Deep Learning

Assignment- Week 12

TYPE OF QUESTION: MCQ/MSQ

Number of questions: 10 **Total mark: 10 X 1 = 10**

QUESTION 1:

During training a Variational Auto-encoder (VAE), it is assumed that $P(z|x) \sim N(0, I)$ i.e., given an input sample, the encoder is forced to map its latent code to $N(0, I)$. After the training is over, we want to use the VAE as a generative model. What should be the best choice of distribution from which we should sample a latent vector to generate a novel example?

a. $N(0, I)$: Normal distribution with zero mean and identity covariance
b. $N(1, I)$: Normal distribution with mean = 1 and identity covariance
c. Uniform distribution between [-1, 1]
d. $N(-1, I)$: Normal distribution with mean = -1 and identity covariance

Correct Answer: a

Detailed Solution:

Since during training, we forced the latent code to follow $N(0, I)$, the decoder has learnt to map latent codes from that distribution only. So, during sampling if we provide vectors from any other distributions, then the encoder will have low probability to have encountered such vectors thereby leading to unrealistic reconstructions. So, we should sample vectors from $N(0, I)$ for using the pre-trained VAE as a generative model.

QUESTION 2:

When the GAN game has converged to its Nash equilibrium (when the Discriminator randomly makes an error in distinguishing fake samples from real samples), what is the probability (of belongingness to real class) given by the Discriminator to a fake generated sample?

a. 1
b. 0.5
c. 0
d. 0.25

Correct Answer: b

Detailed Solution:

For G fixed, the optimal discriminator D is

$$D_G^*(\boldsymbol{x}) = \frac{p_{data}(\boldsymbol{x})}{p_{data}(\boldsymbol{x}) + p_g(\boldsymbol{x})}$$

Proof. The training criterion for the discriminator D, given any generator G, is to maximize the quantity $V(G, D)$

$$V(G, D) = \int_{\boldsymbol{x}} p_{\text{data}}(\boldsymbol{x}) \log(D(\boldsymbol{x}))dx + \int_{z} p_{\boldsymbol{z}}(\boldsymbol{z}) \log(1 - D(g(\boldsymbol{z})))dz$$
$$= \int_{\boldsymbol{x}} p_{\text{data}}(\boldsymbol{x}) \log(D(\boldsymbol{x})) + p_g(\boldsymbol{x}) \log(1 - D(\boldsymbol{x}))dx$$

For any $(a, b) \in \mathbb{R}^2 \setminus \{0, 0\}$, the function $y \rightarrow a\log(y) + b\log(1 - y)$ achieves its maximum in $[0, 1]$ at $\frac{a}{a+b}$. The discriminator does not need to be defined outside of $Supp(p_{\text{data}}) \cup Supp(p_g)$

Nash equilibrium is reached when the generated distribution, $p_g(x)$ equals the original data distribution, $p_{data}(x)$, which leads to $D(x) = 0.5$ for all x.

QUESTION 3:

Why is re-parameterization trick used in VAE?

a. Without re-parameterization, the mean vector of latent code of VAE encoder with tend towards zero
b. Sampling from a VAE encoder latent space is non-differentiable and thus we cannot back propagate gradient during optimization using gradient descent
c. We need to re-parameterize Normal distribution over latent space to Bernoulli distribution
d. None of the above

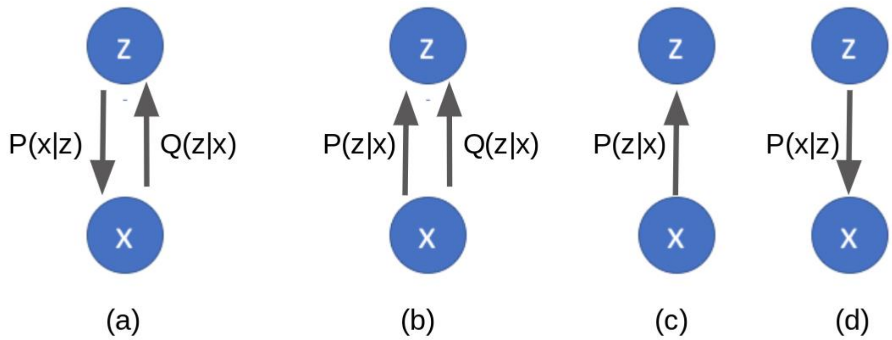
Correct Answer: b

Detailed Solution:

We cannot sample in a differentiable manner from within a computational graph present in a neural network. Re-parameterization enables the sampling function to be present outside the main computational graph which enables us to do regular gradient descent optimization.

QUESTION 4:

Which one of the following graphical models fully represents a Variational Auto-encoder (VAE) realization?



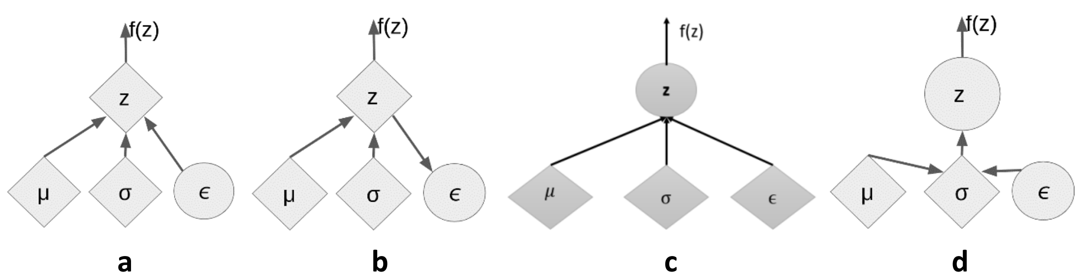
Correct Answer: a

Detailed Explanation:

For practical realization of VAE, we have an encoder $Q(\cdot)$ which receives an input signal, x and generates a latent code, z. This part of the network can be denoted by $Q(z|x)$ and directed from x to z. Next, we have a decoder section which takes the encoded z vector to reconstruct the input signal, x. This part of the network is represented by $P(x|z)$ and should be directed from z to x.

