# REGRESSION METRICS

- MUKESH KUMAR

# AGENDA

- ✓ Mean Squared Error (MSE)
- ✓ Mean Absolute Error (MAE)
- ✓ Root Mean Squared Error (RMSE)
- ✓ R-squared ($R^2$)
- ✓ Adjusted R-squared

# Why error metrics

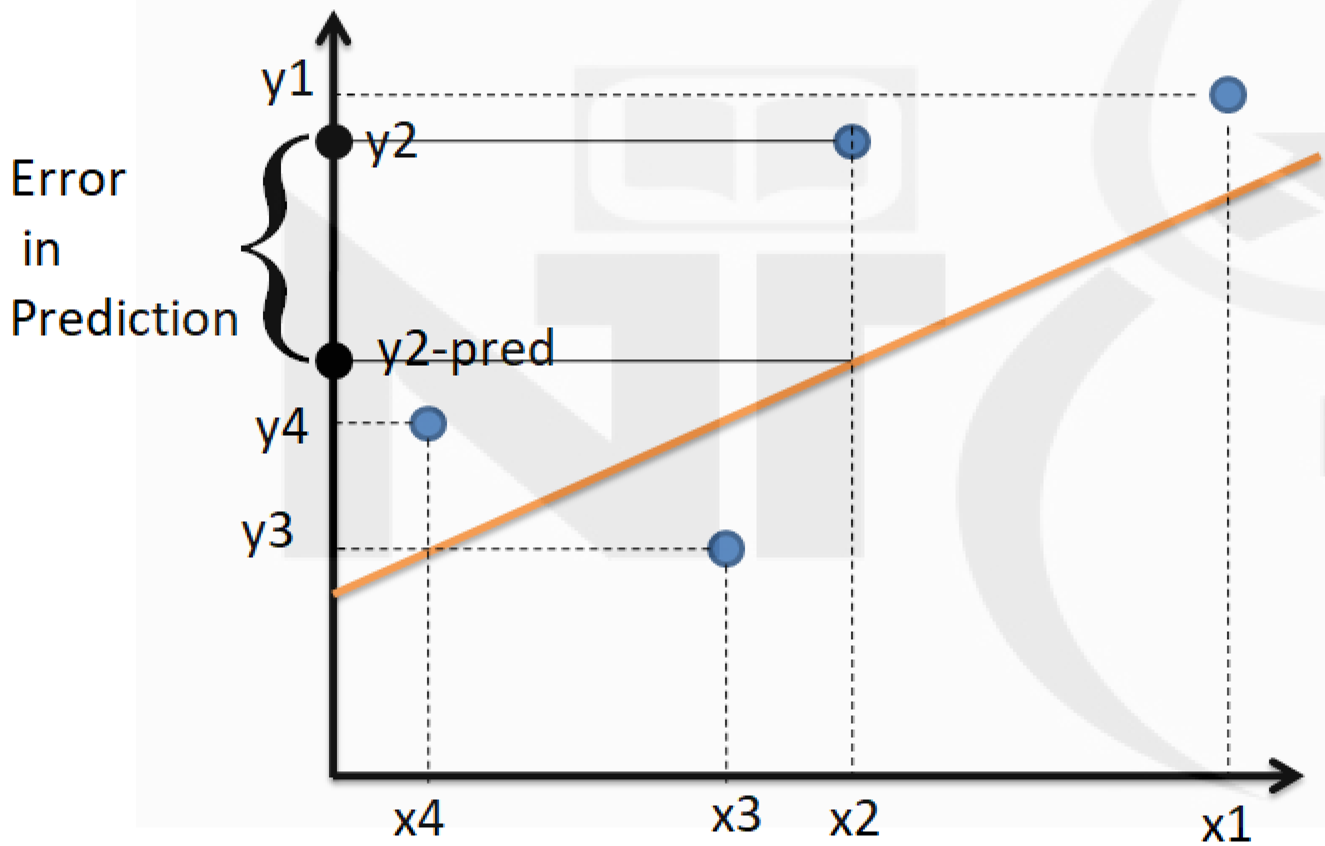
- Once you have built a regression model , how do you know how good the model is?
- There are many metrics to find how good the model is , lets look at few of them

