Supervised Learning

Unit #2 & 3

Marwadi University

Department of Computer Engineering

Machine Learning

Sem 7

01CE0715

4 Credits

Prof. Urvi Bhatt

Course Outcomes

After completion of this course, students will be able to

- Understand machine-learning concepts.
- Understand and implement Classification concepts.
- Understand and analyse the different Regression algorithms.
- Apply the concept of Unsupervised Learning.
- Apply the concepts of Artificial Neural Networks.

Topics - Supervised Learning

Classification Techniques:

- Naive Bayes Classification
- Fitting Multivariate Bernoulli Distribution
- Gaussian Distribution and Multinomial Distribution
- K- Nearest Neighbours
- Decision tree
- Random Forest
- Ensemble Learning
- Support Vector Machines
- Evaluation metrics for Classification Techniques: Confusion Matrix, Accuracy, Precision, Recall, F1 Score, Threshold, AUC-ROC

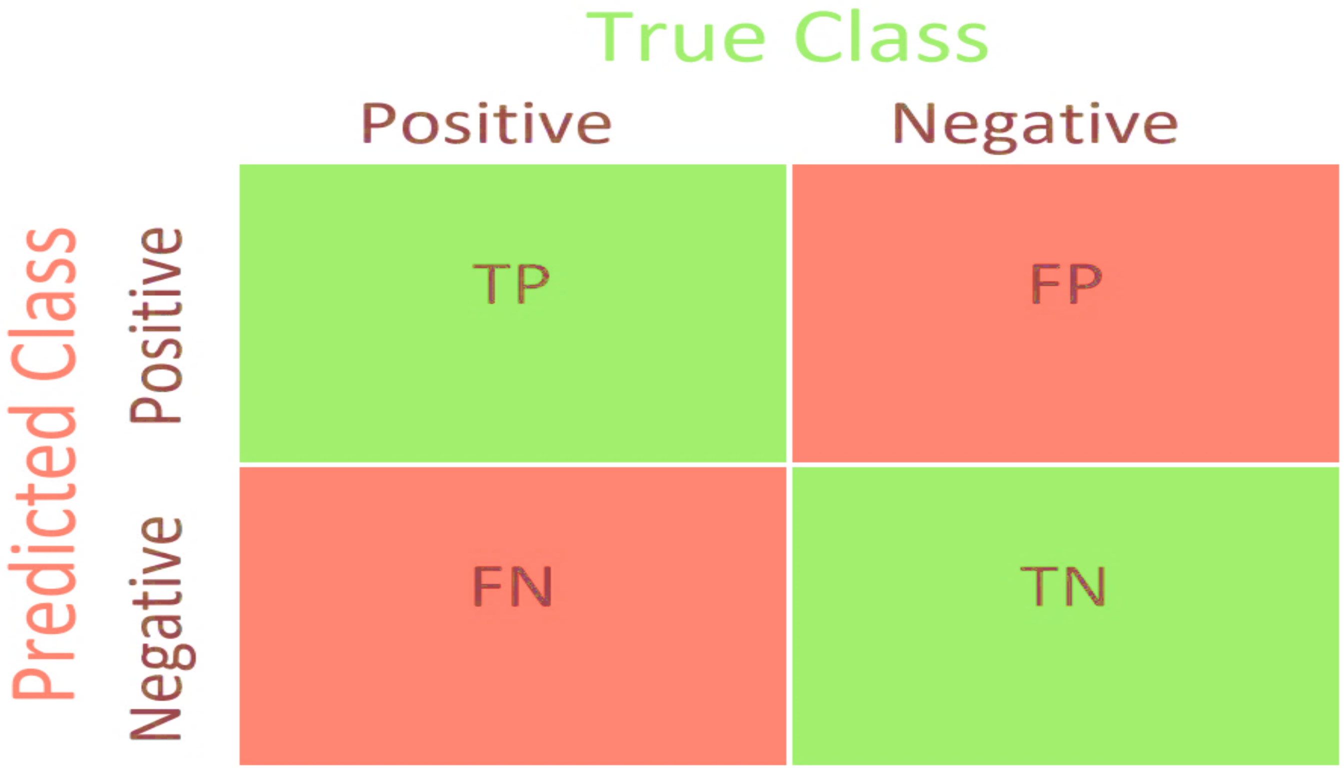
Regression Techniques:

- Basic concepts and applications of Regression
- Simple Linear Regression - Gradient Descent and Normal Equation Method
- Multiple Linear Regression
- Non-Linear Regression
- Linear Regression with Regularization
- Overfitting and Underfitting
- Hyperparameter tuning
- Evaluation Measures for Regression Techniques: MSE, RMSE, MAE, R2

Evaluation metrics for Classification Techniques

Confusion Matrix, Accuracy, Precision, Recall, F1 Score, Threshold, AUC-ROC

Confusion Matrix



- **True Positive (TP): (1✓)** the model predicts 1 and the actual class is 1
- **True Negative (TN): (0✓)** the model predicts 0 and the actual class is 0
- **False Positive (FP): (1✗)** the model predicts 1 but the actual class is 0
- **False Negative (FN): (0✗)** the model predicts 0 but the actual class is 1

Confusion Matrix

Search this file...

