1)Define overfitting and underfitting in machine learning. What are the consequences of each, and howcan they be mitigated?

Ans- Underfitting means that your model makes accurate, but initially incorrect predictions. In this case, train error is large and val/test error is large too. Overfitting means that your model makes not accurate predictions. In this case, train error is very small and val/test error is large.

2) How can we reduce overfitting? Explain in brief.

Ans- How can we reduce overfitting?

Here we will discuss possible options to prevent overfitting, which helps improve the model performance.

Train with more data. ...

Data augmentation. ...

Addition of noise to the input data. ...

Feature selection. ...

Cross-validation. ...

Simplify data. ...

Regularization. ...

Ensembling

3) Explain underfitting. List scenarios where underfitting can occur in ML.

Ans- When a model has not learned the patterns in the training data well and is unable to generalize well on the new data, it is known as underfitting. An underfit model has poor performance on the training data and will result in unreliable predictions.

Underfitting of machine learning models happens when you are not able to reduce the training error. This can happen in some of the following scenarios: When the training set has far fewer observations than variables, this may lead to underfitting or low bias machine learning models

4) Explain the bias-variance tradeoff in machine learning. What is the relationship between bias and variance, and how do they affect model performance?

Ans- Bias-Variance Tradeoff helps optimize the error in our model and keeps it as low as possible. An optimized model will be sensitive to the patterns in our data, but at the same time will be able to generalize to new data. In this, both the bias and variance should be low so as to prevent overfitting and underfitting.

Bias is the difference between our actual and predicted values. Bias is the simple assumptions that our model makes about our data to be able to predict new data.

When the Bias is high, assumptions made by our model are too basic, the model can’t capture the important features of our data. This means that our model hasn’t captured patterns in the training data and hence cannot perform well on the testing data too. If this is the case, our model cannot perform on new data and cannot be sent into production.

Variance is the very opposite of Bias. During training, it allows our model to ‘see’ the data a certain number of times to find patterns in it. If it does not work on the data for long enough, it will not find patterns and bias occurs. On the other hand, if our model is allowed to view the data too many times, it will learn very well for only that data. It will capture most patterns in the data,  but it will also learn from the unnecessary data present, or from the noise.

We can define variance as the model’s sensitivity to fluctuations in the data. Our model may learn from noise. This will cause our model to consider trivial features as important.

5) Discuss some common methods for detecting overfitting and underfitting in machine learning models. How can you determine whether your model is overfitting or underfitting?

Ans- We can determine whether a predictive model is underfitting or overfitting the training data by looking at the prediction error on the training data and the evaluation data. Your model is underfitting the training data when the model performs poorly on the training data.

6) Compare and contrast bias and variance in machine learning. What are some examples of high bias and high variance models, and how do they differ in terms of their performance?

Ans- Bias and variance are inversely connected. It is impossible to have an ML model with a low bias and a low variance. When a data engineer modifies the ML algorithm to better fit a given data set, it will lead to low bias—but it will increase variance

Examples of high-variance machine learning algorithms include: Decision Trees, k-Nearest Neighbors and Support Vector Machines

Let's say returns for stock in Company ABC are 10% in Year 1, 20% in Year 2, and −15% in Year 3. The average of these three returns is 5%. The differences between each return and the average are 5%, 15%, and −20% for each consecutive year.

7) What is regularization in machine learning, and how can it be used to prevent overfitting? Describe some common regularization techniques and how they work.

Ans- Regularization in machine learning is the process of regularizing the parameters that constrain, regularizes, or shrinks the coefficient estimates towards zero. In other words, this technique discourages learning a more complex or flexible model, avoiding the risk of Overfitting.

Regularization is a set of techniques that can prevent overfitting in neural networks and thus improve the accuracy of a Deep Learning model when facing completely new data from the problem domain. In this article, we will address the most popular regularization techniques which are called L1, L2, and dropout.