Evaluation Metrices for ML models

Author: ***Aden Rajput***

Affiliation: ***E-Academy***

Email: adenrajput@gmail.com

Choosing the right evaluation metric in machine learning is essential because it helps you assess how well your model is performing and whether it's suitable for your specific problem. Here's a simplified explanation of some common evaluation metrics and when to use them.

1. Accuracy:

Imagine you have a bag of different-colored marbles, and you want to sort them into red and blue. Accuracy measures how many marbles you correctly sort into their correct colors. If you get 90 out of 100 marbles right, your accuracy is 90%.

> > **When to Use:** Use accuracy when the classes (categories) you're trying to predict are roughly balanced, meaning there are a similar number of examples in each class. It's a simple and easy-to-understand metric.

2. Precision:

Precision is like being very careful. It measures how many of the marbles you sorted as red are actually red. So, if you say 90 marbles are red, but only 85 of them are truly red, your precision is 85%.

> > **When to Use:** Use precision when you want to make sure that when your model makes a positive prediction (like "this email is spam"), it's very likely to be correct. It's important when false positives are costly or undesirable.

3. Recall:

Recall is like not missing anything important. It measures how many of the truly red marbles you correctly identified as red. If there are 100 red marbles, but you only found 85 of them, your recall is 85%.

> > **When to Use:** Use recall when you want to make sure that your model doesn't miss many positive examples. It's important when false negatives are costly or undesirable.

4. F1-Score:

The F1-Score combines precision and recall into one number. It's like getting a balance between being careful and not missing important things. It's a way to find a good compromise between the two. A higher F1-Score means you're doing well in both precision and recall.

> > **When to Use:** Use the F1-Score when you want to consider both false positives and false negatives in your evaluation. It's helpful when you need to find a good balance between precision and recall.

5. Mean Absolute Error (MAE):

Imagine you're guessing the weight of different animals. The MAE tells you, on average, how much your guesses are off by. If your guesses are off by 5 pounds on average, your MAE is 5.

> > **When to Use:** Use MAE when you're dealing with a regression problem (predicting numerical values). It's easy to understand because it gives you the average difference between your predictions and the actual values.

6. Mean Squared Error (MSE):

Similar to MAE, but squared. It penalizes larger errors more. If you're off by 25 pounds for one animal and 1 pound for another, the squared error for the first animal is much bigger.

> > **When to Use:** Like MAE, use MSE for regression problems. It's sensitive to large errors and is often used when you want to emphasize the impact of outliers.

7. R-squared (R^2):

R^2 measures how well your predictions explain the variation in the data. It's like a report card for your model. An R^2 of 1 means your model perfectly explains the data, while 0 means it doesn't explain it at all.

> > **When to Use:** Use R^2 in regression problems to understand how well your model fits the data. A higher R^2 indicates a better fit.

The choice of evaluation metric depends on your specific problem, your priorities, and the nature of your data. It's often a trade-off between different aspects of performance, such as precision and recall. Select the metric that aligns best with what you care about most in your application.