

1. Find the eigenvalues and eigenvectors of a 2x2 matrix

$$A = \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}$$

2. Find the basis of the column space of the matrix

$$A = \begin{pmatrix} 1 & 2 \\ 2 & 4 \end{pmatrix}$$

3. Suppose we have classification problem with 3 classes X, Y and Z. Given that for the probability distribution p,  $p(X) = 0.7$ ,  $p(Y) = 0.2$  and  $p(Z) = 0.1$ . Calculate the entropy of the given probability distribution.
4. State 4 applications of deep learning.
5. State 4 reasons why deep learning is getting so famous?
6. Differentiate between a biological neuron and a perceptron.
7. Consider a perceptron with 3 inputs  $x_1=1$ ,  $x_2=1$ ,  $x_3=0$ , the weights corresponding to the 3 inputs are  $w_1=2$ ,  $w_2=-4$ ,  $w_3=1$  and the activation function is the step function. Find the output value  $Y_{pred}$  of the perceptron.
8. State different nonlinear activation functions.
9. What is the limitation of ReLu activation function? State the variance of ReLu function.
10. Consider a neural network with one input layer, one hidden layer with 2 neurons and one output layer with one neuron.
11. Draw the architecture of the neural network with the given data and calculate the total learnable parameters.
12. Explain how momentum-based gradient descent works and how does it differs from basic gradient descent?
13. What is the Adam optimizer and how does it improve over standard gradient descent?  
How PCA works in dimensionality reduction? State the disadvantages of PCA.
14. What is autoencoder? How does it work?
15. What are the loss functions used in autoencoder, elaborate?
16. Give the formula to calculate the output size in CNN with stride and padding.
17. Consider a CNN with the following layers and calculate the total no. of parameters.

Conv layer: 16 filters, 3\*3 size, 3 input channels

Conv layer: 32 filters, 3\*3 size, 16 input channels

Fully connected layer: 128 input units, 64 output units

Batch normalization after each convolutional layer.

18. Discuss the factors which affect the parameter calculation in CNN.

19. Derive the gradient descent update rule for linear regression.

20. Explain the working principle of momentum based gradient descent. How it differs from basic gradient descent?

21. Minimize  $f(x) = (x-3)^2$ . Assume  $x_0 = 0$ , learning rate = 0.1.

22. Find the local minima and maxima of  $f(x) = x^3 - 3x + 2$ .