10-423/10-623 Gen AI	**Name:**
Spring 2025	**Andrew ID:**
Quiz 2	**Room:**
02/17/24	**Seat:**
Time Limit: 15 minutes	**Exam Number:**

Instructions:

- Verify your name and Andrew ID above.
- This exam contains **??** pages (including this cover page).
  The total number of points is **??**.
- Clearly mark your answers in the allocated space. If you have made a mistake, cross out the invalid parts of your solution, and circle the ones which should be graded.
- Look over the exam first to make sure that none of the **??** pages are missing.
- No electronic devices may be used during the exam.
- Please write all answers in pen or *darkly* in pencil.
- You have 15 minutes to complete the exam. Good luck!

Run LaTeX again to produce the table

1 Deep Models for Vision (?? points)

1.1. (1 point) **True or False:** As we go deeper into a CNN, the weights of *later* convolutional layers learn to detect features of *larger* patches of the input images.

- ◯ True
- ◯ False

1.2. (1 point) **Select one:** Consider a 2D convolution layer with input image size $M \times M$ with C_{in} channels. Let N_W be the original number of weight parameters. If we double the input image width and height to be $2M \times 2M$ and change nothing else about the layer, what is total number of weights in this layer?

- ◯ $N_W/2$
- ◯ $N_W/4$
- ◯ $\sqrt{2}N_W$
- ◯ $2N_W$
- ◯ $4N_W$
- ◯ None of the above

1.3. (1 point) **Select all that apply:** Which aspects of an encoder-only Transformer model need to be substantially changed to convert it to a basic Vision Transformer (ViT) model?

Select as few options as necessary.

- ☐ Tokenization
- ☐ Position embedding
- ☐ Attention blocks
- ☐ Transformer blocks
- ☐ Optimization algorithm
- ☐ None of the above

2 GANs (?? points)

2.1. (1 point) **True or False:** The discriminator's role in a GAN is to determine whether the noise vector was obtained by adding noise to a real image or by sampling noise from the generator model.

◯ True

◯ False

2.2. (2 points) **Select all that apply:** GANs learn by trying to find a θ and ϕ that optimize a minimax problem for a generator G_θ and a discriminator D_ϕ:

$$\min_\theta \max_\phi J(\theta, \phi), \qquad \text{where } J(\theta, \phi) = \log\left(D_\phi(\mathbf{x}^{(i)})\right) + \log\left(1 - D_\phi(G_\theta(\mathbf{z}^{(i)}))\right),$$

$\mathbf{x}^{(i)}$ is a random training image, and $\mathbf{z}^{(i)}$ is a random noise vector. Which of the following techniques could be used to optimize this learning problem?

□ Alternate between a step in the direction of $\nabla_\phi J(\theta, \phi)$ and a step opposite the gradient of $\nabla_\theta J(\theta, \phi)$.

□ Alternate between a step opposite the direction of $\nabla_\phi J(\theta, \phi)$ and a step in the gradient of $\nabla_\theta J(\theta, \phi)$.

□ Jointly step in the direction $(\nabla_\phi J(\theta, \phi), -\nabla_\theta J(\theta, \phi))$

□ Jointly step in the direction $(-\nabla_\phi J(\theta, \phi), \nabla_\theta J(\theta, \phi))$

□ None of the above

2.3. (1 point) **Select one:** Which of the following best describes how an image is generated from a trained GAN?

◯ A neural network creates the mean and covariance parameters of a Gaussian, and an image is sampled from that Gaussian.

◯ Gaussian noise is repeatedly subtracted away from a randomly sampled noise vector until an image is left remaining.

◯ A noise vector is sampled from a Gaussian, then a deterministic neural network transforms the noise vector into an image.

◯ A noise vector is constructed by a neural network and then an image is sampled from a nonlinear distribution that conditions on that noise vector.

3 Diffusion Models (?? points)

3.1. (1 point) **True or False:** To train a diffusion model, we find the parameters for the reverse process model that maximize the sum of the log-likelihoods of the training images, i.e. $\hat{\boldsymbol{\theta}} = \arg\max \sum_{i=1}^{N} \log p_{\boldsymbol{\theta}}(\mathbf{x}^{(i)})$ where $p_{\boldsymbol{\theta}}$ is the reverse process, and $\mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(N)}$ are the N training images.

◯ True

◯ False

3.2. (1 point) **True or False:** The forward process of a diffusion model *and* the (learned) reverse process are both stochastic.

◯ True

◯ False

3.3. (2 points) **Select all that apply:** Why does the Denoising Diffusion Probabilistic Model (DDPM) use a UNet model? Recall that the structure of the exact reverse process $q(\mathbf{x}_{t-1} \mid \mathbf{x}_t, \mathbf{x}_0)$ is a Gaussian of the form $\mathcal{N}(\tilde{\mu}_q(\mathbf{x}_t, \mathbf{x}_0), \sigma_t^2 \mathbf{I})$.

□ Because a UNet is a parameter efficient encoder-only Transformer model.

□ Because the inputs and outputs of a UNet can be of the same dimension.

□ In order to approximate $\tilde{\mu}_q(\mathbf{x}_t, \mathbf{x}_0)$ through various parameterizations.

□ In order to approximate $\sigma_t^2 \mathbf{I}$ through various parameterizations.

□ None of the above

4 VAEs (?? points)

4.1. (2 points) **Select all that apply.** Which of the following would we like to *minimize* when training a variational autoencoder, where $q_\phi(\mathbf{z} \mid \mathbf{x})$ is the encoder, $p_{\boldsymbol{\theta}}(\mathbf{x} \mid \mathbf{z})$ is the decoder, $\mathbf{z}^{(i)} \sim q_\phi(\mathbf{z} \mid \mathbf{x}^{(i)})$, and $\hat{\mathbf{x}}^{(i)} \sim p_\theta(\mathbf{x} \mid \mathbf{z}^{(i)})$.

- ☐ $\frac{1}{N}\sum_{i=1}^{N} \|\mathbf{x}^{(i)} - \hat{\mathbf{x}}^{(i)}\|_2^2 - KL\left(q_{\boldsymbol{\phi}}(\mathbf{z} \mid \mathbf{x}) \mid\mid \mathcal{N}(\mathbf{0}, \mathbf{I})\right)$
- ☐ $\mathbb{E}_{\mathbf{z} \sim q_{\boldsymbol{\phi}}(\mathbf{z}|\mathbf{x})}\left[-\log p_{\boldsymbol{\theta}}(\mathbf{x} \mid \mathbf{z})\right] - KL\left(q_{\boldsymbol{\phi}}(\mathbf{z} \mid \mathbf{x}) \mid\mid p_{\boldsymbol{\theta}}(\mathbf{z})\right)$
- ☐ $-\text{ELBO}(q_\phi)$
- ☐ None of the above

4.2. (1 point) **True or False:** The reparameterization trick is used to avoid having a random function on the computation path between the generator network weights and the objective.

- ◯ True
- ◯ False