	Prediction	Actual value	Type	Explanation
1	Prediction	Actual value	Type	Explanation
2	1	1	True Positive	Predicted Positive and was Positive
3	0	0	True Negative	Predicted Negative and was Negative
4	1	0	False Positive	Predicted Positive but was Negative
5	0	1	False Negative	Predicted Negative but was Positive

med_prf_1.csv hosted with ❤ by GitHub   view raw

Examples of True/False Positive and Negative

Example:

- The cases in which the patients actually have heart disease and our model also predicted as having it are called the **True Positives.** For our matrix, True Positives = 43
- The cases in which the patients actually did not have heart disease and our model also predicted as not having it are called the **True Negatives.** For our matrix, True Negatives = 33.

Example:

- However, there are some cases where the patient actually has no heart disease, but our model has predicted that they do. This kind of error is the Type I Error, and we call the values **False Positives.** For our matrix, False Positives = 8
- Similarly, there are some cases where the patient actually has heart disease, but our model has predicted that he/she doesn't. This kind of error is a Type II Error, and we call the values **False Negatives.** For our matrix, False Negatives = 7

Accuracy

- Accuracy is the ratio of the total number of correct predictions and the total number of predictions.
- Can you guess what the formula for Accuracy will be?

$$\text{Accuracy} = \frac{\text{True Positive} + \text{True Negative}}{(\text{True Positive} + \text{False Positive} + \text{True Negative} + \text{False Negative})}$$

Accuracy

- For our model, Accuracy will be = 0.835.
- There might be other situations where our accuracy is very high, but our precision or recall is low

Precision

- Precision defines of all the predictions y=1, which ones are correct.
- In the simplest terms, Precision is the ratio between the True Positives and all the Positives. For our problem statement, that would be the measure of patients that we correctly identify as having a heart disease out of all the patients actually having it. Mathematically:

$$Precision = \frac{TP}{TP + FP}$$

$$Precision = \frac{True\ Positive(TP)}{True\ Positive(TP) + False\ Positive(FP)}$$

Precision

- What is the Precision for our model?
- Yes, it is 0.843, or when it predicts that a patient has heart disease, it is correct around 84% of the time.
- Precision also gives us a measure of the relevant data points.

Recall / Sensitivity/ True Positive Rate

- Recall defines of all the actual y=1, which ones did the model predict correctly
- The recall is the measure of our model correctly identifying True Positives. Thus, for all the patients who actually have heart disease, recall tells us how many we correctly identified as having a heart disease. Mathematically:

$$Recall = \frac{TP}{TP + FN}$$

$$Recall = \frac{True\ Positive(TP)}{True\ Positive(TP) + False\ Negative(FN)}$$

Recall / Sensitivity/ True Positive Rate

- For our model, Recall = 0.86.
- Recall also gives a measure of how accurately our model is able to identify the relevant data.
- We refer to it as Sensitivity or True Positive Rate.

F1-Score

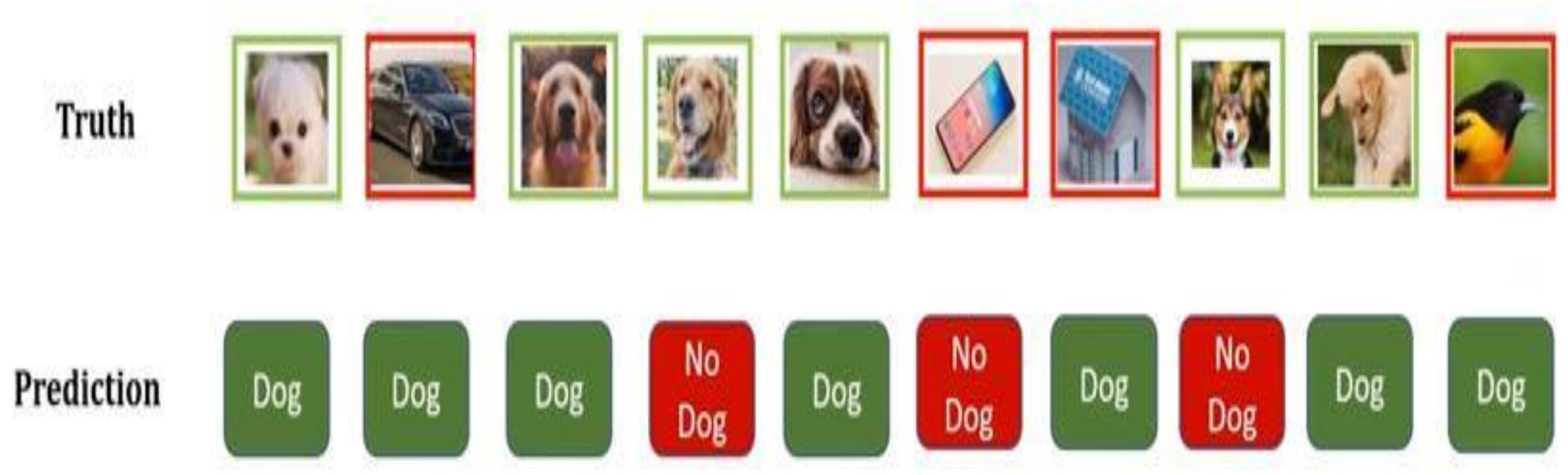
- F1-Score is a measure combining both precision and recall.
- It is generally described as the harmonic mean of the two. Harmonic mean is just another way to calculate an "average" of values, generally described as more suitable for ratios (such as precision and recall) than the traditional arithmetic mean.
- The formula used for F1-score in this case is:

