1. Let $X \in \mathbb{R}$ be a random variable with mean μ_x and variance σ_x^2. Similarly, let $Y \in \mathbb{R}$ be a random variable with mean μ_y and variance σ_y^2. If X and Y are independent random variables, then E[XY] = ?

   A. $\mu_x \cdot \mu_y$
   B. μ_x
   C. μ_y
   D. $\sigma_x \cdot \sigma_y$

   Solution: Option A is the correct answer

2. Continuing the above question, Var(XY) = ?

   A. $\sigma_x^2\sigma_y^2$
   B. $\sigma_x^2\sigma_y^2 + \sigma_x^2\mu_y^2 + \sigma_y^2\mu_x^2$
   C. $\mu_x^2\mu_y^2$
   D. $\sigma_x\mu_x + \sigma_y\mu_y$

   Solution: Option B is the correct answer

$$
\begin{aligned}
Var(XY) &= E[(XY)^2] - \{E[XY]\}^2 && (\because Var(A) = E[A^2] - \{E[A]\}^2) \\
&= E[X^2Y^2] - \{E[X]E[Y]\}^2 && (\because \text{X and Y are independent}) \\
&= E[X^2]E[Y^2] - E[X]^2E[Y]^2 \\
&= (\sigma_x^2 + \mu_x^2)(\sigma_y^2 + \mu_y^2) - \mu_x^2\mu_y^2 && (\because E[A^2] = Var(A) + \{E[A]\}^2) \\
&= \sigma_x^2\sigma_y^2 + \sigma_x^2\mu_y^2 + \sigma_y^2\mu_x^2 + \mu_x^2\mu_y^2 - \mu_x^2\mu_y^2 \\
&= \sigma_x^2\sigma_y^2 + \sigma_x^2\mu_y^2 + \sigma_y^2\mu_x^2
\end{aligned}
$$

3. Is ReLU activation function differentiable at origin?

   A. True
   B. False

Solution: Option B is the correct answer.

4. Is ReLU activation function a non-linear activation function?

   A. True

   B. False

Solution: Option A is the correct answer.

5. The maxout activation function is given by $f(x) = max(w_1^T x + b_1, w_2^T x + b_2)$. You can obtain the ReLU function from this by setting:

   A. $b_1 = b_2 = 0$

   B. $w_1 = b_1 = b_2 = 0$

   C. $w_1 = w_2 = 0$

Solution: Option B is the correct answer.

6. The symmetry breaking problem occurs only if all the weights in a layer are initialized to zero.

   A. True

   B. False

Solution: Option B is the correct answer.

7. The batch normalization layer does not introduce any new parameters.

   A. True

   B. False

Solution: Option B is the correct answer.

8. A deep neural network with linear activation functions throughout and no bias parameters is equivalent to a shallow neural network

A. True

B. False

Solution: Option A is the correct answer.

9. We can use backpropogation to train a deep neural network only if all the hidden layers in the network have the same activation function.

A. True

B. False

Solution: Option B is the correct answer.

10. Consider the functions $f(x) = \frac{1}{1+e^{-}x}$ and $g(x) = tanh(x)$. Which of the following statements is true ?

A. $g(x) = 2 * f(x) - 1$

B. $g(x) = 2 * f(2x) - 1$

C. $g(x) = f(\frac{x}{2}) - 0.5$

D. $g(x) = f(\frac{x}{2}) - 1$

Solution: Option B is the correct answer.

11. Suppose there are 16 neurons in layer1 and the 64 neurons in layer 2. The weight matrix connecting the neurons in layer 1 to layer 2 will be $W \in \mathbb{R}^{16\times64}$. Further, assume that you are using the logistic activation function in all the hidden layers of your network. If you are using Xavier initialization to initialize the weights of this layer then you will draw the weights from a unit Gaussian and multiply them by ?

A. $\frac{1}{4}$

B. $\frac{1}{8}$

C. $\frac{1}{16}$

D. $\frac{1}{64}$

Solution: Option A is the correct answer.

12. Continuing the above question but this time assuming that you are using the ReLU activation function in all the hidden layers of your network. Now, if you are using He initialization to initialize the weights of the layer mentioned in the previous question then you will draw the weights from a unit Gaussian and multiply them by ?

A. $\frac{1}{32\sqrt{2}}$

B. $\frac{1}{8\sqrt{2}}$

C. $\frac{1}{4\sqrt{2}}$

D. $\frac{1}{2\sqrt{2}}$

Solution: Option D is the correct answer.