# Mean Absolute Error (MAE)

Formula: $\mathrm{MAE} = \frac{1}{n} \sum_{i=1}^{n} |y_i - \hat{y}_i|$

**Use Case**:

- MAE measures the average magnitude of the errors without considering their direction.
- It is less sensitive to outliers compared to MSE.


y1
y2
Error
in
Prediction
y2-pred
y4
y3
x4
x3
x2
x1

## Original MAE Formula:

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^{n} |y_i - \hat{y}_i|$$

## Expanded Summation:

For a dataset with $n$ observations, the summation $\sum_{i=1}^{n} |y_i - \hat{y}_i|$ expands as follows:

$$\text{MAE} = \frac{1}{n} \left( |y_1 - \hat{y}_1| + |y_2 - \hat{y}_2| + |y_3 - \hat{y}_3| + \cdots + |y_n - \hat{y}_n| \right)$$

# Advantages of MAE:

- **Easy Interpretability:** MAE is intuitive and easy to understand, representing the average error in the same units as the target variable.
- **Less Sensitive to Outliers:** MAE doesn't exaggerate large errors since it doesn't square them, making it more robust to outliers compared to MSE.
- **No Exaggeration of Errors:** MAE treats all errors linearly, offering a balanced view of average performance without overemphasizing large deviations.

# Disadvantages of MAE:

- **Equal Weight to All Errors:** MAE treats small and large errors equally, which might not be ideal when larger errors are more significant in your context.

- **Non-differentiable at Zero:** MAE can be less convenient for optimization in gradient-based algorithms due to its non-differentiability at zero.

- **Under-penalization of Large Errors:** MAE may under-penalize large errors, making it less suitable when large deviations are particularly problematic

# Mean Squared Error (MSE)

Formula: $\mathrm{MSE} = \frac{1}{n}\sum_{i=1}^{n}\left(y_i - \hat{y}_i\right)^2$

**Use Case**:

- MSE is the most commonly used error function in linear regression.
- It penalizes larger errors more heavily due to the squaring term, making it sensitive to outliers.

## Original MSE Formula:

$$\text{MSE} = \frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$$

## Expanded Summation:

For a dataset with $n$ observations, the summation $\sum_{i=1}^{n} (y_i - \hat{y}_i)^2$ expands as follows:

$$\text{MSE} = \frac{1}{n} \left( (y_1 - \hat{y}_1)^2 + (y_2 - \hat{y}_2)^2 + (y_3 - \hat{y}_3)^2 + \cdots + (y_n - \hat{y}_n)^2 \right)$$

# Advantages of MSE:

- **Sensitive to Large Errors:** MSE squares the errors, giving more weight to larger errors, making it useful when you want to penalize large deviations more heavily.

- **Mathematically Convenient:** MSE is differentiable and widely used in gradient-based optimization algorithms, making it suitable for many machine learning models, particularly neural networks.

- **Reflects Variability in Error Distribution:** MSE captures the variability in the error distribution, providing insights into how widely the errors are spread.

# Disadvantages of MSE:

- **Less Interpretable:** The squaring of errors means MSE is not in the same units as the target variable, making it less intuitive to interpret compared to MAE.

- **Highly Sensitive to Outliers:** MSE's emphasis on larger errors makes it very sensitive to outliers, which can disproportionately affect the metric.

- **May Over-penalize Large Errors:** By squaring the errors, MSE can exaggerate the impact of large errors, which might not be desirable in all applications.

# Root Mean Squared Error (RMSE)

Formula: $\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(y_i - \hat{y}_i\right)^2}$

**Use Case**:

- RMSE is the square root of MSE and has the same units as the target variable, making it more interpretable.
- Like MSE, it is sensitive to outliers.

## Original RMSE Formula:

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}$$

## Expanded Summation for RMSE:

For a dataset with $n$ observations, the summation $\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$ inside the RMSE formula expands as follows:

$$\text{RMSE} = \sqrt{\frac{1}{n}\left((y_1 - \hat{y}_1)^2 + (y_2 - \hat{y}_2)^2 + (y_3 - \hat{y}_3)^2 + \cdots + (y_n - \hat{y}_n)^2\right)}$$

# Advantages of RMSE:

- **Same Units as Target Variable:** RMSE is in the same units as the target variable, making it more interpretable than MSE while still reflecting error magnitude.

- **Sensitive to Large Errors:** Like MSE, RMSE squares the errors, giving greater weight to larger deviations, which is useful when larger errors need to be emphasized.

