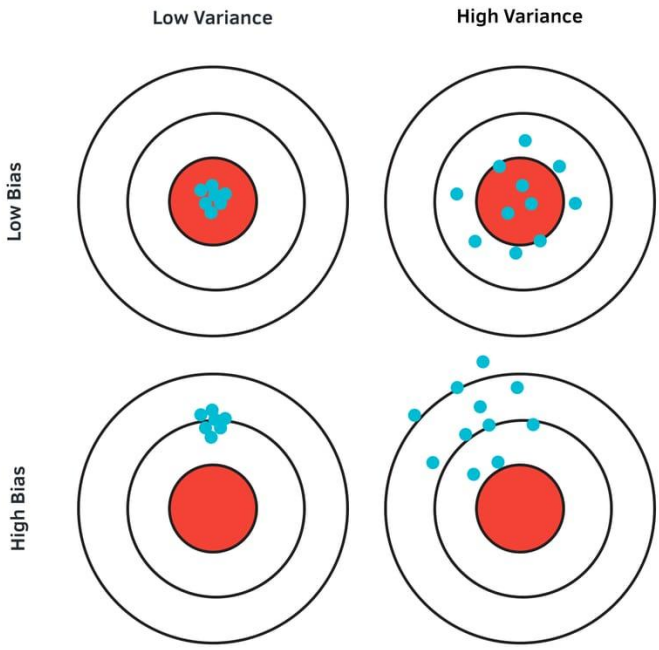
ASSIGNMENT-1

1.Variance and Bias (Diagram, overfit, underfit)-For best fit model should we have low bias or high variance, low bias or low variance, high bias or high variance, low bias or high variance.



Bias-Variance Tradeoff

1.What is variance?

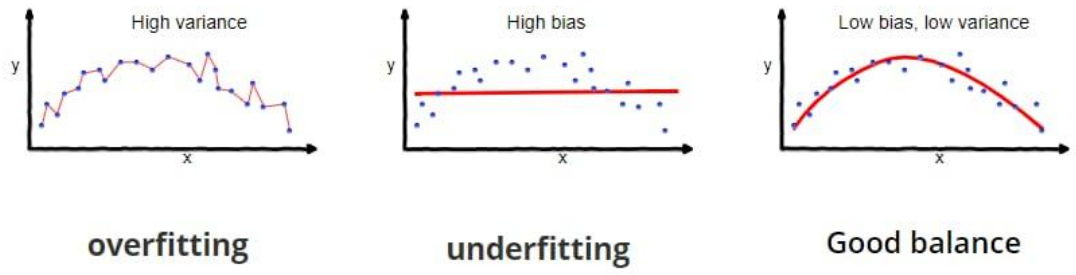
>Variance in machine learning refers to how much a model's predictions change when it is trained on different datasets.

>If a small change in the training data causes large changes in the model's predictions, the model is said to have high variance.

>High variance usually occurs when the model is too complex and learns not only the underlying pattern but also the noise in the data, which leads to overfitting.

2.What is bias?

>Bias in machine learning refers to the error that occurs when a model makes overly simple assumptions about the data. It measures how far the model's predictions are from the actual true values on average.



>A model with high bias fails to capture the underlying patterns in the data, leading to underfitting, where both training and test accuracy are low.

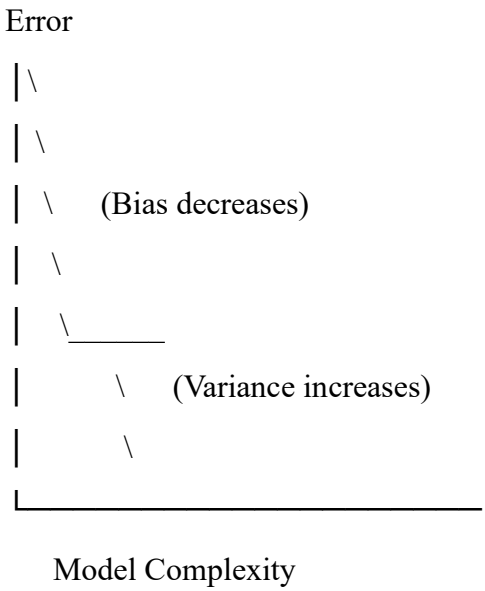
3.Overfitting and Underfitting:-

Overfitted Model:-

>overfitted model is a model that learns the training data too well, including random noise, outliers, and small fluctuations that do not represent the true underlying pattern.

> for examples As a result, it performs extremely well on the training dataset but performs poorly on new, unseen data.

> It performs very well on training data but poorly on unseen test data. Overfitting happens when the model has **low bias and high variance**, meaning it is highly sensitive to small changes in the training dataset and does not generalize well.



Left → Underfitting

Middle → Best Fit

Right → Overfitting

Underfitted model:-

>An underfitted model is a model that is too simple to learn the underlying pattern of the data.

> It cannot capture the important relationships between input and output variables, so it performs poorly on both the training data and new (test) data.

> Underfitting usually occurs when the model has **high bias and low variance**, meaning it makes strong assumptions about the data and does not adapt well to its complexity.

>For example, if the actual data follows a curved pattern but we use a straight-line model to represent it, the model will miss the true relationship.

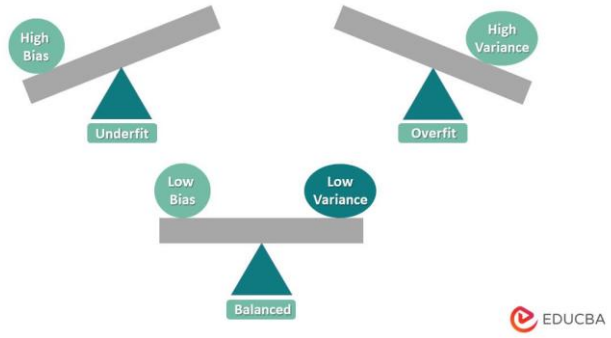
4.Balanced Model:-

>A balanced model is a model that achieves the right trade off between bias and variance. It is neither too simple (which causes underfitting) nor too complex (which causes overfitting). Instead, it captures the true underlying pattern in the data without learning the noise.

>This means the model performs well on both the training data and unseen test data, showing good generalization ability.

>In a balanced model, the bias is low enough to represent the actual relationship in the data, and the variance is controlled so that the model does not change drastically with small variations in the training dataset.

For example:- if the data follows a curved trend, a balanced model will fit a smooth curve that follows the overall pattern without bending excessively to pass through every single data point.



1 Underfitted Model (High Bias, Low Variance)

- **Problem:** Model too simple to learn the data pattern.
- **Effect:** Poor training and test accuracy.
- **Example:**Fitting a straight line to data that is clearly curved.

2 Overfitted Model (Low Bias, High Variance)

- **Problem:** Model too complex, learns noise in the training data.
- **Effect:** Very high training accuracy, poor test accuracy.
- **Example:**A polynomial curve that twists to pass through every data point.

3 Balanced Model (Low Bias, Controlled Variance)

- **Problem:** None — this is the ideal model.
- **Effect:** Good training and test accuracy, generalizes well.
- **Example:**-Polinomial regression of moderate degree.