BIAS-VARIANCE TRADE OFF

- Our main aim of supervised machine learning algorithm is to reduce the model error.
- In any Machine learning algorithm, we need to mainly focus on 3 important concepts/nomenclature:
  - Predicted values — Values resulted from the model
  - Actual values — Original Values
  - Error — Difference between actual and predicted output
- To calculate the error, we do summation of the reducible and irreducible error also known as Bias-Variance decomposition
  - **Irreducible Error** — Those error cannot be reduced irrespective of any algorithm that we use in the model. It is caused by unusual variables that have a direct influence on the output.
  - **Reducible Error** — As the name indicates, we need to optimize this error. it has two components that is Bias and variance.
  - So, the presence of bias and variance influences the model's accuracy in several ways like overfitting and underfitting

What is Bias?

- Bias refers to distance between predicted and actual values. i.e., how far we have predicted the values from actual values

  - $\mathbf{Bias[f(X)] = E[\hat{Y}] - Y}$

- **High Bias** — if the average prediction values are far away from actual values
- **Low Bais** — Distance between prediction and actual values are minimal
- High Bias will cause the algorithm to miss a dominant pattern or relationship between input and output variables.
- If bias is too high, model performs very badly and accuracy will be low, which causes underfitting
- Bias can be reduced by,
  - Using a better ML algorithm
  - Using a larger model (i.e. with more parameters)
  - Training for more iterations if training was stopped earlier
  - Using a larger training data set
  - Reducing regularization (if exists)
- Example for high bias,
  - Using a straight line to model a quadratic polynomial distribution

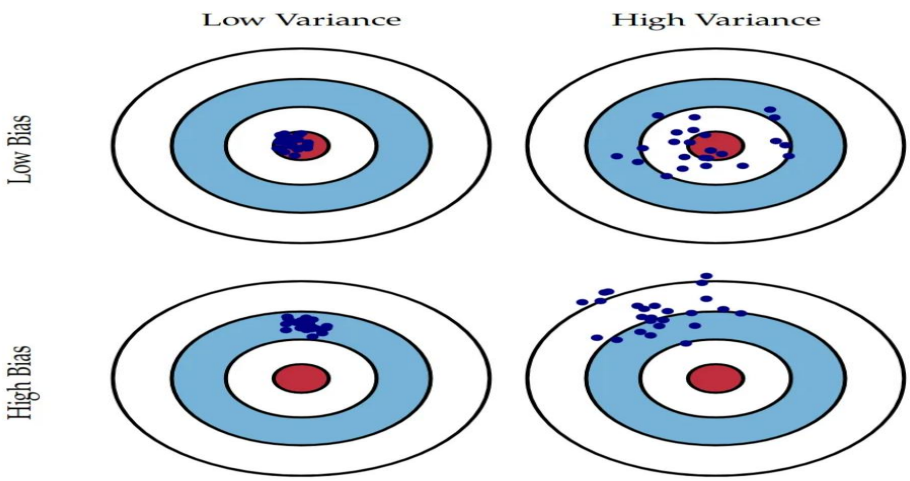
What is Variance?

- If model which predicts well with training dataset and fails with independent unseen data (Testing dataset), then it is evident that model has a variance.

  - $\mathbf{Variance[f(X)] = E[\ \{E(\hat{Y}) - \hat{Y}\}^2\ ]}$

- i.e., it conveys how scattered the predicted values are from the actual values
- We can say that How much the data is scattered from each other
- High Variance — data is scattered significantly, the model has trained with lot of noise and irrelevant data that causes the overfitting
- Low Variance — Less scattered
- Variance can be reduced by
  - Using a larger training dataset can reduce the variance as the errors get cancelled out.
  - Reducing the number of parameters can also reduce variance as the less significant insights (like noise) will not be included in the model.
  - Using Early Stopping to stop training at an optimal point
  - Dropout is used in Deep Learning models
- Example for high variance,
  - Using an 8 degree polynomial to model a linear distribution.

Bias and Variance as the form of Errors



Graphical representation of bias and variance

- Lets understand bias and variance with visual representation.
- Here bull's eye (target) nothing but actual data points and are at the center
  - LV — Low Variance
  - HV — High Variance.
  - LB — Low bias and
  - HB — High Bias

Case (i) — LB & LV (ideal and Best Scenario — Best Model)
- LB — All predicted data are at bull's eye. so distance between predicted and actual data points are very less
- LV — No scattered of data points. they are close to each other and at bull's eye.

