THE HONG KONG UNIVERSITY OF SCIENCE & TECHNOLOGY

Machine Learning

Homework 4

Due Date: See course website

Your answers should be typed, not handwritten. You can submit a Word file or a pdf file. Submissions are to be made via Canvas. Note that penalty applies if your similarity score exceeds 40. To minimize your similarity score, don't copy the questions.

Question 1: Here is the objective function of VAE:

$$\mathbf{L}(\mathbf{x}^{(i)}, \theta, \phi) = E_{\mathbf{z} \sim q_\phi(\mathbf{z}|\mathbf{x}^{(i)})}\left[\log p_\theta(\mathbf{x}^{(i)}|\mathbf{z})\right] - KL[q_\phi(\mathbf{z}|\mathbf{x}^{(i)})||p_\theta(\mathbf{z})]$$

Explain why the first term is called the reconstruction error.

Question 2: In class, we discussed two loss functions for the generator G in GAN:

$$\begin{aligned} L_1 &= \frac{1}{m}\sum_{i=1}^{m}[-\log D(G(z^i))] \\ L_2 &= \frac{1}{m}\sum_{i=1}^{m}\log[1 - D(G(z^i))] \end{aligned}$$

(a) What is the common goal that both loss functions aim to achieve?

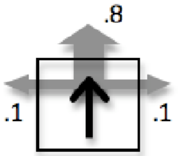
(b) What are the advantages and disadvantages of each loss function? Why? How are the disadvantages mitigated in practice?

(c) Assume the discriminator D is optimal. Which loss function is an approximation of the Jensen-Shannon divergence between the data distribution and the generator distribution?

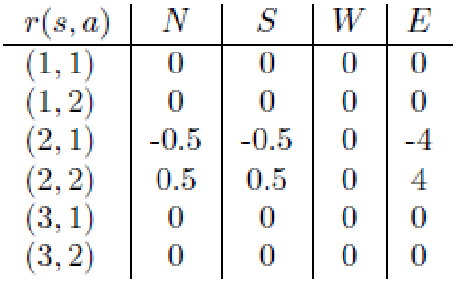
Question 3: What are the inputs and output of the noise predictor in denoising diffusion probabilistic models (DDPM)? What is the loss function used to train it?

Question 4: Consider an agent that acts in the gridworld shown below. The agent always starts in state $(1, 1)$, marked with the letter S. There are two terminal goal states, $(3, 2)$ with reward $+5$ and $(3, 1)$ with reward -5. Rewards are 0 in non-terminal states. (The reward for a state is received as the agent moves into the state.) The transition function is such that the intended agent movement (North, South, West, or East) happens with probability 0.8. With probability 0.1 each, the agent ends up in one of the states perpendicular to the intended direction. If a collision with a wall happens, the agent stays in the same state.

		+5
S		-5



The expected immediate reward function $r(s, a) = \sum_{s'} r(s, a, s')P(s'|s, a)$ is as follows:



$r(s,a)$	N	S	W	E
$(1,1)$	0	0	0	0
$(1,2)$	0	0	0	0
$(2,1)$	-0.5	-0.5	0	-4
$(2,2)$	0.5	0.5	0	4
$(3,1)$	0	0	0	0
$(3,2)$	0	0	0	0

(a) Assume the initial value function $Q_0(s,a) = 0$ for all states s and actions a. Let $\gamma = 0.9$. The Q-function Q_1 after the first value iteration is the same as $r(s,a)$. What is the Q-function Q_2 after the second value iteration? What is the greedy policy π_2 based on Q_2. In case of ties, list all tied actions.

(b) Let $Q(s,a) = 0$ for all s and a initially. Using the Q-learning rule (with $\alpha = 0.1$ and $\gamma = 0.9$), update the Q-function by considering the following experience tuples one by one and in the order presented. (There is no need to consider how the tuples are obtained.) Show the function after each update.

s	a	r	s'
(2, 2)	E	5	(3, 2)
(2, 1)	N	0	(2, 2)
(1, 2)	E	0	(2, 2)
(1, 1)	N	0	(1, 2)

Give the greedy policy based on the latest Q function. In case of ties, list all tied actions.

Question 5: Here is the parameter update rule for Deep Q-Networks:

$$\theta \leftarrow \theta - \alpha \nabla_\theta ([r(s,a) + \gamma \max_{a'} Q(s', a'; \theta^-)] - Q(s, a; \theta))^2$$

What do s, a, $r(s,a)$ and s' stand for? What about θ^-? What is the objective that the update rule is intended to achieve?

Question 6: Here is the update rule for the actor in the Actor-Critic algorithm:

$$\theta \leftarrow \theta + \alpha \nabla_\theta \log \pi_\theta(a|s) \hat{A}^\pi(s,a),$$

where $\hat{A}^\pi(s,a) \leftarrow r + \gamma \hat{V}^\pi_\phi(s') - \hat{V}^\pi_\phi(s)$.

What do s, a, $r(s,a)$ and s' stand for? What about $\hat{V}^\pi_\phi(s)$? Intuitively, what does the update rule try to achieve?

The following two questions are for self-practice

Question 7: (a) In the context of deep image classification, what is adversarial attack?

(b) The CW attack finds an adversarial example $\mathbf{x}'$ for a benign example $\mathbf{x}$ by solving the following optimization problem:

$$\min_{\mathbf{x}'} c||\mathbf{x} - \mathbf{x}'||_2^2 + l(\mathbf{x}')$$
$$\text{s.t. } \mathbf{x}' \in [0,1]^n$$

where $l(\mathbf{x}') = \max\{\max_{i \neq t} Z_i(\mathbf{x}') - Z_t(\mathbf{x}'), -\kappa\}$.

What do the terms $Z_t(\mathbf{x}')$, $Z_i(\mathbf{x}')$ stand for? What are the key ideas behind this attack?

Question 8: In some pixel-level explanations, we need to compute $\frac{\partial z_c(\mathbf{x})}{\partial x_i}$. The backpropagation algorithm for the task is given in Lecture 16. Suppose the last layer of the model is a Softamx layer. How would you change to algorithm if we are to compute $\frac{\partial \log P(y=c|\mathbf{x})}{\partial x_i}$?