Week-6 Questions

Prof . Prathosh AP and Chandan J

May 2025

1. Which of the following is the correct formula for the KL divergence between the encoder distribution $q(z|x) = \mathcal{N}(\mu, \sigma^2)$ and the prior $p(z) = \mathcal{N}(0, 1)$ in a VAE?

   (A) $\frac{1}{2}(\sigma^2 + \mu^2 - 1)$
   (B) $\frac{1}{2}(\mu^2 + \sigma^2 - \log \sigma^2 - 1)$
   (C) $\mu^2 + \sigma$
   (D) $\log(\sigma^2) + \mu$

   Answer: (B)

2. Increasing the β value in β-VAE:

   (A) Decreases the importance of KL divergence
   (B) Encourages latent variables to deviate from the prior
   (C) Encourages disentanglement by penalizing latent capacity
   (D) Has no effect on the latent space

   Answer: (C)

3. Assuming a latent space of dimension 2 in a VAE trained on MNIST, which of the following is true?

   (A) KL divergence will be zero
   (B) Decoder may struggle to reconstruct due to low capacity
   (C) Training loss will vanish
   (D) The posterior becomes non-Gaussian

   Answer: (B)

4. The KL divergence term in a VAE becomes very small (close to zero). This indicates:

   (A) Encoder overfits
   (B) Decoder collapsed

(C) β too large

(D) Latent Posterior matches the prior

Answer: (D)

(A)

```
std = torch.exp(logvar / 2)
eps = torch.randn_like(std)
z = mu + eps * std
```

(B)

```
z = torch.randn(mu.size()) * logvar + mu
```

(C)

```
z = mu + torch.randn_like(mu) + logvar
```

(D)

```
z = mu * torch.randn_like(logvar)
```

Answer: (A)

5. Which of the following decoder blocks is appropriate for a VAE trained on MNIST using binary cross-entropy loss?

(A)

```
nn.Sequential(
    nn.Linear(20, 400),
    nn.ReLU(),
    nn.Linear(400, 784),
    nn.Tanh()
)
```

(B)

```
nn.Sequential(
    nn.Linear(20, 400),
    nn.ReLU(),
    nn.Linear(400, 784),
    nn.Sigmoid()
)
```

(C)

```
nn.Sequential(
    nn.Linear(20, 400),
    nn.ReLU(),
    nn.Linear(400, 784),
    nn.ReLU()
)
```

(D)
```
nn.Sequential(
    nn.Linear(20, 784)
)
```

Answer: (B)

6. Which of the following code snippets correctly flattens an encoder output $z \in \mathbb{R}^{[B,D,H,W]}$ to compute distances to codebook vectors?

(A)
```
flat_z = z.view(-1, z.size(1))
```

(B)
```
flat_z = z.permute(0, 2, 3, 1).reshape(-1, z.size(1))
```

(C)
```
flat_z = z.squeeze()
```

(D)
```
flat_z = z.reshape(-1)
```

Answer: (B)

7. Which code samples 16 latent vectors from a trained codebook of size 512 with embedding dim 64?

(A)
```
        indices = torch.randint(0, 512, (16,))
        z_q = model.codebook(indices)
```

(B)
```
        z_q = torch.randn(16, 64)
```

(C)
```
        z_q = model.encoder(torch.randn(16, 3, 32, 32))
```

(D)
```
        z_q = torch.ones(16, 64)
```

Answer: (A)

8. Theoretically it is not possible to use a deep neural network as an encoder and reduce the dimension of latent space to one. Is the this statement true??

(A) Yes

(B) No

Answer: (A)

9. VAE is

   (A) An optimization problem where the optimal encoder and decoder are found by finding the 'arg min' of the reconstruction error between data and encoded - decoded data

   (B) An optimization problem where the optimal encoder and decoder are found by finding the 'arg min' of the reconstruction error between data and decoded- encoded data

   (C) An optimization problem where the optimal encoder and decoder are found by finding the 'arg max' of the reconstruction error between data and encoded - decoded data

   (D) An optimization problem where the optimal encoder and decoder are found by finding the 'arg max' of the reconstruction error between data and decoded - encoded data

   Answer: (B)

10. The basic idea in a VQ VAE is

    (A) Use a discrete latent space with a fixed finite size dictionary and replace the output of encoder with the element with greatest distance in the dictionary

    (B) Use a discrete latent space with a fixed finite size dictionary and replace the output of encoder with the element with least distance in the dictionary

    (C) Use a discrete latent space with a learnable finite size dictionary and replace the output of encoder with the element with greatest distance in the dictionary

    (D) Use a discrete latent space with a learnable finite size dictionary and replace the output of encoder with the element with least distance in the dictionary

    Answer: (D)