$$2 * \frac{\text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}}$$

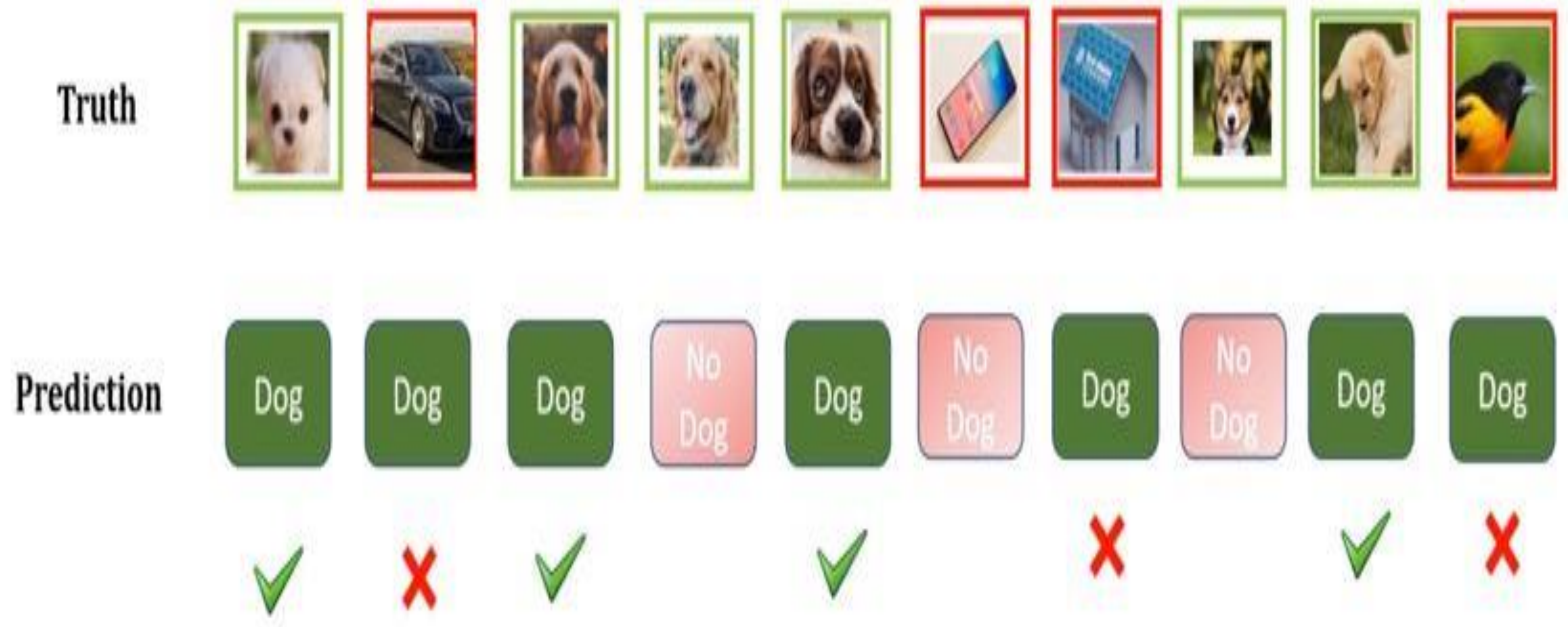
F1-Score

- For our model, F1-Score = ??

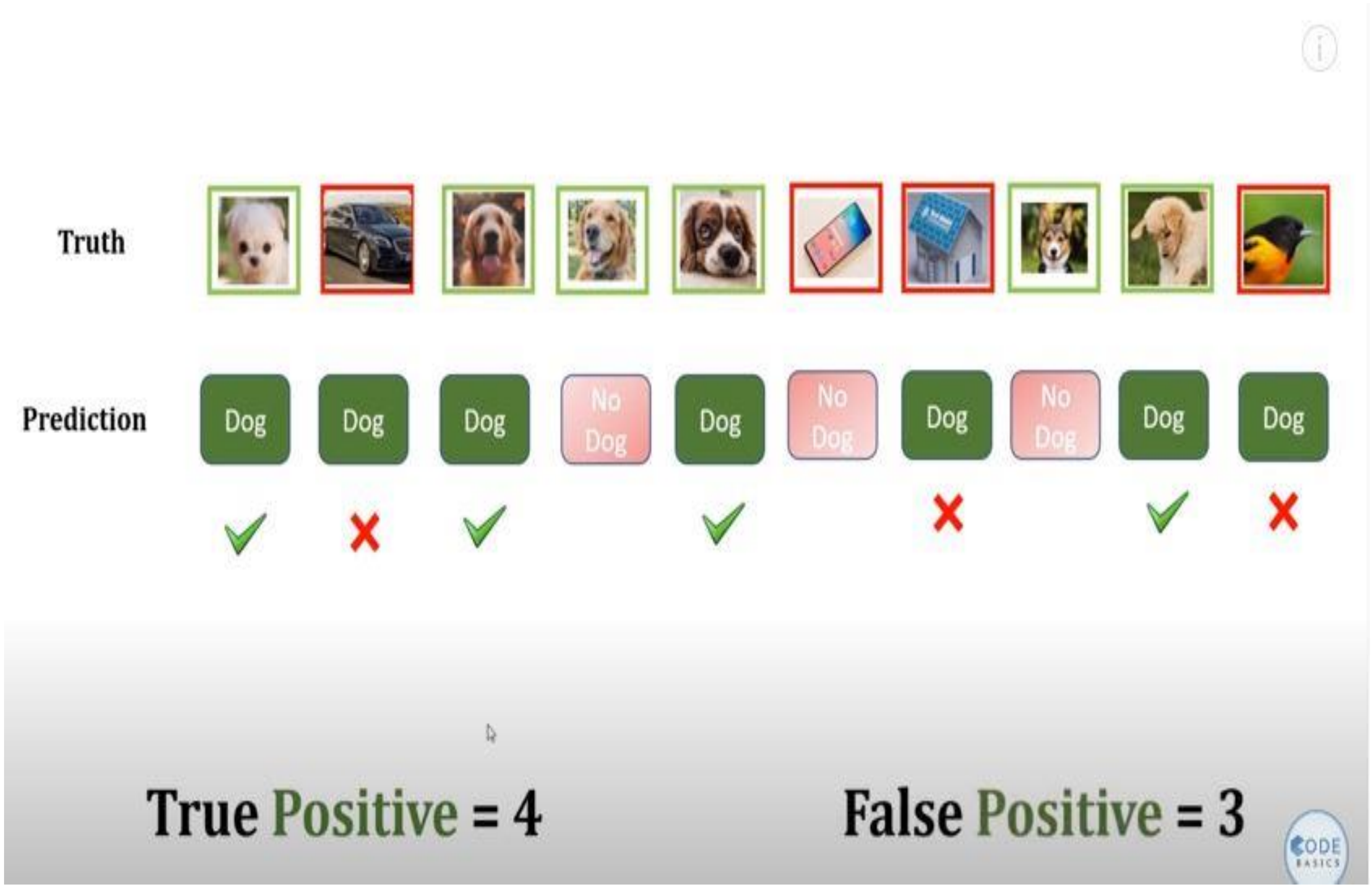
Example to understand precision and recall



