Evaluation Metrics for Regression models- MAE Vs MSE Vs RMSE vs RMSLE

by akhilendra leave a comment

If you are working on a regression-based machine learning model like linear regression, one of the most importar tasks is to select an appropriate evaluation metric.

In fact, if you are working on a machine learning projects in general or preparing to become a data scientist, it's kind of must for you to know the top evaluation metrics.

These are also called loss functions.

There are two kinds of machine learning problems – classification and regression.

And these have different kind of loss functions.

In this post, I am going to talk about regression's loss functions.

Since every project or data set is different, we must select appropriate evaluation metrics. Usually, more than 1 metrics is required to evaluate a machine learning model.

Instead of including all the loss functions or evaluation metrics for regression machine learning models, I will try t focus on top loss functions.

Evaluation Metrics or Loss functions for Regression

- Mean absolute error (MAE)
- Mean squared error (MSE)
- Root mean square error (RMSE)
- Root mean square log error (RMSLE)

Before we start with loss functions, you need to understand what we are trying to do here. In a typical regression-based machine learning model, our model will produce continuous values (predicted value).

Our primary objective is to keep these predicted values closer to actual values.

Predicted values are denoted by y hat ().

Actual values are denoted by y.

Error = y - y hat

$$\text{Residual Error} = y - \hat{y}$$

So, whenever we are talking about error in this post, we are talking about this error. And yes, ideal condition (hypothetical one) is that this error (difference) is 0, which means our model can predict all values correctly (whic is not going to happen).

Let’s start with mean absolute error.

Mean absolute error (MAE)

In simple terms, mean absolute error is the sum of absolute/positive errors of all values. So, if there are 5 values in our data set, we find out the difference between the actual value and predicted values for all 5 values and take their positive value. So even if the difference between actual and predicted value is negative, we take positive value for calculation.

So we take the positive value of all errors, add them and find out their mean.

Mean absolute error illustration;

Actual Value (y)	Predicted Value (y hat)	Error (difference)	Absolute Error	
100	130	-30	30	
150	170	-20	20	
200	220	-20	20	
250	260	-10	10	
300	325	-25	25	
			21	**Mean**
Note- You take the absolute value of error which is the positive value, therefore -30 becomes 30				

MAE is the sum of **absolute** differences between actual and predicted values. It doesn’t consider the direction, that is, positive or negative.

When we consider directions also, that is called Mean Bias Error (MBE), which is a sum of errors(difference).

Formula for mean absolute Error or MAE is represented by;

$$MAE = \frac{\sum_{i=1}^{n}|y_i-\hat{y}_1|}{n}$$

Mean Square Error (MSE)

Mean square error is always positive and a value closer to 0 or a lower value is better. Let’s see how this this is calculated;

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

Let’s use the last illustration to understand it better.

Actual Value (y)	Predicted Value (y hat)	Error (difference)	Squared Error	
100	130	-30	900	
150	170	-20	400	
200	220	-20	400	
250	260	-10	100	
300	325	-25	625	
			485	Mean

So if we were to run a model with different parameters/independent variables, model with lower MSE will be deemed better.

We will look at its comparison with other loss functions in a while in this post. First quic

Root mean square error (RMSE)

Square root of MSE yields root mean square error (RMSE). So it's formula is quite similar to what you have seen with mean square error, it's just that we need to add a square root sign to it;

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

It is the standard deviation of error (residual error).

it indicates the spread of the residual errors. It is always positive, and a lower value indicates better performance Ideal value would be 0 but it is never achieved.

Actual Value (y)	Predicted Value (y hat)	Error (difference)	Squared Error	

100	130	-30	900	
150	170	-20	400	
200	220	-20	400	
250	260	-10	100	
300	325	-25	625	
			485	**Mean**
			22.02271555	**Square root of mean**

Effect of each error on RMSE is directly proportional to the squared error therefore, RMSE is sensitive to outliers and can exaggerate results if there are outliers in the data set.

Before moving to their comparison, I just want to mention one more evaluation metric and that is Root mean squared log error (RMSLE)

Root mean squared log error (RMSLE)

Root mean squared log error is basically RMSE but calculated at logarithmic scale. So, if you understand the above mentioned 3 evaluation metrics, you won't have any problem understanding RMSLE or most other evaluation metric or loss functions used in regression-based machine learning model.

