDPs: Policy Iteration (From Recitation)

Note that this question was introduced as part of recitation (Recitation Q6), but must be turned in as part of this week's written homework.

Hint: Check your recitation resources!

Recall the racing MDP from the prior problem. In this game, the car repeatedly moves a random number of spaces that is equally likely to be 2, 3, or 4. The car can either Move or Stop if the total number of spaces moved is less than 6.
If the total spaces moved is 6 or higher, the game automatically ends, and the car receives a reward of 0. When the car Stops, the reward is equal to the total spaces moved (up to 5), and the game ends. There is no reward for the Move action.
States: $\{0, 2, 3, 4, 5, Done\}$

Transition	Reward
$T(s, Stop, Done) = 1,\ \ \forall s \neq Done$	$R(s, Stop, Done) = s,\ \ \forall s \leq 5$
$T(0, Move, s') = \frac{1}{3},\ \ \forall s' \in \{2, 3, 4\}$	$R(s, a, s') = 0$ otherwise
$T(2, Move, s') = \frac{1}{3},\ \ \forall s' \in \{4, 5, Done\}$	
$T(3, Move, 5) = \frac{1}{3}$	
$T(3, Move, Done) = \frac{2}{3}$	
$T(4, Move, Done) = 1$	
$T(5, Move, Done) = 1$	
$T(s, a, s') = 0$ otherwise	

Now recall the policy evaluation and policy improvement equations, which together make up policy iteration.

$$\text{Policy Evaluation: } V^{\pi}_{k+1}(s) = \sum_{s'} T(s, \pi(s), s')[R(s, \pi(s), s') + \gamma V^{\pi}_{k}(s')]$$

$$\text{Policy Improvement: } \pi_{new}(s) \leftarrow \text{argmax}_a \sum_{s'} T(s, a, s')[R(s, a, s') + \gamma V^{\pi_{old}}(s')]$$

Perform two iterations of policy iteration for one step of this MDP, starting from the fixed policy below. Use the initial $\gamma = 1$. Note that following the notation above, we are treating π_0 as π_{old} and π_1 as π_{new}.

States	**0**	**2**	**3**	**4**	**5**
π_0	Move	Stop	Move	Stop	Move
$V_0^{\pi_0}$					
$V_1^{\pi_0}$					
$V_2^{\pi_0}$					
$V_3^{\pi_0}$					
π_1					
$V_0^{\pi_1}$					
$V_1^{\pi_1}$					
$V_2^{\pi_1}$					
$V_3^{\pi_1}$					
π_2					

4 [8 pts] Pay Attention to Dimensions! (From Recitation)

Note that this question was introduced as part of recitation (Recitation Q3), but must be turned in as part of this week's written homework.

1. **[3 pts]** As a refresher, remember that $V = XW_V$, $Q = XW_Q$, and $K = XW_K$.

   Let's say our input sequence is $X \in \mathbb{R}^{T \times d_{model}}$ and our W_K, W_Q, W_V matrices are all in $\mathbb{R}^{d_{model} \times d_k}$.

   (a) What would be the dimensions of $S = QK^T$?

   Dimensions of S

   (b) How about $A = g_{sm}(S)$? g_{sm} applies the softmax function.

   Dimensions of A

   (c) In the matrix version of attention, this softmax function is applied row or column wise?

   ○ Row-wise ○ Column-wise

   What does each row in S represent?

   Answer

2. **[4 pts]** Let's say that now $d_{model} = 2$, $T = 2$, and $d_k = 2$.

   (a) What would be the dimensions of $S = QK^T$?

   Dimensions of S

   (b) How about $A = g_{sm}(S)$? g_{sm} applies the softmax function.

   Dimensions of A

(c) How about the W_K, W_Q, W_V matrices?

Dimensions

(d) How about K, Q, V?

Dimensions

3. **[1 pts]** If $T = 3$, which of the following would change dimensions (select multiple).

- ☐ W_K ☐ W_V ☐ W_Q
- ☐ K ☐ V ☐ Q
- ☐ S ☐ A ☐ X

5 Collaboration Policy

After you have completed all other components of this assignment, report your answers to the following collaboration questions.

1. Did you receive any help whatsoever from anyone in solving this assignment? If so, include full details including names of people who helped you and the exact nature of help you received.

Your Answer

2. Did you give any help whatsoever to anyone in solving this assignment? If so, include full details including names of people you helped and the exact nature of help you offered.

Your Answer

3. Did you find or come across code that implements any part of this assignment? If so, include full details including the source of the code and how you used it in the assignment.

Your Answer