

# Machine Learning



*Two-Pager on loss functions and metrics*



# Loss function vs Metric - Definitions

## Loss Function

- Role : Guides the training and optimization process. It is the function that the algorithm (e.g., Gradient Descent) seeks to minimize.
- Requirement : Must be differentiable across its domain to allow the calculation of the gradient (e.g., MSE is preferred over MAE for optimization).
- Usage : Calculated repeatedly on the training data (Empirical Risk) during each iteration to update model weights.

## Metric

- Role : Quantifies the model's final performance and generalization ability for human interpretation and comparison
- Requirement : Must be intuitive and reflect the business or problem objective (e.g., R2, MAE, Accuracy).
- Usage : Calculated only on the Validation and Test data after training to provide an unbiased score.

## Example :

- MSE is a loss function when used to train OLS regression models as it guides the optimization. Yet, it can also be used by the user as a metric to find the average squared error
- The term "loss function" is used when the function is actively directing the learning algorithm (optimization). The term "metric" is used when the function is passively evaluating the final model's performance for humans.

# Main Regression metrics

Metric	Formula	Intuitive Meaning (Simple Explanation)	Outlier Sensitivity	Key Advantages and Disadvantages
<b>MAE</b> ( <i>Mean Absolute Error</i> )	$MAE = \frac{1}{n} \sum_{i=1}^n  y_i - \hat{y}_i $	$y_i - \hat{y}_i$ "Absolute difference between prediction and true value"		"The average error, giving no special penalty." It's the absolute average deviation between prediction and reality.
<b>MSE</b> ( <i>Mean Squared Error</i> )	$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2$	"The average of the squared errors, where big mistakes count double."	<b>Very High/Quadratic.</b> Quadratically penalizes large errors.	Ideal loss function as it is differentiable (used in OLS).  Difficult to interpret (unit is Y squared).
<b>RMSE</b> ( <i>Root Mean Squared Error</i> )	$RMSE = \sqrt{MSE}$	"The MSE error scaled back to the original unit."	<b>Very High/Quadratic.</b> Identical to MSE sensitivity.	Unit of Y (easy to compare with the target).  Extremely sensitive to outliers.
<b>R<sup>2</sup></b> ( <i>Coefficient of Determination</i> )	$R^2 = 1 - \frac{\sum (y_i - \bar{y})^2}{\sum (y_i - \hat{y}_i)^2}$	"The percentage of the target's variance that your model explains."	<b>Not directly sensitive</b> (measures overall fit quality).	Indicates the overall goodness-of-fit (0 to 1).  <b>Always increases when adding irrelevant predictors;</b> requires Adjusted R <sup>2</sup> for true comparison.