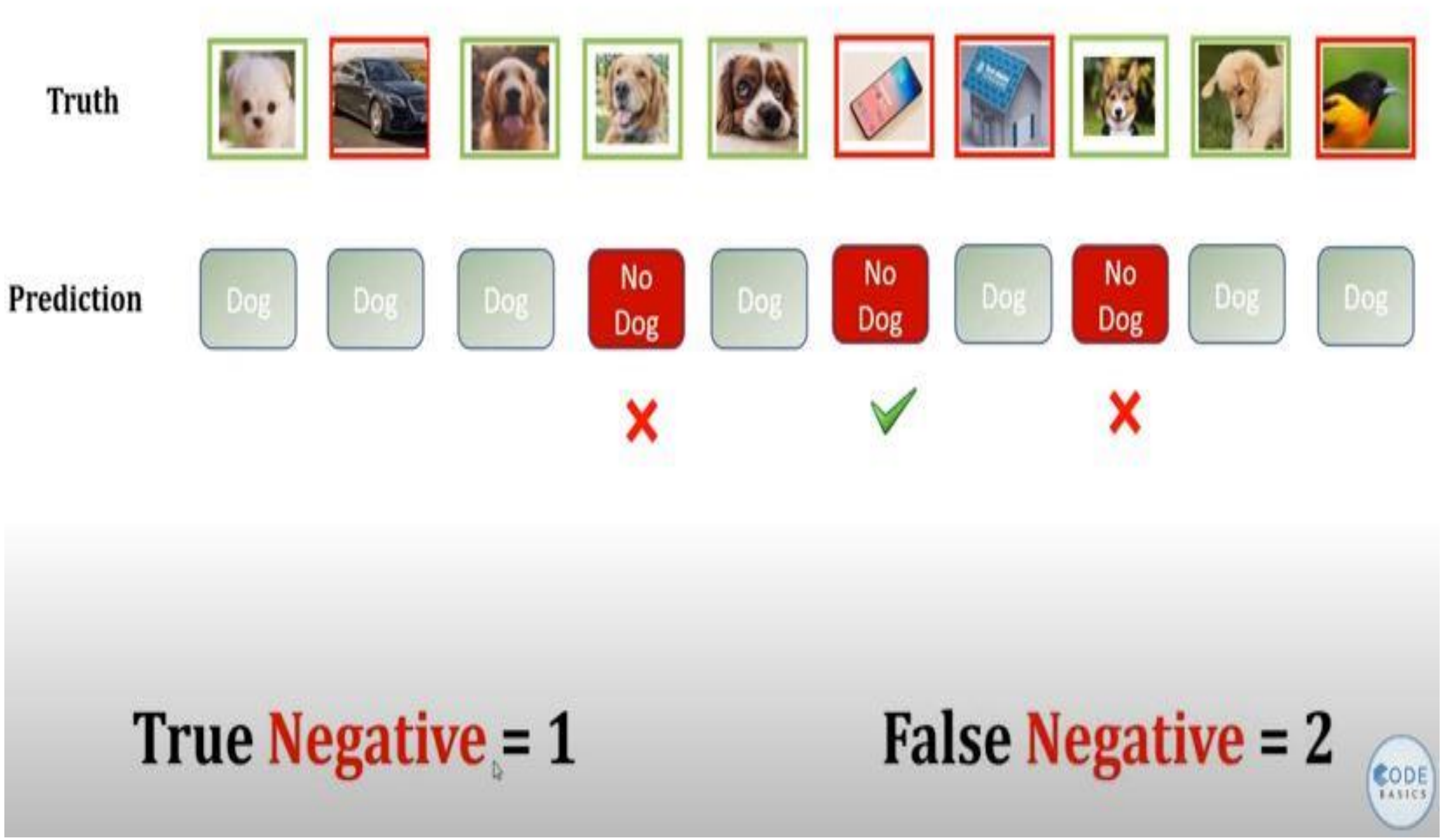
Example (Contd.)



