Quiz: Variational Autoencoders

STA414/2104 - Winter 2026

1. Which of the following describes a primary issue with standard, deterministic autoencoders when evaluating them as generative models?

   (a) They naturally suffer from posterior collapse where the latent variable is completely ignored.

   (b) The encoder outputs a probability distribution rather than a fixed code, making direct reconstruction impossible.

   (c) Proximity in the data space does not directly enforce proximity in the feature space, leading to unrealistic outputs when sampling from gaps between representations.

   (d) They require an annealing schedule to gradually introduce the KL divergence term during optimization.

   Correct Answer: (c)
   Rationale: In deterministic autoencoders, proximity in feature space is not directly enforced for inputs in close proximity in data space. This means the latent space may not be continuous or allow easy interpolation. If the space has discontinuities (e.g., gaps between clusters) and you sample a variation from there, the decoder will simply generate an unrealistic output.

2. What is the primary computational limitation that necessitates the use of an approximate posterior in Variational Autoencoders (VAEs)?

   (a) Backpropagating through the stochastic sampling layer requires an exhaustive search without the reparametrization trick.

   (b) Computing the exact marginal likelihood of a new data point involves an intractable integral over the hidden variables.

   (c) The model is forced to rely on Markov Chain Monte Carlo (MCMC) during the forward generation pass, which scales poorly.

   (d) The dimensionality reduction process of the encoder increases the computational complexity of evaluating the prior $p(z)$ to an exponential time bound.

   Correct Answer: (b)
   Rationale: In VAEs, evaluating the exact likelihood $p(x)$ requires computing the integral $\int p(x|z)p(z)dz$ over the latent variable z. For non-linear deep neural networks, this integration is intractable and highly inefficient to approximate with methods like MCMC. This necessitates the introduction of an approximate distribution $q_\phi(z|x)$ to optimize a tractable lower bound instead.

3. In the optimization objective (loss function) of a Variational Autoencoder, what is the role of the Kullback-Leibler (KL) divergence term?

   (a) It calculates the exact likelihood of the new data point by integrating over the hidden variable z.

   (b) It manages the reconstruction error by measuring the geometric distance between the decoder's output and the input image.

   (c) It forces the neural network to output exactly the same deterministic code for identical inputs.

   (d) It regularizes the encoder to stay close to the prior distribution, preventing the network from simply memorizing the training data.

   Correct Answer: (d)
   Rationale: The VAE maximization objective (ELBO) contains two main parts. While the first term handles the reconstruction loss, the second term (KL divergence) regularizes the encoder not to be too far from the prior. This ensures that the network cannot simply memorize the data.

4. When designing a VAE for complex high-dimensional data like facial images, what is the most likely visible effect of this choice on your generated samples?

   (a) The samples will exhibit perfect disentanglement of independent real-world factors like head pose or hair color.

   (b) The generated samples will typically appear unrealistic, often manifesting as overly smooth predicted means combined with overly noisy final outputs.

   (c) The model will perfectly memorize the training data, leading to exact reconstructions but zero variation in the latent space.

   (d) The samples will be overly smooth.

   Correct Answer: (b)
   Rationale: Samples generated from VAEs are generally not perfect. These often leads to predicted means that are overly smooth, and final samples that appear overly noisy when the noise is added.

5. Consider a modified Variational Autoencoder objective where the KL divergence term (in the ELBO) is multiplied by a constant weight β. If we set a constant $0 < \beta < 1$, what is the primary consequence on the learned model compared to a standard VAE ($\beta = 1$)?

   (a) It enforces stricter regularization on the latent space, resulting in highly disentangled and uncorrelated latent features.

   (b) It prioritizes the reconstruction loss over adherence to the prior, allowing the model to achieve higher-fidelity reconstructions at the cost of a less continuous latent space for generation.

   (c) It intentionally induces posterior collapse by forcing the encoder to simply predict the prior distribution.

   (d) It ensures that exact likelihood evaluation becomes computationally tractable.

   Correct Answer: (b)
   Rationale: The VAE objective balances reconstruction loss and a KL divergence regularization term. By setting a constant $\beta < 1$, the penalty for the approximate posterior diverging from the prior is weakened. This allows the network to prioritize encoding specific information needed for precise input reconstruction, leading to sharper outputs. However, this comes at the expense of a well-behaved, continuous latent space, making the model less suitable for randomly generating realistic new samples from the prior.