Case (ii) — LB & HV (Model are somewhat accurate but inconsistent)
- LB — As similar to case (i) LB. i.e., distance between predicted and actual data points are very less

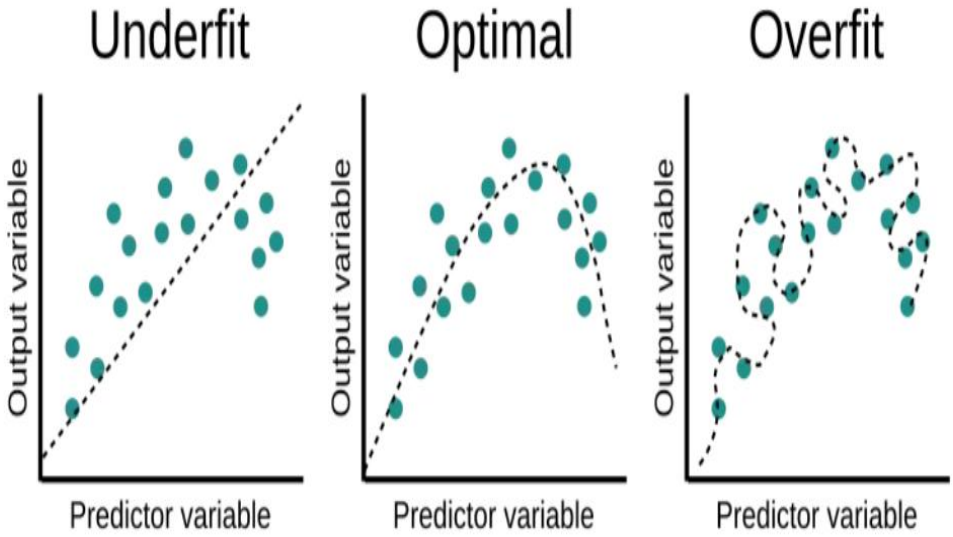
- HV — Data points are scattered to each other and are away from actual data

Case (iii) — HB & LV (Model are Consistent but inaccurate)
- LV- As similar case (i) LV, No scattered of data points.
- HB — Predicted data points are far away from blue's eye. So the distance between predicted and actual data is High(Error). i.e., Prediction is somewhere and Actual is somewhere

Case (iv) — HB & HV (Model will be inconsistent & inaccurate)
- HB — As similar to case (iii) HB.
- HV — As similar to case (ii) HV.
- for better model, the predicted data points as to be close to the actual data points (nothing but Less Error) and also it as to be less scattered. i.e., case (i)

How Bias and Variance will cause Overfitting and Underfitting?

- Dataset is divided into training and testing data.
- we will build a model using training dataset and validate a model using testing dataset.
- Bias is resulting from training error
- variance is related to testing error



UNDERFITTING

- Underfitting refers when the data model is unable to capture the relationship between the input and output variables accurately. i.e., model that can neither model the training data nor generalize to new data(test data).
- From underfitting graph, the best fit line is not covering major data points. i.e., the best fit line is away from data points zone (only some data points are around

best fit line but majority of data points are scattered from it). therefore accuracy of training as well testing will be much lesser. This phenomenon is known as UNDERFITTING.

Characteristics of Underfitting:

- The model performs poorly on both training and test datasets.
- It fails to capture the complexity of the dataset.
- Adding more training data does not improve performance significantly.
- It occurs when models are too simple (e.g., using linear regression for a non-linear dataset).

Example

- if training accuracy of the model is 65% and testing accuracy is 50%.
- Then resulted error rate is high because model is unable to capture the relationship between the input and output variables accurately (usually underfit models performs bad for both training and testing data).

OPTIMAL CONDITION

- In Optimal Condition, it is said to best/optimal model, when it performs well for training and as well as for testing dataset.
- Example: if the training accuracy of the model is 85% and testing accuracy of the model is 80–90%. resulting in less error rate.
- Because model is able to capture the relationship between the input and output variables accurately.