Example (Contd.)



Example (Contd.)



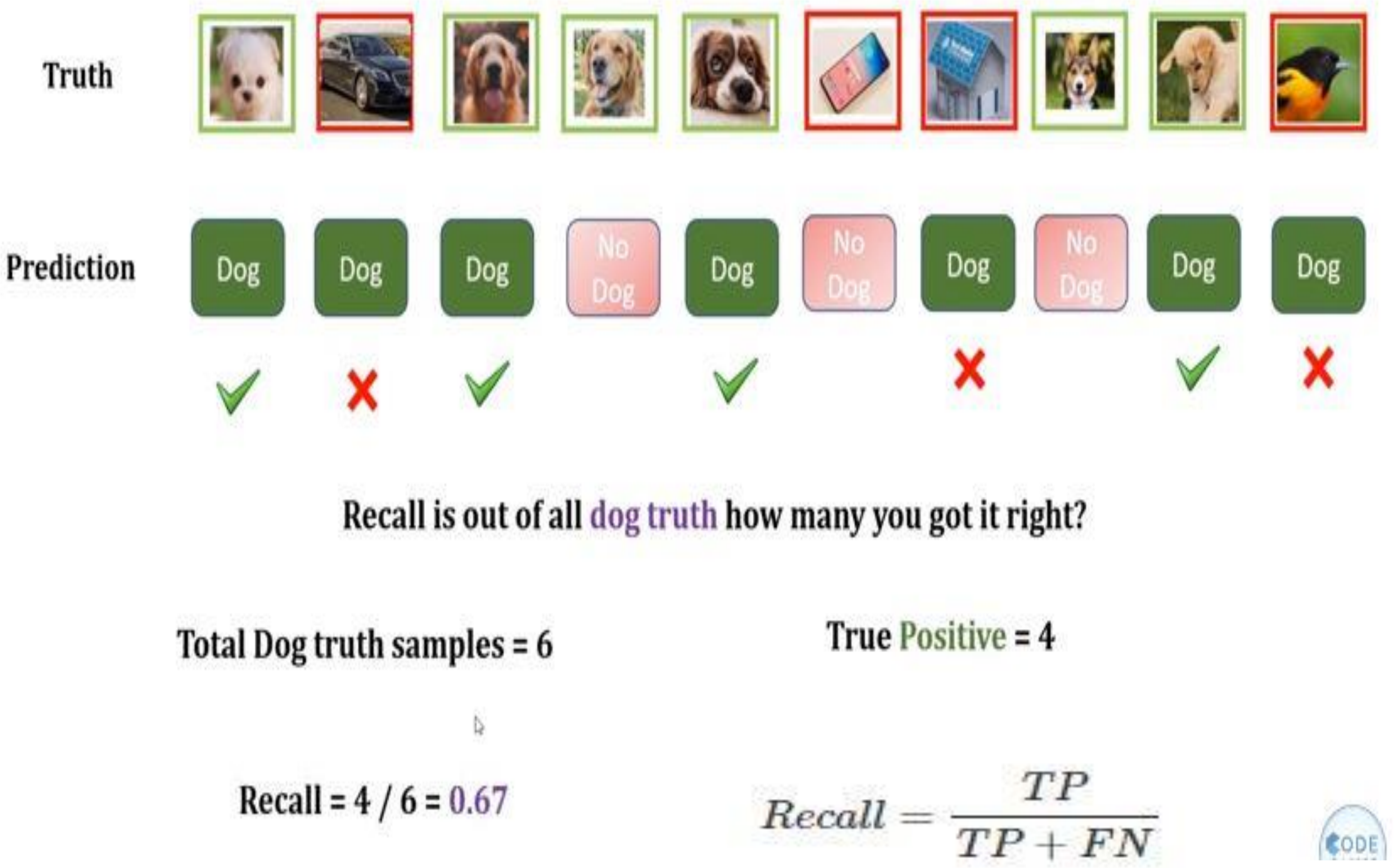
Example (Contd.)

$$Accuracy = \frac{\text{True Positive} + \text{True Negative}}{(\text{True Positive} + \text{False Positive} + \text{True Negative} + \text{False Negative})}$$

Example (Contd.)

$$Precision = \frac{TP}{TP + FP}$$

Example (Contd.)



Example (Contd.)

For precision, think about predictions as your base

For recall, think about truth as your base

Threshold / The Tradeoff

Precision and Recall are two important measures in evaluating the performance of a machine learning model, especially in classification problems.

- **Precision** focuses on how many of the items selected by the model are relevant.
- **Recall** focuses on how many of the relevant items are selected by the model.
- By changing the threshold value for the classifier confidence, one can adjust the precision and recall for the model.

