Q1-1: Select the correct option.

A. Autoencoders are a supervised learning technique.

B. Autoencoder's output is exactly the same as the input.

1. Both the statements are TRUE.
2. Statement A is TRUE, but statement B is FALSE.
3. Statement A is FALSE, but statement B is TRUE.
4. Both the statements are FALSE.

Q1-1: Select the correct option.

A. *Autoencoders are a supervised learning technique.*

B. *Autoencoder's output is exactly the same as the input.*

1. Both the statements are TRUE.
2. Statement A is TRUE, but statement B is FALSE.
3. Statement A is FALSE, but statement B is TRUE.
4. Both the statements are FALSE.

Autoencoders are an unsupervised learning technique.

The output of an autoencoder are indeed pretty similar, but not exactly the same.

Q1-2: Select the correct option.

A. One way to implement undercomplete autoencoder is to constrain the number of nodes present in hidden layer(s) of the neural network.
B. Autoencoders are capable of learning nonlinear manifolds (a continuous, non-intersecting surface.)

1. Both the statements are TRUE.
2. Statement A is TRUE, but statement B is FALSE.
3. Statement A is FALSE, but statement B is TRUE.
4. Both the statements are FALSE.

Q1-2: Select the correct option.

A. *One way to implement undercomplete autoencoder is to constrain the number of nodes present in hidden layer(s) of the neural network.*

B. *Autoencoders are capable of learning nonlinear manifolds (a continuous, non-intersecting surface.)*

1. Both the statements are TRUE.
2. Statement A is TRUE, but statement B is FALSE.
3. Statement A is FALSE, but statement B is TRUE.
4. Both the statements are FALSE.

Having less number of nodes in hidden layer(s) limits the amount of information that can flow through the network.

Autoencoders can be viewed as a generalization of PCA (another dimensionality reduction method which discovers lower dimensional hyperplane -- *will study in later lectures...*)

Q2-1: Select the correct option about Sparse autoencoders.

A. Sparse autoencoders introduces information bottleneck by reducing the number of nodes at hidden layers.

B. The idea is to encourage network to learn an encoding and decoding which only relies on activating a small number of neurons.

1. Both the statements are TRUE.
2. Statement A is TRUE, but statement B is FALSE.
3. Statement A is FALSE, but statement B is TRUE.
4. Both the statements are FALSE.

Q2-1: Select the correct option about Sparse autoencoders.

A. Sparse autoencoders introduces information bottleneck by reducing the number of nodes at hidden layers.

B. The idea is to encourage network to learn an encoding and decoding which only relies on activating a small number of neurons.

1. Both the statements are TRUE.
2. Statement A is TRUE, but statement B is FALSE.
3. Statement A is FALSE, but statement B is TRUE.
4. Both the statements are FALSE.

Sparse autoencoders introduces an information bottleneck without requiring a reduction in the number of nodes at hidden layers.

It encourages network to learn an encoding and decoding which only relies on activating a small number of neurons. Interestingly, this is a different approach towards regularization, as we normally regularize the weights of a network, not the activations!

Q2-2: Select the correct option about Denoising autoencoders.

A. *The loss is between the original input and the reconstruction from a noisy version of the input.*
B. *Denoising autoencoders can be used as a tool for feature extraction.*

1. Both the statements are TRUE.
2. Statement A is TRUE, but statement B is FALSE.
3. Statement A is FALSE, but statement B is TRUE.
4. Both the statements are FALSE.

Q2-2: Select the correct option about Denoising autoencoders.

A. *The loss is between the original input and the reconstruction from a noisy version of the input.*
B. *Denoising autoencoders can be used as a tool for feature extraction.*

1. Both the statements are TRUE.
2. Statement A is TRUE, but statement B is FALSE.
3. Statement A is FALSE, but statement B is TRUE.
4. Both the statements are FALSE.

The denoising encourages the encoder to keep important information but forget about spurious information about the input. Then the hidden representation can be viewed as preserving useful features from the input.

Q3-1: Select the correct option.

A. *Boltzmann machines are non-deterministic generative Deep learning models with 3 types of nodes: visible, hidden and output nodes.*

B. *Boltzmann machines fall into the class of Unsupervised learning.*

1. Both the statements are TRUE.
2. Statement A is TRUE, but statement B is FALSE.
3. Statement A is FALSE, but statement B is TRUE.
4. Both the statements are FALSE.

Q3-1: Select the correct option.

A. *Boltzmann machines are non-deterministic generative Deep learning models with 3 types of nodes: visible, hidden and output nodes.*

B. *Boltzmann machines fall into the class of Unsupervised learning.*

1. Both the statements are TRUE.
2. Statement A is TRUE, but statement B is FALSE.
3. Statement A is FALSE, but statement B is TRUE.
4. Both the statements are FALSE.

Boltzmann machine has only 2 types of nodes: visible and hidden. It doesn't have output nodes.

Q3-2: Select the correct option about Restricted Boltzmann Machines (RBM).

A. *RBM is ‘restricted’ to have only the connections between the visible and the hidden units.*
B. *RBM performs discriminative learning similar to what happens in a classification problem.*
C. *If number of visible nodes = nV, number of hidden nodes = nH, then number of connections in RBM = nV* nH*

1. True, True, True
2. True, False, True
3. False, False, True
4. True, False, False

Q3-2: Select the correct option about Restricted Boltzmann Machines (RBM).

A. *RBM is ‘restricted’ to have only the connections between the visible and the hidden units.*

B. *RBM performs discriminative learning similar to what happens in a classification problem.*

C. *If number of visible nodes = nV, number of hidden nodes = nH, then number of connections in RBM = nV* nH*

1. True, True, True
2. True, False, True
3. False, False, True
4. True, False, False

RBM estimates the probability distribution, instead of learning a prediction function. That is, it tries to find parameter values for $W\ b\ c$, so that the corresponding probability distribution $p(x; W, b, c)$ defined by the RBM model can approximate the true probabiliyt distribution over data points x. This is known as generative learning, as opposed to discriminative learning that happens in a classification problem (which tries to learn parameter θ for a prediction function $y = f(x; \theta)$ mapping input x to label y).