QUESTION 5:

Which one of the following computational graphs correctly depict the re-parameterization trick deployed for practical Variational Auto-encoder (VAE) implementation? Circular nodes represent random nodes in the models and the quadrilateral nodes represent deterministic nodes.



Correct Answer: a

Detailed Solution:

With the re-parameterization trick, the only random component in the network is the node of $\in$ which is sampled from $N(0, I)$. The other nodes of μ and σ are deterministic. Since $\in$ is sampled from outside the computational graph, the overall z vector also becomes deterministic component for a given set of μ, σ and $\in$. Also, if z is not deterministic, we cannot back propagate gradients through it. Also, in the computation graph, the forward arrows will emerge from μ,σ,ϵ towards z for computing the z vector.

QUESTION 6:

For the following min-max game, at which state of (x, y) do we achieve the Nash equilibrium (the state where change of one variable does not alter the state of the other variable)?

$$\min_{x} \max_{y} V(x,y) = xy$$

a. X = 0 y = -1
b. X = 0 , y= 0
c. X = 0 y = 1
d. X = ∞(infinite), y = 0

Correct Answer: b

Detailed Solution:

The Nash equilibrium is x=y=0. This is the only state where the action of one player does not affect the other player's move. It is the only state that any opponents' actions will not change the game outcome.

QUESTION 7:

Which of the following losses can be used to optimize for generator's objective (while training a Generative Adversarial network) by MINIMIZING with gradient descent optimizer? Consider cross-entropy loss,

$$CE(a, b) = - [a*\log(b) + (1-a)*\log(1-b)]$$

and D(G(z)) = probability of belonging to real class as output by the Discriminator for a given generated sample G(z).

a. CE(1, D(G(z)))
b. CE(1, -D(G(z)))

c. CE(1, 1 - D(G(z)))
d. CE(1, 1 / D(G(z)))

Correct Answer: a

Detailed Solution:

Except for option (a) none of the other objective function are minimized at D(G(z)) = 1 which is the goal of the Generator, i.e. to force the Discriminator to output probability=1 for a generated sample. Loss function in option (a) is the only choice which keeps on decreasing as D(G(z)) increases. Also, it is required that D(G(z)) ∈ [0,1].

QUESTION 8:

While training a Generative Adversarial network, which of the following losses **CANNOT** be used to optimize for discriminator objective (while only sampling from the distribution of generated samples) by **MAXIMIZING** with gradient ASCENT optimizer? Consider cross-entropy loss,

$$CE(a, b) = - [\ a*\log(b) + (1-a)*\log(1-b)]$$

and D(G(z)) = probability of belonging to real class as output by the Discriminator for a given generated sample, G(z) from a noise vector, z.

a. CE(1, D(G(z)))
b. -CE(1, D(G(z)))
c. CE(1, 1 + D(G(z)))
d. -CE(1, 1 - D(G(z)))

Correct Answer: b

Detailed Solution:

During optimization of discriminator, when we sample from the distribution of fake/generated distribution, we want D(G(z)) = 0. Since we want to use gradient ASCENT optimization, the objective function should increase as we approach D(G(z)) = 0 while the objective value should decrease with increase in value of D(G(z)). Apart from option (b), all other options satisfy the above conditions.

QUESTION 9:

For training VAE, we want to predict an unknown distribution of latent code given an observed sample, i.e., P(z|x), but we approximate it with some distribution Q(z|x) which we can control by varying some known parameters. Which of the following loss functions is used as a loss to minimize?

a. $-\sum_z Q(z|x) log \frac{P(x,z)}{Q(z \vee x)}$
b. $-\sum_x Q(z|x) log \frac{P(x,z)}{Q(z \vee x)}$
c. $\sum_z P(z|x) log \frac{P(x,z)}{Q(z \vee x)}$
d. None of the above

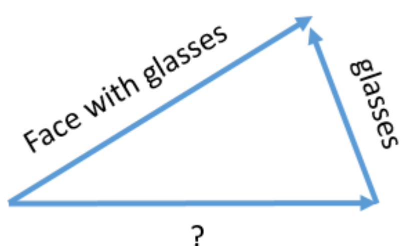
Correct Answer: a

Detailed Solution:

Since we are trying to approximate P(z|x) with Q(z|x), we will try to minimize the KL divergence, KL(Q(z|x) || P(z|x)) which eventually leads to maximization of the well-known Variational lower bound of $\sum_z Q(z|x) log \frac{P(x,z)}{Q(z \vee x)}$.

So, we will minimize $-\sum_z Q(z|x) log \frac{P(x,z)}{Q(z \vee x)}$. See the lecture videos for detailed derivations.

QUESTION 10:



Above figure shows latent vector subtraction of two concepts of "face with glasses" and "glasses". What is expected from the resultant vector?

a. glasses
b. face without glasses
c. face with 2 glasses
d. None of the above

Correct Answer: b

Detailed Solution:

It is expected that VAE latent space follows vector arithmetic. Thus the resultant vector is a vector subtraction of the two concepts which will result in the final vector to represent a face without glasses.

Face with glasses - ? = glasses

? = (face with glasses) – (glasses)

? = face without glasses

**************END*******