While calculating RMSLE, 1 is added as constant to actual and predicted values because they can be 0 and log of 0 is undefined. Overall formula remains same. Standard denotation for RMSLE is;

$$\text{RMSLE}=\sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(\log\left(y_i+1\right)-\left(\widehat{y+1}\right)\right)^2}$$

In this illustration, I have used log for calculation;

Actual Value (y)	Predicted Value (y hat)	Actual + 1	Predicted + 1	log (Actual)	Log (Predicted)	Error (difference)	Squared Error	
100	130	101	131	2.004321374	2.117271296	-0.112949922	0.012757685	
150	170	151	171	2.178976947	2.23299611	-0.054019163	0.00291807	
200	220	201	221	2.303196057	2.344392274	-0.041196216	0.001697128	
250	260	251	261	2.399673721	2.416640507	-0.016966786	0.000287872	
300	325	301	326	2.478566496	2.5132176	-0.034651104	0.001200699	

							0.061418977	**Squre root of mean**

Let’s look their difference now.

MAE vs MSE vs RMSE Vs RMSLE

In terms of comparison, primary differences are between MAE & MSE because they both are calculated in different ways. RMSE & RMSLE are extension of MSE therefore they share lots of properties with MSE.

Mean absolute Error (MAE)	**Mean square Error (MSE)**	**Root mean square error (RMSE)**	**Root mean square log Error (RMSLE)**
It doesn’t account for the direction of the value. Even if value is negative, positive value is used for calculation.	It does account for positive or negative value.	It does account for positive or negative value.	It does account for positive or negative value.
	RMSE & MSE share many properties with MSE because RMSE is simply the square root of MSE.	RMSE & MSE share many properties with MSE because it is simply the square root of MSE.	
MAE is less biased for higher values. It may not adequately reflect the performance when dealing with large error values.	MSE is highly biased for higher values.	RMSE is better in terms of reflecting performance when dealing with large error values.	
		RMSE is more useful when lower residual values are preferred.	
MAE is less than RMSE		RMSE tends to be	

as the sample size goes up.		higher than MAE as the sample size goes up.	
MAE doesn't necessarily penalize large errors.	MSE penalize large errors.	RMSE penalize large errors.	RMSLE doesn't penalize large errors. It is usually used when you don't want to influence the results if there are large errors. RMSLE penalize lower errors.
MAE is more useful when the overall impact is proportionate to the actual increase in error. For example- if error values go up to 6 from 3, actual impact on the result is twice. It is more common in financial industry where a loss of 6 would be twice of 3.		RMSE is more useful when the overall impact is disproportionate to the actual increase in error. For example- if error values go up to 6 from 3, actual impact on the result is more than twice. This could be common in clinical trials, as error goes up, overall impact goes up disproportionately.	
		When actual and predicted values are low, RMSE & RMSLE are usually same.	When actual and predicted values are low, RMSE & RMSLE are usually same.
		When either of actual or predicted values	When either of actual or predicted value

		are high, RMSE > RMSLE.	are high, RMSE > RMSLE.

MAE vs MSE vs RMSE Vs RMSLE Conclusion

I have mentioned only important differences. If there is no valid point for one, I haven't included in the above table and that's why we have empty cells in the table.

Few important points to remember when using loss functions for your regression;

- Never compare apple with oranges, that is, never compare different metrics with each other. For example don't compare values of MSE with MAE or others. They would be different.
- Try to use more than 1 loss function.
- Always calculate evaluation metrics (loss functions) for both testing and training data set.
- Compare evaluation metrics between test and training data set. There shouldn't be a huge difference between them. If there is, there is a problem with your model. For example- if you are using RMSE, calculate RMSE for testing and training data set. There should be huge difference between these values for this data set.
- If you have outlier in the data and you want to ignore them, MAE is a better option but if you want to account for them in your loss function, go for MSE/RMSE.

Questions or feedback? Please leave your comments.

filed under: data science tagged with: data science, evaluation metrics, loss function, machine learning, mae, mean absolute error, mean square error, mse, regression model, rmse, rmsle, root mean square error, root mean square log error

About akhilendra

Hi, I'm Akhilendra and I write about Product management, Business Analysis, Data Science, IT & Web. Join me on Twitter, Facebook & Linkedin

Speak Your Mind

Name *

Email *

Website

Post Comment

Search this website … Search

Access courses on Android

Recent Posts

- Top machine learning algorithms for beginners
- Business analyst-Product owner-Scrum master interview questions
- How to become a great Scrum Master? Skills-Roles & Responsibilities
- GATE EXAM 2021 IITB-Exam dates, Eligibility & Details
- Website Redesign – When Is The Best Time To R

Got questions?

- Learning Computer File Structure

What is Activation Function & Why you need it in Neural Networks

Amazing SaaS Startup Growth Strategies You Shouldn't Miss

How to find jobs (Naukri)

What Business Owners Need to Know About Scrum Agile Implementation

Important Stuff

Articles

Business

Business

Got questions?

Data Science

Gadgets

Internet Marketing

IPO

Personal Finance

Programming

SEO

Social Media Marketing

Technology

Web Hosting

Wordpress