OVERFITTING

- Overfitting refers to a model that models the training data too well but fails with testing data.
- From Overfitting graph, when we gave a input to the model, training output will comes out in the form of above graph(overfit graph). The line will cover entire data points and when we test it using unseen data(test data) it will fail. This phenomenon is known as OVERFITTING.

Characteristics of Overfitting:

- The model performs exceptionally well on training data but poorly on test data.
- It captures noise and anomalies instead of general patterns.
- The model has too many parameters and is overly complex.
- It fails to generalize well to new datasets.

Example

- if training accuracy of model is 95% and testing accuracy is 70%.
- The error rate between training and testing is very high because model is behaving well with training but fails with testing

THE BIAS-VARIANCE TRADEOFF

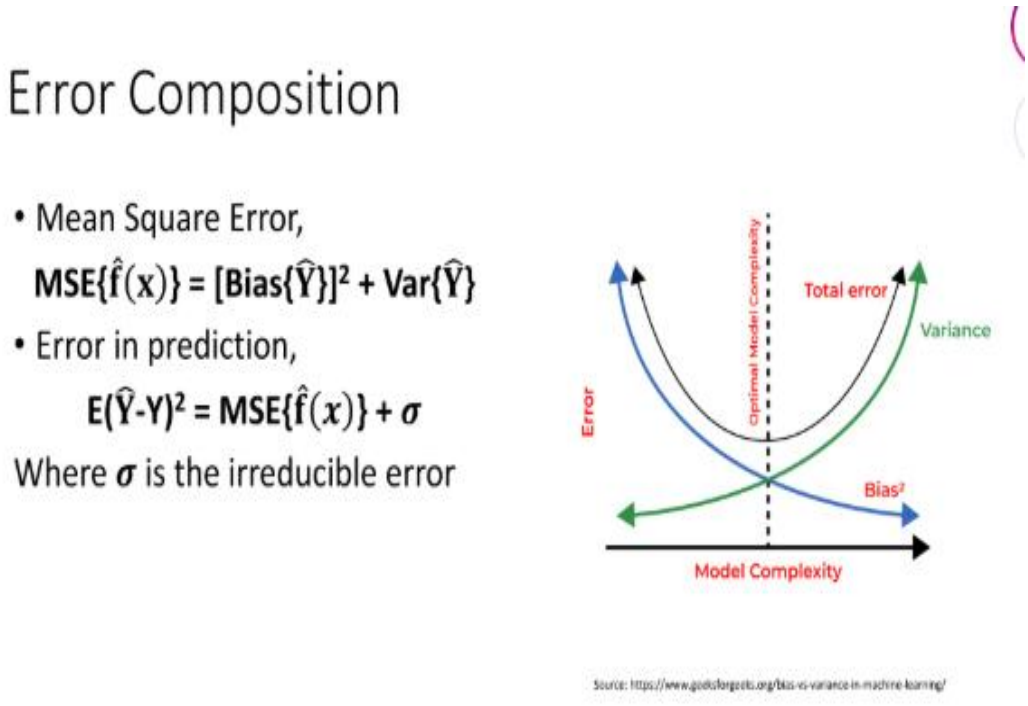
- The bias variance tradeoff describes the balance between a model being too simple and too complex.
- A simple model may miss important patterns (high bias), while a very complex model may learn noise from training data (high variance).
- The aim is to balance both so the model performs well on new data.
- Simple models usually have high bias and low variance, which may cause underfitting.
- Complex models usually have low bias but high variance, which may cause overfitting.
- Balanced model achieves an optimal point where both bias and variance are reasonably low.
- Goal of machine learning is to minimize the total prediction error on unseen data.

The total prediction error can be expressed as:

$$\textbf{TotalError} = \textbf{Bias}^2 + \textbf{Variance} + \textbf{IrreducibleError}$$

Here:

- Bias^2: Error caused by incorrect assumptions in the model.
- Variance: Error caused by sensitivity to training data.
- Irreducible Error: Random noise in the data that cannot be eliminated.

