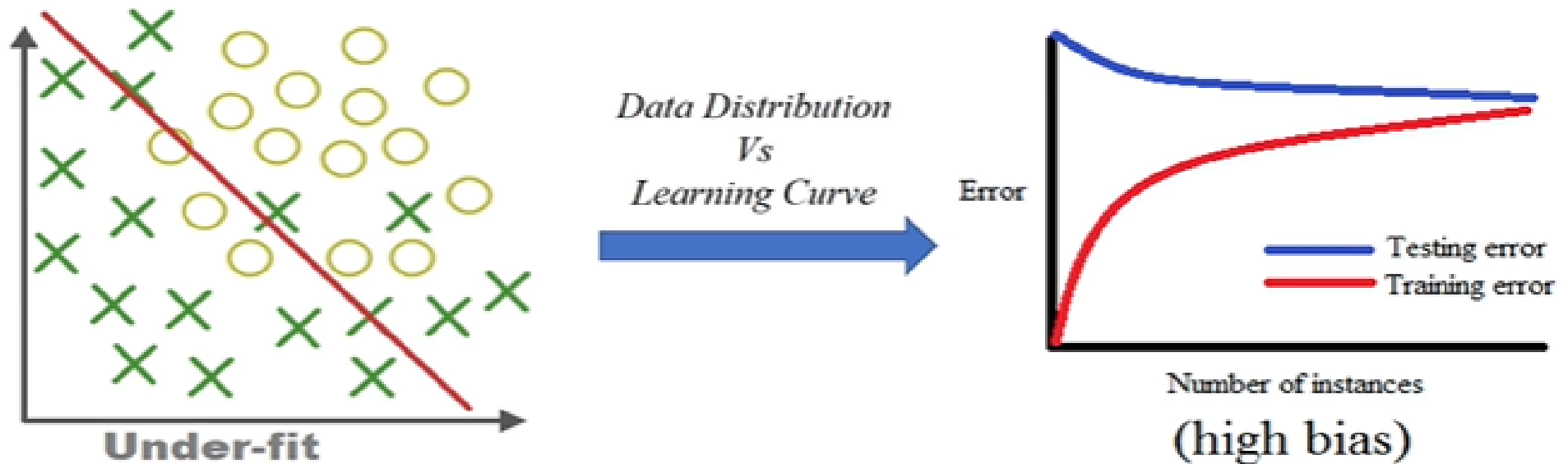


## **Learning Curve in Machine Learning**

- Learning curve shows changes in the training and validation errors with respect to the number of training examples used in training the model.
- If a model is balanced, both errors converge to small values when the training sample size increases.

# Bias

- **Bias** is defined as the mistake (error) caused by the model's simple assumptions in fitting the data.
- A high bias represents the model is unable to capture training data patterns that leads to under-fitting.
- A good model should have minimal bias.

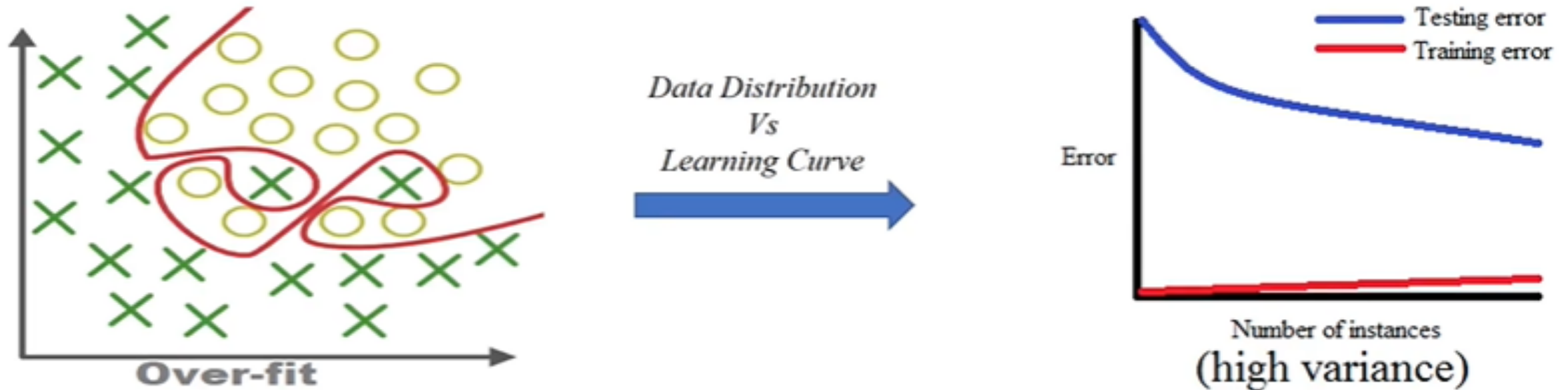


## Underfitting

- If a model contains **high bias**, model is **underfitting** with the data.
- In this case, both errors fail to decrease even increases the training set.

# Variance

- **Variance** is defined as the error caused by the complicated model tries to match the training data.
- When a model has a high variance, it passes over the majority of data points that leads the data to overfit.
- A good model should have minimal variance.

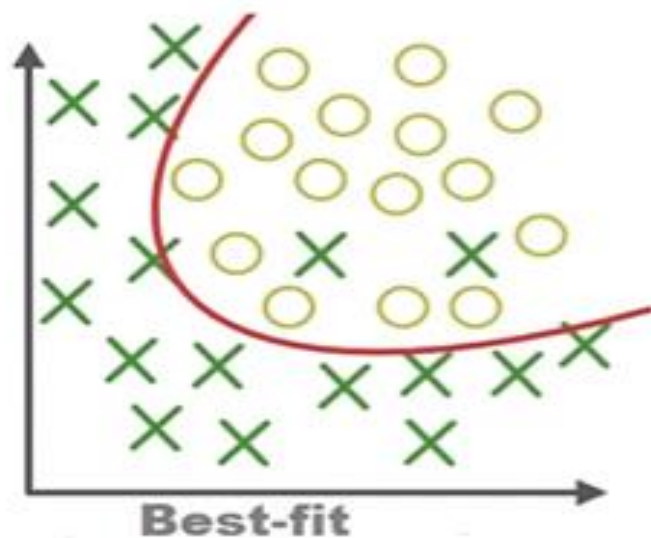


## Overfitting

- Sometime model performs well with the training data but does not perform well with the test data i.e., model is **overfitting** with the training data due to data with **high variance** and **Low bias**.
- In this case, increasing the training sample size decreases the training error but it fails to decrease the validation (test) error.

## Best fit - Bias and Variance Trade-off

- A high bias represents the model is unable to capture training data patterns that leads to under-fitting.
- When a model has a high variance, it passes over the majority of data points that leads the data to overfit.
- As the model's **complexity** grows, the bias reduces while the variance increases, and vice versa.



*Data Distribution  
Vs  
Learning Curve*



### Best fit

- A good model (Best-fit) should have **minimal** variance and bias. However, both is nearly impossible.
- So, a **trade-off** must be made to build a strong model that performs well on both train and unseen data.