Threshold / The Tradeoff

Want to Increase Precision Value → Try to Get Low False Positive → But Value of False Negative will Increase → Getting the less value For Recall

Precision ↑ → FP ↓ → FN ↑ → Recall ↓

Want to Increase Recall Value → Try to Get Low False Negative → But Value of False Positive will Increase → Getting the less value For Precision

Recall ↑ → FN ↓ → FP ↑ → Precision ↓

$$Precision = \frac{TP}{TP + FP}$$

$$Recall = \frac{TP}{TP + FN}$$

Video that are actually safe for Kids are correctly Identify as safe

Want to Increase Precision Value

Precision ↑

→

Try to Get Low False Positive

FP ↓

→

But Value of False Negative will Increase

FN ↑

→

Getting the less value For Recall

Recall ↓

- Goal: Mark Video as safe (no Violent or Adult Content)

Want to Increase Precision Value

(Video that are actually safe for Kids are correctly Identify as safe)

Precision ↑

→

Try to Get Low False Positive

(Videos that are not safe for the Kids but are incorrectly identify as safe by Model)

FP ↓

→

But Value of False Negative will Increase

(We are very cautions about marking of Video as safe so, Some time safe videos being wrongly blocked or marked as a unsafe due to any reason.

FN ↑

→

Getting the less value For Recall

Recall ↓

$$Precision = \frac{TP}{TP + FP}$$

$$Recall = \frac{TP}{TP + FN}$$

Patients Disease detection and if detected not having disease it should be correct only

Want to Increase Recall Value

Recall ↑

→

Try to Get Low False Negative

FN ↓

→

But Value of False Positive will Increase

FP ↑

→

Getting the less value For Precision

Precision ↓

Want to Increase Recall Value

(Patients Disease detection and if detected not having disease it should be correct only)

Recall ↑

→

Try to Get Low False Negative

(Identify as Patients having a disease but which is incorrect. So try to reduce case where Patients not having disease and we have correct identification)

FN ↓

→

But Value of False Positive will Increase

(we try to identify not having disease perfectly so might be we get small problem we are going to mark them as disease so in future we don't have any problem. Means making as a disease is More.

FP ↑

→

Getting the less value For Precision

Precision ↓

YouTube's restricted mode as an example to explain high precision

- **Goal:** Ensure that videos marked as safe for kids are indeed safe (no violent or adult content).

- True Positives (TP): Videos that are actually safe for kids and are correctly identified as safe by the model.
- True Negatives (TN): Videos that are not safe for kids and are correctly identified as not safe by the model
- False Positives (FP): Videos that are not safe for kids but are incorrectly identified as safe by the model.
- False Negatives (FN): Videos that are safe for kids but are incorrectly identified as not safe by the model.
- .

Example (Contd.)

- **Precision** is the ratio of true positives to the sum of true positives and false positives:

$$Precision = \frac{TP}{TP + FP}$$

- High precision means that when the model identifies a positive case, it is very likely correct.
- True Positives (TP): Videos that are actually safe for kids and are correctly identified as safe by the model.
- False Positives (FP): Videos that are not safe for kids but are incorrectly identified as safe by the model.

Example.. (Contd.)

To achieve high precision, YouTube's restricted mode model would:

- Only mark videos as safe if it is very confident they are appropriate for kids.
- This is achieved by being very cautious, which might result in **some safe videos being wrongly blocked (more false negatives),** but it ensures that almost no inappropriate videos are marked as safe.

Example.. (Contd.)

- If **more false negatives are there, Value of Recall will be low.**

$$Recall = \frac{TP}{TP + FN}$$

So, if we want to achieve a High Precision , the value of Recall is getting Low.

YouTube's restricted mode as an example to explain high recall

- **Goal:** Ensure that all inappropriate videos (violent or adult content) are correctly identified and blocked from kids.

- True Positives (TP): Videos that are actually Not safe for kids and are correctly identified as Not safe by the model.
- True Negatives (TN): Videos that are safe for kids and are correctly identified as safe by the model
- False Positives (FP): Videos that are safe for kids but are incorrectly identified as Not safe by the model.
- False Negatives (FN): Videos that are not safe for kids but are incorrectly identified as safe by the model.

YouTube's restricted mode as an example to explain high recall

- **Goal:** Ensure that all inappropriate videos (violent or adult content) are cor kids.

$$Recall = \frac{TP}{TP + FN}$$

High Recall Focus:

- **High Recall** means that the model successfully identifies most, if not all, inappropriate videos. This reduces the number of false negatives (FN).

Example.. (Contd.)

To achieve high recall, YouTube's restricted mode model would:

- Be very inclusive and strict about blocking videos.
- Aim to catch every single inappropriate video, even if it means sometimes blocking safe videos.

Result:

- **Few false negatives (FN):** Most inappropriate videos will be correctly identified and blocked.
- **More false positives (FP):** Some safe videos might be wrongly blocked, but this is a tradeoff to ensure high recall.

Example.. (Contd.)

- If **more false Positive are there, Value of Precision will be low.**

$$Precision = \frac{TP}{TP + FP}$$

- So, if we want to achieve a High Recall, the value of Precision is getting Low.

Precision / Recall tradeoff

- Unfortunately, you can't have both precision and recall high. If you increase precision, it will reduce recall, and vice versa. This is called the **precision/recall tradeoff**.