- **Widely Used and Accepted:** RMSE is a standard metric in regression tasks and is commonly used for evaluating model performance, making it a familiar and widely understood measure.

# More on Gradient Vs loss

- Jupyter Notebook: Gradient Vs loss.ipynb

# Disadvantages of RMSE:

- **Sensitive to Outliers:** RMSE is highly sensitive to outliers due to the squaring of errors, which can lead to an overestimation of model error if outliers are present.

- **Less Intuitive Interpretation:** Although RMSE is in the same units as the target variable, the squaring of errors can make it less intuitive to interpret compared to simpler metrics like MAE.

- **May Over-penalize Large Errors:** RMSE can overemphasize large errors, which might not be desirable in situations where all errors should be treated more equally.

# R-squared ($R^2$)

$$R^2 = 1 - \frac{SS_{RES}}{SS_{TOT}} = 1 - \frac{\sum_i (y_i - \hat{y}_i)^2}{\sum_i (y_i - \overline{y})^2}$$

**Use Case**:

- $R^2$ represents the proportion of the variance in the dependent variable that is predictable from the independent variables.

- While not an error function per se, it is often used to assess the goodness-of-fit of the model.

# Advantages of R-squared:

- **Easy Interpretation:** R-squared provides a clear interpretation of model performance by representing the proportion of variance in the target variable explained by the model.
- **Standardized Measure:** R-squared is a widely recognized and commonly used metric in regression analysis, making it easy to compare different models.
- **Indicates Goodness of Fit:** A higher R-squared value indicates a better fit between the model and the data, showing how well the model captures the underlying trend.

## Original R-squared Formula:

$$R^2 = 1 - \frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{\sum_{i=1}^{n} (y_i - \bar{y})^2}$$

## Expanded Summation for R-squared:

For a dataset with $n$ observations, the formula can be expanded as follows:

$$R^2 = 1 - \frac{\frac{1}{n}\left((y_1 - \hat{y}_1)^2 + (y_2 - \hat{y}_2)^2 + (y_3 - \hat{y}_3)^2 + \cdots + (y_n - \hat{y}_n)^2\right)}{\frac{1}{n}\left((y_1 - \bar{y})^2 + (y_2 - \bar{y})^2 + (y_3 - \bar{y})^2 + \cdots + (y_n - \bar{y})^2\right)}$$

# Disadvantages of R-squared:

- **Insensitive to Overfitting:** R-squared does not account for model complexity, meaning it can increase with additional predictors, even if they do not contribute meaningfully, potentially leading to overfitting.
- **Misleading with Non-linear Relationships:** R-squared assumes a linear relationship between variables, making it less useful or misleading when the true relationship is non-linear.
- **Does Not Measure Predictive Power:** A high R-squared value doesn’t necessarily mean that the model will perform well on new, unseen data, as it only reflects the fit to the training data.

# Adjusted R-squared

Formula: $\text{Adjusted } R^2 = 1 - \left(\frac{1-R^2}{n-k-1}\right) \times (n-1)$

**Use Case**:

- N- number of samples
- K - features
- Adjusted $R^2$ accounts for the number of predictors in the model, preventing overfitting by penalizing the inclusion of unnecessary variables.

# Advantages of Adjusted R-squared:

- **Accounts for Model Complexity:** Adjusted R-squared adjusts for the number of predictors in the model, penalizing the addition of unnecessary variables, which helps prevent overfitting.

- **Better Comparison Between Models:** It allows for a more accurate comparison of models with different numbers of predictors, as it adjusts for the model complexity.

- **Reflects True Fit:** Adjusted R-squared provides a more realistic measure of model performance by considering both the goodness of fit and the number of predictors, making it a better indicator of how well the model generalizes.

# Disadvantages of Adjusted R-squared:

- **Complexity in Calculation:** Adjusted R-squared is less intuitive and slightly more complex to calculate than the standard R-squared, which might make it harder to interpret for beginners.

- **Limited Use in Non-linear Models:** Similar to R-squared, Adjusted R-squared assumes a linear relationship and may not be as informative or useful in non-linear models.

- **Does Not Account for All Overfitting:** While it adjusts for the number of predictors, Adjusted R-squared does not fully account for all forms of overfitting, particularly if irrelevant variables still slightly improve the fit.