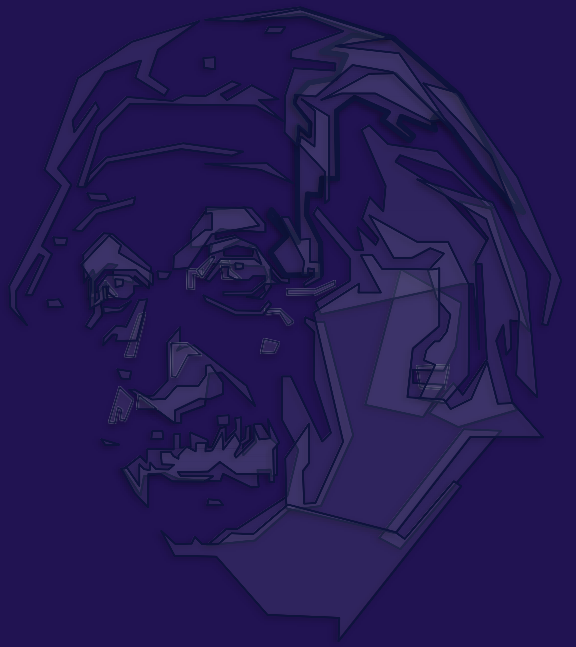
Machine Learning



Two-Pager on loss functions and metrics

Thomas Milcent - 2025

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	Outlier Sensitivity	Key Advantages and Disadvantages
MAE *(Mean Absolute Error)*	$\frac{1}{n}\sum_{i=1}^{n}\lvert y_i - \hat{y}_i\rvert$	**“ Absolute difference between prediction and true value”**		**"The average error, giving no special penalty."** It's the absolute average deviation between prediction and reality.
MSE *(Mean Squared Error)*	$\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$	**"The average of the squared errors, where big mistakes count double."**	**Very High/Quadratic. Quadratically penalizes large errors**.	- Ideal loss function as it is differentiable (used in OLS). - Difficult to interpret (unit is Y squared).
RMSE *(Root Mean Squared Error)*	$\sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}$	**"The MSE error scaled back to the original unit."**	**Very High/Quadratic.** Identical to MSE sensitivity.	- Unit of Y (easy to compare with the target). - Extremely sensitive to outliers.
R2 *(Coefficient of Determination)*	$1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2}$	**"The percentage of the target's variance that your model explains."**	**Not directly sensitive** (measures overall fit quality).	- Indicates the overall goodness-of-fit (0 to 1). - **Always increases when adding irrelevant predictors**; requires Adjusted R2for true comparison.