AUC-ROC

- An **ROC curve** (**receiver operating characteristic curve**) is a graph showing the performance of a classification model at all classification thresholds. This curve plots two parameters:
  - True Positive Rate
  - False Positive Rate
- **True Positive Rate** (**TPR**) is a synonym for recall and is therefore defined as follows:

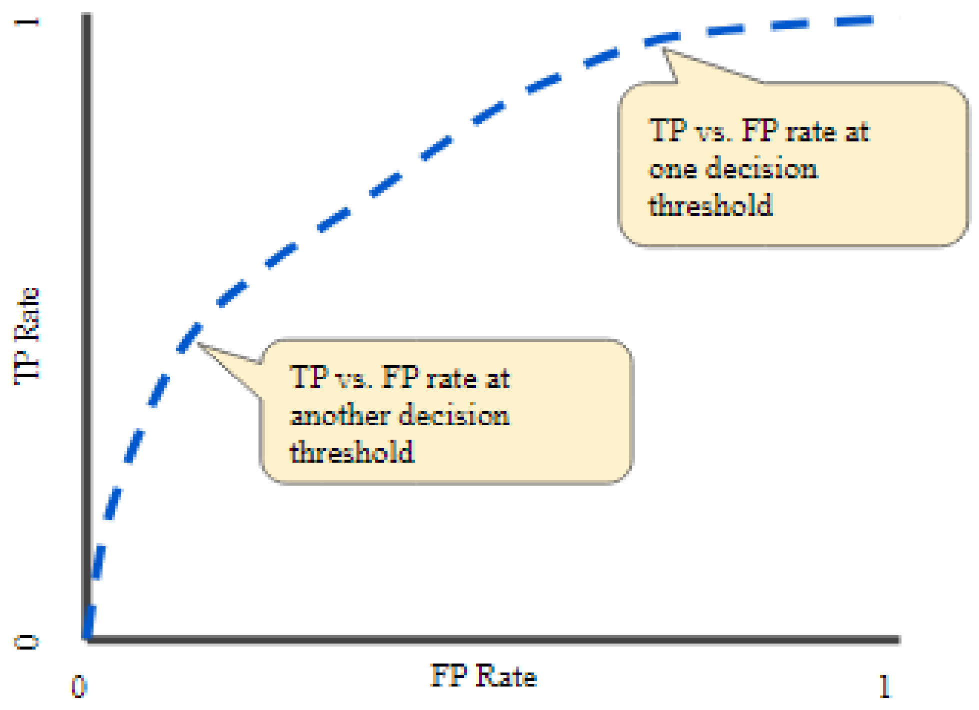
$$TPR = \frac{TP}{TP + FN}$$

- **False Positive Rate** (**FPR**) is defined as follows:

$$FPR = \frac{FP}{FP + TN}$$

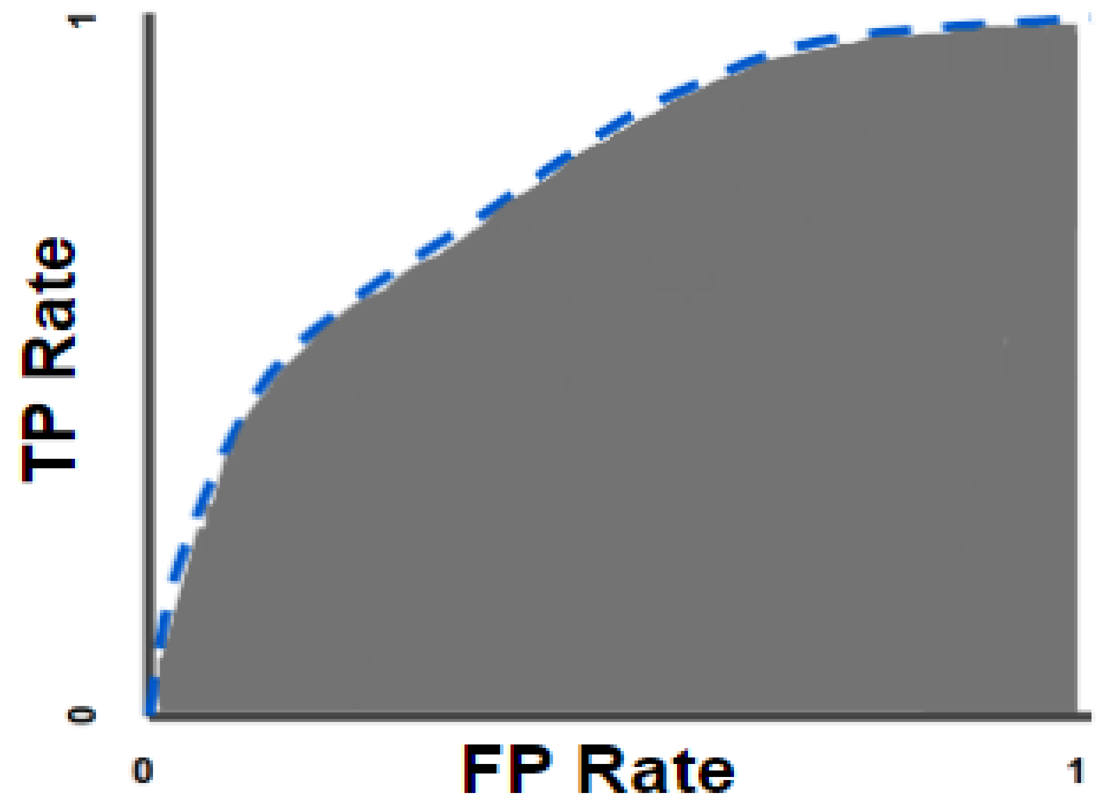
AUC-ROC

- An ROC curve plots TPR vs. FPR at different classification thresholds. Lowering the classification threshold classifies more items as positive, thus increasing both False Positives and True Positives. The following figure shows a typical ROC curve.



