# Evaluation Metrics

---

## Metrics

1. Bias (Linear Regression)
2. Coefficient of Determination / R squared (Linear Regression)
3. Adjusted R Squared (Linear Regression)
4. Mean Squared Error (Linear Regression)
5. Root mean Squared Error (Linear Regression)
6. Root Mean Squared Log Error (Linear Regression)
7. Mean Absolute Error (Linear Regression)
8. Explained variance (Linear Regression)
9. Mean Absolute Percentage Error (Linear Regression)
10. AUC ROC (Classification)
11. AUC PR
12. F measure
13. Accuracy (Classification)
14. Average Precision (Classification)
15. Precision (Classification)
16. Recall (Classification)
17. IoU (Multi-label / Image classification)
18. Confusion Matrix (Classification)
19. ROC curve (Sensitivity/Recall Vs Specificity)
20. PR curve
21. Cumulative gain curve
22. Lift curve
23. Calibration curve

24. Precision by label (Multi-class)
25. Recall by label (Multi-class)
26. F1 score by label (Multi-class)
27. Precision at k (Ranking/ Recommender Systems)
28. Mean Average Precision (Ranking/ Recommender Systems)
29. Normalized Discounted Cumulative Gain (Ranking/ Recommender Systems)

## Bias And Variance

Statistical or Mathematical models error can be split into reducible and irreducible error. Reducible error comprises of

1. Error due to squared bias
2. Error due to variance

### Bias

Bias is the amount that a model's prediction differs from the target value.
Quantifies how well the model predictions match the ground truths.
If the bias is low, the model matches the training data very well.
However, imagine you could repeat the whole model building process more than once: each time you gather new data and run a new analysis creating a new model.
Due to randomness in the underlying data sets, the resulting models will have a range of predictions. Bias measures how far off in general these models' predictions are from the correct value.
Cause: It results from simplifying the assumptions of data used in model so that target function is easier to approximate.
Effects:

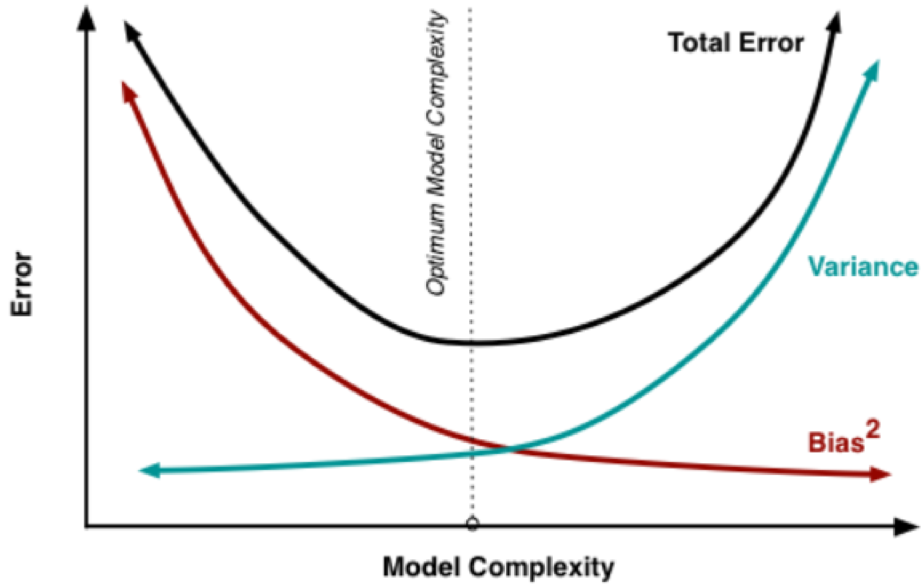
1. Underfitting in nature.
2. Simplified model
3. High error rate

### Variance

The error due to variance is taken as the variability of a model prediction for a given data point. Again, imagine you can repeat the entire model building process multiple times. The variance is how much the predictions for a given point vary between different realizations of the model.
High bias will result in low variance which makes the model suboptimal.
Causes of high variance: Noisy data set, complex models

### Trade-off between Bias and Variance

1. Increasing complexity of model: Reduces bias and increases variance : Used when model is underfitting.
2. Increasing the size of dataset: Large dataset tends to generalize easily; Hence, this method is preferred when dealing with overfitting models.
3. Bagging(**B**ootstrap **Agg**regatin**g**) reduces variance




## Coefficient of Determination R-squared Error

This metric represents the proportion of variance explained by the model.
$R^2$ corresponds to the degree to which the variance in the dependent variable (the target) can be explained by the independent variables (features).

$$
\begin{aligned}
R^2 &= 1 - \frac{\sum_{i=1}^{N}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{N}(y_i - \bar{y}_i)^2} \\
&= 1 - \frac{\text{(Unexplained Variance)}}{\text{(Total Variance)}}
\end{aligned}
$$

From the formula, it can be seen comparing the fit of the model with simple mean model where the prediction for each observation is same as the mean of all observations.
Dependent variable(y) has variance $\frac{\sum_{i=1}^{N}(y_i - \bar{y}_i)^2}{N}$ which cannot be explained by the simple model.
Higher $R^2$ value means a greater fit.
It assumes every feature helps in explaining the variation in target.
If not linear model, $R^2$ may be negative.
It does not determine the goodness of fit of the model. Since it does not give any measure of bias, highly overfitted model may have higher $R^2$.

## Adjusted R squared

Since $R^2$ metric assumes every feature helps in explaining the variance, it might explain 100% of variance but in reality it is not learning but overfitting.

Hence, we penalise adding features that are not useful for predicting the target. The value of the adjusted $R^2$ decreases if the increase in the $R^2$ caused by adding new features is not significant enough.

$$R_a^2 = 1 - [\frac{(n-1)}{(n-k-1)}(1-R^2)]$$

where n is number of observations and k is number of independent variables.