

Exercise Session 11

World Models and Generative World Modeling

Prepared by Liangze Jiang and Eeshaan Jain

Overview

Task 1. Multiple Choice Questions on World Modeling

1

Task 1. Multiple Choice Questions on World Modeling.

This problem set contains 8 multiple-choice questions. The purpose is to test your conceptual understanding of world models and provide a light warm-up for the final exam. Please also expect to see many other question formats in the final exam, similar to those you have seen in other exams.

For each question, select the **single** best answer from the options provided.

1. In the **World Models** paper, the architecture is divided into three components. Which is *not* one of them?
 - (a) V (Vision model)
 - (b) M (Memory/RNN model)
 - (c) C (Controller)
 - (d) P (Planning module)
2. The Vision (V) component in World Models *cannot* be implemented as:
 - (a) A Variational Autoencoder (VAE)
 - (b) A standard Autoencoder without variational inference
 - (c) A Generative Adversarial Network (GAN) encoder
 - (d) A supervised classifier trained on object labels
3. The key innovation of JEPA compared to generative models is:
 - (a) Predicting in representation space rather than pixel space
 - (b) Using a stop-gradient mechanism to prevent collapse
 - (c) Eliminating the need for a decoder
 - (d) Does not require labels for the training examples
4. In JEPA, what does the “joint-embedding” refer to?
 - (a) Multiple modalities embedded in the same space
 - (b) Both context and target are embedded in a shared representation space

- (c) Using the same encoder for multiple downstream tasks
- (d) Joint training of encoder and decoder

5. Which problem does JEPA explicitly aim to avoid?

- (a) Supervised learning with labels
- (b) Collapse to trivial solutions
- (c) Predicting irrelevant details in high-dimensional spaces
- (d) Using contrastive learning with negative samples

6. In implicit world models, the decoder component (if present) is used for:

- (a) Decision-making and planning
- (b) Generating synthetic training data
- (c) Auxiliary training losses or visualization
- (d) Computing the policy gradient

7. In explicit world models, simulated trajectories are:

- (a) Always in latent space only
- (b) In the same modality as perception
- (c) Only used for visualization purposes
- (d) Computed without using action information

8. Which statement about world model representations is *true*?

- (a) Explicit world models never use latent representations
 - (b) Explicit world models reconstruct future observations during imagined rollouts
 - (c) Implicit world models produce latent states that must be lower-dimensional than the observation space.
 - (d) Implicit world models always use lower-dimensional representations than explicit models
-