Interview Questions

1. Discuss a deep learning project you have recently worked on? What methodologies have you employed?
2. Can you explain what vanishing gradient is? Why is this harmful?
   1. As more layers using certain activation functions(sigmoid) are added to neural networks, the gradients of the loss function approach zero, making the network hard to train.
   2. ReLu is a better alternative
3. How does deep learning differ from other machine learning algorithms?
   1. DL is a subfield of machine learning in AI that comes with networks capable of learning unsupervised from unlabeled or unstructured data.
4. Can you give some real-life examples where auto-encoders can be applied?
   1. Image denoising
   2. Feature extraction
   3. Image compression
   4. Recommender systems
5. What Is Data Normalization, and Why Do We Need It?
   1. The process of standardizing and reforming data is called “Data Normalization.” It’s a pre-processing step to eliminate data redundancy. Often, data comes in, and you get the same information in different formats. In these cases, you should rescale values to fit into a particular range, achieving better convergence.
6. What Do You Understand by Backpropagation?
   1. Backpropagation is a technique to improve the performance of the network. It backpropagates the error and updates the weights to reduce the error employing chain rule.
7. What Will Happen If the Learning Rate Is Set Too Low or Too High?
   1. if learning rate is too low, training of the model will progress very slowly as we are making minimal updates to the weights. It will take many updates before reaching the minimum point.
   2. If the learning rate is set too high, this causes undesirable divergent behavior to the loss function due to drastic updates in weights.
8. Have you employed any data augmentation technique before?
9. How do you combat overfitting?
   1. Reduce model complexity
   2. Use larger dataset
   3. Apply regularization
   4. Use dropout layers
10. How would you build, train, and deploy a system that detects if multimedia and/or ad content violates terms or contains offensive materials? What data would you collect for training?
    1. Its essentially an object detection/recognition task
    2. I could use transfer learning on YOLO pretrained on Imagenet
    3. Deploy it to any device that’s connected to a video feed.
    4. Collect Dataset containing class of target objects that need to be identified

Phase2

* + ACC, Precision, Recall/sensitivity/TP, TN (specificity)
  + If we have only two classes, confusion matrix
  + ROC curve (Receiver Operating Characteristics)
    - Perfect classifier: no false positive rate, all negatives are classified as negative.
    - Random decision
    - Low True positive rate
    - Something wrong
* Suppose you have a model trained on the ImageNet data set for a classification task. Then you feed the model with a blank image where every pixel is the same white colour. For this input, the network will output the same score for each class. This statement is True?
  + False
* What's the purpose of the backpropagation algorithm?
  + It computes all gradients required for the optimization of the network
* What is the purpose of the momentum used in different optimizers?
  + It stabilises the training by computing the moving average over the previous gradient.
* What approaches can be used to reduce overfitting?
  + Data Augmentation [y]
  + Dropout [y]
  + Low learning rate [N]
* You train a model with batch gradient descent that is supposed to differentiate images with apples and with oranges, but you don’t get a satisfactory performance. You inspect your training data and see that the training data is not shuffled. Would it help to shuffle your training data? Explain your answer briefly?
  + No, it would not help because batch gradient descent uses the whole training set for an update.
* You train one network with l2 regularisation and a second network with l1 regularization. How do you expect the respective weights to behave compared to training a network without normalization?
  + For the network trained with l2 normalisation, we expect smaller weights compared to the network trained without normalization (add. explanation: since large weights are penalized quadratically), whereas for the l1 normalized network, the weights are expected to be more sparse (add. explanation: some weights are close to 0 while others are not).
* Why is YOLO called a ”Single-Shot-Detector” (SSD)? You already work and expert in it.
  + SSDs combine the prediction of the bounding box and the classification of each box into one model and one forward pass call.

* **DETECTION:** You receive close-up images from shore regions and you are responsible for developing a system that automatically classifies whether the current image contains trash or not. Assume that you have training data with bounding-box annotations for each piece of trash. Briefly explain whether this is a form of weakly supervised learning.
  + No, labels contain more information than the task requires.
* One stage vs two stage object detection? What is it?
  + OS: Object classification and bounding-box regression are done directly without using pre-generated region proposals (candidate object bounding boxes).
  + TS: ***Generation of region proposals***, e.g. by selective search as in R-CNN and Fast R-CNN, or by a Region Proposal Network (RPN) as in Faster R-CNN. ***Object classification*** *for each region proposal*. Additionally, other things can be done such as bounding-box regression for refining the region proposals, binary-mask prediction, etc.
  + Two-stage detectors usually reach better accuracy but are slower than one-stage detectors.
* You are participating in the student Deep Learning challenge of your university. The task is to classify whether a person is smiling in portrait images, but you have only little training data available. You choose a network with 3 convolutional layers together with ReLU activation and one fully connected layer at the end. You observe that you have a good performance on your training data but poor performance on your validation data, so your tutor recommends using regularization techniques.
  + What is the general purpose of regularization?
  + Solution: Bias/variance trade-off, counter overfitting, generalizing on unseen data, enforcing prior
* **How would you differentiate between Multilabel and MultiClass classification?**