Q1.

Define overfitting and underfitting in machine learning. What are the consequences of each, and how can they be mitigated?

- Overfitting :
  - when model learns training data too well , but poorly on new unseen data.
  - High accuracy on training set but low accuracy on test set.
  - low bias and high variance
  - Regularization
  - Cross-Validation.
- Underfitting :
  - when a model performs poorly on both training set and new data.
  - high bias and high variance
  - inability to capture the underlying trends in the data
  - Increase model complexity
  - Feature Engineering

Q2.

How can we reduce overfitting? Explain in brief.

- Cross-Validation :
  - Use techniques like k-fold cross-validation to assess the model's performance on different subsets of the data.
- Feature Selection :
  - Identify and remove irrelevant or redundant features that may contribute to overfitting.
- Regularization :
  - Introduce regularization terms in the models objective function to penalize overly complex models.

Q3.

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

- Underfitting occurs when a machine learning model is too simplistic to capture the underlying patterns in the data.

- Insufficient Model Complexity

- Limited Features

- Inadequate Training

- Small Training Dataset.

Q4.

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

- It refers to the balance that needs to be struck between two types of errors, bias error and variance error, when training a machine learning model.

- As you decrease bias, you tend to increase variance and vice versa.

- Impact :

  - High bias : Performs poorly on both training and test data.
  - High variance : Performs well on training data but poorly on test data.

Q5.

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

- Cross-Validation
- Learining Curves
- Model Complexity Curves
- Error Analysis
- Regularization Inspection

Determining :

```
- Overfitting : Training error is significantly lower than
validation error.
- Underfitting : Validation error remains high even as training
progresses.
```

Q6.

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?

- Bias :
    - High bias associated with underfitting.
    - The model is too simplistic and fails to capture the underlying patterns in the data.
    - Performance is poor on both the training and test sets.
- Variance :
    - High variance is associated with overfitting.
    - The model is too complex and may fit the noise in the training data.
    - Performance is good on the training set but poor on the test set.

Examples :

- High Bias Model : Linear Regression model applied to highly nonlinear dataset.
- High Variance Model : Decision tree with a large depth applied to a small dataset.

Performance :

```
- High Bias:

    - Training Set: Poor performance
    - Test Set: Poor performance

- High Variance:

    - Training Set: Good performance
    - Test Set: Poor performance
```

Q7.

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

- Regularization in machine learning is a set of techniques used to prevent overfitting and improve the generalization performance of a model

Common Regularization Techniques :

- L1 Regularization
- L2 Regularization

- Elastic Net Regularization
- Dropout