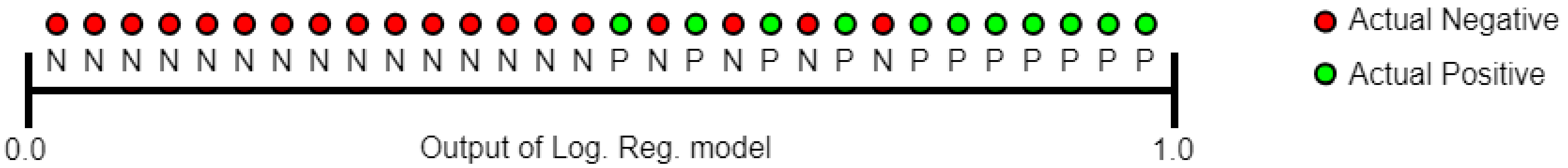
AUC-ROC

- **AUC** stands for "Area under the ROC Curve." That is, AUC measures the entire two-dimensional area underneath the entire ROC curve (think integral calculus) from (0,0) to (1,1).



AUC-ROC

- AUC represents the probability that a random positive (green) example is positioned to the right of a random negative (red) example.
- AUC ranges in value from 0 to 1. A model whose predictions are 100% wrong has an AUC of 0.0; one whose predictions are 100% correct has an AUC of 1.0.



Evaluation Measures for Regression Techniques

MSE, RMSE, MAE, R2

Evaluation Measures for Regression Techniques

Common regression evaluation metrics for regression include

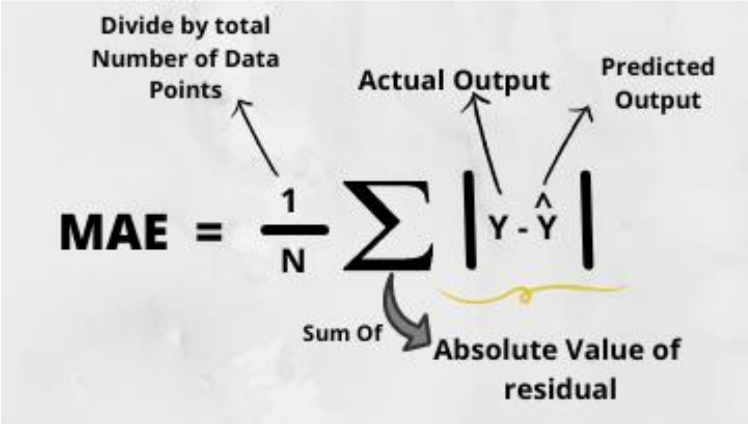
- Mean Absolute Error (MAE)
- Mean Squared Error (MSE)
- Root Mean Squared Error (RMSE)
- R-squared (Coefficient of Determination) (R2)

Why We Require Evaluation Metrics?

- In simple words, Regression can be defined as a Machine learning problem where we have to predict continuous values like price, Rating, Fees, etc.
- It is necessary to obtain the accuracy on training data, But it is also important to get a genuine and approximate result on unseen data otherwise Model is of no use.

Mean Absolute Error (MAE)

- MAE is a very simple metric which calculates the absolute difference between actual and predicted values.
- MAE is basically a mistake made by the model known as an error.
- so, sum all the errors and divide them by a total number of observations And this is MAE.
- we aim to get a minimum MAE because this is a loss.



Mean Squared Error (MSE)

- Mean squared error states that finding the squared difference between actual and predicted value.
- It represents the squared distance between actual and predicted values.
- We perform squared to avoid the cancellation of negative terms and it is the benefit of MSE.

$$MSE = \frac{1}{n} \Sigma \underbrace{\left(y - \hat{y} \right)^2}_{\text{The square of the difference between actual and predicted}}$$

Root Mean Squared Error (RMSE)

- As RMSE is clear by the name itself, that it is a simple square root of mean squared error.

R Squared (R2) (Coefficient of Determination)

- MAE and MSE depend on the context as we have seen whereas the R2 score is independent of context.
- with help of R squared we have a baseline model to compare a model which none of the other metrics provides.
- The same we have in classification problems which we call a threshold which is fixed at 0.5.
- Hence, R2 squared is also known as Coefficient of Determination or sometimes also known as Goodness of fit.

R Squared (R2) (Coefficient of Determination)

- The value of R-square lies between 0 to 1.
- Where we get R-square equals 1 when the model perfectly fits the data and there is no difference between the predicted value and actual value.
- However, we get R-square equals 0 when the model does not predict any variability in the model and it does not learn any relationship between the dependent and independent variables.

$$\textbf{R2 Squared} = 1 - \frac{SSr}{SSm}$$

SSr = Squared sum error of regression line

SSm = Squared sum error of mean line

R Squared (R2) (Coefficient of Determination)

- So we can conclude that as our regression line moves towards perfection, R2 score move towards one. And the model performance improves.
- The normal case is when the R2 score is between zero and one like 0.8 which means your model is capable to explain 80 per cent of the variance of data.

$$\textbf{R2 Squared} = 1 - \frac{SSr}{SSm}$$

SSr = Squared sum error of regression line

SSm = Squared sum error of mean line

Any Queries..??

Thank you