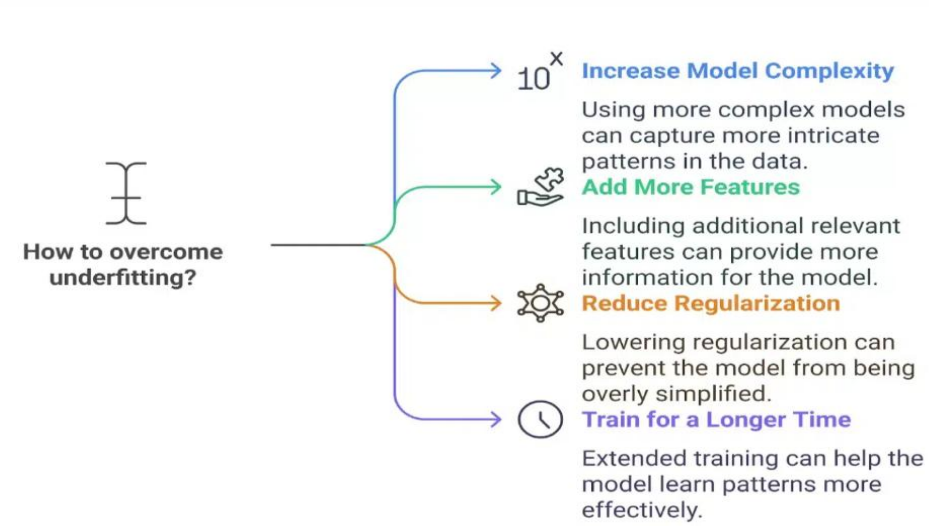


Importance of Bias Variance Tradeoff

- Understanding the bias variance tradeoff helps in building better machine learning models.
- Helps models perform well on unseen data.

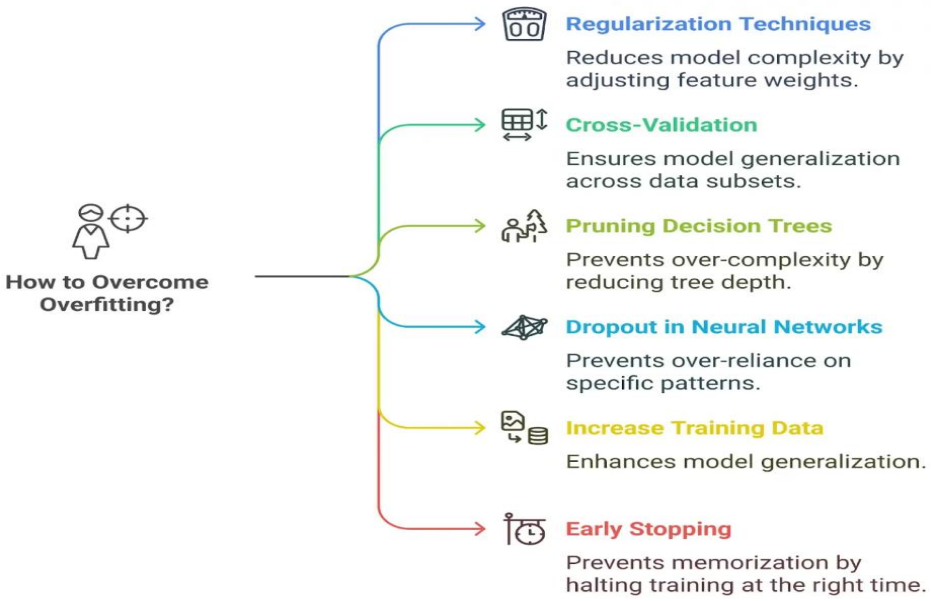
- Helps choose appropriate algorithms and complexity.
- Helps achieve better model performance.

How to Overcome Underfitting



-

How to Overcome Overfitting



Methods to Manage Bias Variance Tradeoff

- Several techniques can help balance bias and variance in machine learning models so that the model performs well on unseen data.
  - Cross Validation

- Regularization
- Feature Selection
- Ensemble Methods
- Adjusting Model Complexity

Real-World Applications

- Overfitting and underfitting are crucial concerns across various domains:
- Healthcare: Predicting diseases based on medical data requires generalizable models.
- Finance: Stock market predictions should avoid memorizing past fluctuations.
- Autonomous Vehicles: Object detection systems must generalize across diverse environments.
- Natural Language Processing: Sentiment analysis models should not memorize specific training phrases.